



Europa-Universität
Flensburg

Europa-Universität Flensburg

Institut für Gesundheits-, Ernährungs- und Sportwissenschaften

Abteilung Gesundheitspsychologie und Gesundheitsbildung

Entwicklung und Evaluation eines Peer Review- Verfahrens zur Qualitätssicherung der Sozialmedizinischen Begutachtung

Dissertation zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.)

Vorgelegt von:	André Strahl, M.Sc. Rehabilitationspsychologie
Abgabedatum:	22.03.2019
Erstgutachter:	Prof. Dr. Petra Hampel
Zweitgutachter:	Prof. Dr. Matthias Morfeld

*„Man hat niemals eine sehr beschäftigte Person gesehen,
die unglücklich war.“*

Dorothy Dix (1861-1951)

Auf dem langen Weg zur Fertigstellung dieser Arbeit haben mich über die Jahre viele Menschen begleitet, denen ich an dieser Stelle meinen Dank aussprechen möchte.

Zunächst möchte ich Herrn PD Dr. Heiner Vogel danken, der mir direkt nach dem Studium meine erste Arbeitsstelle anbot und mir mit seinen Projekten den Grundstein für die vorliegende Arbeit und meinen Einstieg in die Welt der Wissenschaft und Forschung überhaupt erst ermöglicht hat.

In diesem Zuge möchte ich mich zudem bei allen Projektbeteiligten von der Deutschen Rentenversicherung für ihre Unterstützung und Mitarbeit im Rahmen der verschiedenen sozialmedizinischen Projekte bedanken. Besonderen Dank gebührt den Projektkoordinatoren von der Deutschen Rentenversicherung Bund Frau Dr. Helga Mai, Frau Dr. Anette Müller-Garnn und Herrn Dipl.-Psych. Jörg Gehrke, ohne die die Entwicklung, Umsetzung und Evaluation des Qualitätssicherungsverfahrens nicht möglich gewesen wäre.

Darüber hinaus danke ich meinem wertgeschätzten Würzburger Kollegen Herrn Dipl.-Psych. Christian Gerlich, der mir damals wie heute mit Rat und Tat zur Seite steht und stets für die kompliziertesten Probleme eine pragmatische Lösung parat hat.

Mein ganz besonderer Dank gilt Frau Prof. Dr. Petra Hampel, die durch ihren Vorschuss an Vertrauen in meine Person, ihre Ratschläge, ihr Engagement, ihre schnellen E-Mailantworten und ihre strukturierte und zielorientierte Arbeitsweise die Promotion an der Europa-Universität Flensburg für mich erst ermöglicht hat.

Zudem danke ich meinem Zweitgutachter Herrn Prof. Dr. Matthias Morfeld, der mich bereits seit dem 3. Semester meines Bachelor-Studiums beruflich begleitet. Durch seine erfrischende Art, Didaktik und Themenauswahl in den Vorlesungen und Seminaren hat er es in nur vier Semestern geschafft, aus mir einen begeisterten Anhänger der (Rehabilitations-)Forschung und Wissenschaft zu machen, obgleich ich vor unserem ersten Treffen fest davon überzeugt war, später einmal als klinischer Psychologe in einer Psychiatrie zu arbeiten.

Zu guter Letzt danke ich meiner Freundin, Verlobten und Bald-Ehefrau Katharina Barleben für ihre Unterstützung, Motivation und ihrer steten und vorbildlichen Lebensführung im Sinne der Maslowschen Individualbedürfnisse und Selbstverwirklichung.

In der vorliegenden Arbeit wird aus Gründen der besseren Lesbarkeit auf eine geschlechtsspezifisch differenzierte Schreibweise verzichtet. Dies bedeutet, dass die Verwendung der männlichen Form stets auch das weibliche Geschlecht umfasst.

Inhaltsverzeichnis

Zusammenfassung	1
1 Theoretischer Teil	4
1.1 Problemstellung und Überblick	4
1.2 Die gesetzliche Rentenversicherung	8
1.2.1 Leistungen der gesetzlichen Rentenversicherung	10
1.2.2 Rente wegen Erwerbsminderung	12
1.3 Chronische Erkrankungen	16
1.3.1 Häufige Diagnosen im Zusammenhang mit EM-Rente	16
1.3.2 Entwicklung der Morbidität	22
1.3.3 Risiko- und Schutzfaktoren	25
1.4 Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF)	28
1.4.1 Modell der ICF	29
1.4.2 Anwendung der ICF in der Sozialmedizinischen Begutachtung	30
1.5 Sozialmedizinische Begutachtung	35
1.5.1 Rahmenbedingungen der Begutachtung	36
1.5.2 Das ärztliche Gutachten	39
1.5.3 Evidenz in der Sozialmedizinischen Begutachtung	42
1.6 Qualitätssicherung in der Sozialmedizinischen Begutachtung	46
1.6.1 Definition und Begriffsbestimmung	47
1.6.2 Maßnahmen zur Qualitätssicherung in der Begutachtung	49
1.6.3 Peer Review-Verfahren	51
1.7 Peer Review in der Sozialmedizinischen Begutachtung	57
1.7.1 Manual zum Peer Review-Verfahren	58
1.7.2 Explorative Analyse der Interrater-Reliabilität	61
1.8 Zusammenfassung und Relevanz des Forschungsvorhabens	63
2 Aufbau des empirischen Teils	67
3 Teilstudie 1: Systematische Literaturrecherche	69
3.1 Fragestellung	72
3.2 Untersuchungsdesign	73
3.2.1 Elektronische Datenbankrecherche	73
3.2.2 Suchstrategie	74
3.2.3 Beurteilung der Studienqualität	79
3.2.4 Beurteilung der Evidenz	79

3.3	Ergebnisse	81
3.3.1	Merkmale und inhaltliche Charakteristika der eingeschlossenen Studien ...	82
3.3.2	Qualität der eingeschlossenen Studien.....	101
3.3.3	Evidenz der eingeschlossenen Studien	101
3.4	Diskussion	106
3.4.1	Allgemeine Diskussion.....	106
3.4.2	Beantwortung der Fragestellung	109
3.5	Handlungsempfehlungen	112
3.6	Kernbotschaft	115
4	Teilstudie 2: Interrater-Reliabilität des Peer Review Verfahrens	116
4.1	Fragestellung	117
4.2	Untersuchungsdesign	118
4.2.1	Rekrutierung	119
4.2.2	Schulung	119
4.2.3	Peer Review	120
4.2.4	Kalibrierung	125
4.2.5	Reliabilitätsanalyse	125
4.2.6	Qualitative Inhaltsanalyse	129
4.3	Ergebnisse	132
4.3.1	Übereinstimmung der Probegutachten	132
4.3.2	Interrater-Reliabilität der Einzelkriterien	133
4.3.3	Interrater-Reliabilität des übergeordneten Kriteriums	134
4.3.4	Beurteilungsunterschiede zwischen den Peers.....	136
4.3.5	Qualitative Inhaltsanalyse der Einzelkriterien.....	136
4.3.6	Qualitative Inhaltsanalyse des übergeordneten Kriteriums.....	158
4.3.7	Exemplarische Verknüpfung qualitativer und quantitativer Daten	160
4.3.8	Kalibrierungsworkshop	162
4.4	Diskussion	164
4.4.1	Allgemeine Diskussion.....	164
4.4.2	Beantwortung der Fragestellung	169
4.5	Handlungsempfehlungen	173
4.5.1	Empfehlungen für das Manual zum Peer Review	173
4.5.2	Empfehlung für die Peer-Schulung	174
4.6	Kernbotschaft	176
5	Teilstudie 3: Entwicklung und Evaluation einer Peer-Schulung	177
5.1	Fragestellung	179
5.2	Untersuchungsdesign	180

5.2.1	Entwicklung des standardisierten Peer-Schulungsprogramms.....	181
5.2.2	Durchführung der Peer-Schulung	187
5.2.3	Formative Evaluation	188
5.2.4	Summative Evaluation	188
5.3	Ergebnisse	191
5.3.1	Aufbau des standardisierten Peer-Schulungsprogramms	191
5.3.2	Umsetzung des standardisierten Peer-Schulungsprogramms.....	193
5.3.3	Formative Evaluation	195
5.3.4	Beurteiler-Übereinstimmung	198
5.3.5	Beurteiler-Reliabilität.....	200
5.3.6	Übereinstimmung mit dem Peer und Experten-Konsens.....	201
5.3.7	Effektivität der standardisierten Schulung	202
5.4	Diskussion	204
5.4.1	Allgemeine Diskussion.....	204
5.4.2	Beantwortung der Fragestellung	211
5.5	Handlungsempfehlungen.....	213
5.6	Kernbotschaft	216
6	Zusammenfassende Diskussion.....	217
6.1	Einordnung in den Public Health Action Cycle	222
6.2	Implikationen	225
7	Literaturverzeichnis.....	230
	Anhangsverzeichnis	248
	Anhang 1.....	249
	Anhang 2.....	252
	Anhang 3.....	262
	Anhang 4.....	265
	Anhang 5.....	273
	Anhang 6.....	308
	Anhang 7.....	316
	Anhang 8.....	318
	Anhang 9.....	320
	Anhang 10	352
	Abbildungsverzeichnis.....	362
	Tabellenverzeichnis.....	364
	Abkürzungsverzeichnis.....	368
	Persönliche Versicherung des Bewerbers.....	371

Zusammenfassung

Theoretischer Hintergrund. Infolge des demographischen Wandels, der geprägt ist durch eine steigende Lebenserwartung und eine sinkende Geburtenziffer, wird sich die Bevölkerungsstruktur der Bundesrepublik Deutschland in den kommenden Jahrzehnten nachhaltig verändern. Derzeit ist noch nicht abzusehen, ob im Rahmen dieser Entwicklungen auch die Lebensarbeitszeit sukzessiv über der 67. Lebensjahr hinaus gesteigert wird. Mit steigender Lebenserwartung und längerer Lebensarbeitszeit treten jedoch degenerative Prozesse und vormals eher latent ausgeprägte chronische Erkrankungen bei der noch arbeitenden Bevölkerung in den Vordergrund. Versicherte der Deutschen Rentenversicherung (DRV) können dann als Folge gesundheitlicher Einschränkung eine Rehabilitationsmaßnahme oder eine Rente wegen Erwerbsminderung (EM-Renten) nach §§ 43 SGB VI beantragen. Im Rahmen einer Sozialmedizinischen Begutachtung durch einen gutachterlichen, neutralen Sachverständigen werden die aktuellen körperlichen und psychischen Befunde, die Beeinflussbarkeit der aktuellen Leistungsfähigkeit durch eine Rehabilitationsmaßnahme sowie Aussagen zur weiteren Prognose der Grunderkrankung verknüpft, um das Restleistungsvermögen quantitativ und qualitativ zu bewerten. Dieser äußerst komplexe und individuelle Beurteilungsprozess unterliegt hinsichtlich der Diagnosen und Leistungsbeurteilungen infolge verschiedener Störfaktoren einer gewissen Beurteilervarianz. Neben einer fehlenden wissenschaftlichen Grundlage der Entscheidungsfindung und -begründung, existiert auch kein trägerübergreifendes Qualitätssicherungskonzept für Sozialmedizinische Gutachten im EM-Rentenverfahren der DRV.

Fragestellungen. Das vorliegende Forschungsvorhaben griff die unklare Evidenzlage hinsichtlich des Sozialmedizinischen Begutachtungsprozesses und der fehlenden trägerübergreifenden Qualitätssicherung sozialmedizinischer Gutachten auf und hatte zum Ziel, wissenschaftlich fundierte Literatur zur Evidenzbasierung zu identifizieren sowie ein trägerübergreifendes Qualitätssicherungskonzept für die Sozialmedizinische Beurteilung von Anträgen zur EM-Rente auf Basis eines Peer Review-Verfahrens zu evaluieren.

Methodik. Zur Beantwortung der Fragestellung wurde drei voneinander unabhängige Studiendesigns konzipiert, die sich am Public Health Action Cycle mit seinen vier konsekutiven Bearbeitungsschritten (Problemdefinition, Strategieformulierung, Umsetzung, Evaluation) orientierten. Für Teilstudie 1 wurde eine systematische Literaturrecherche in sechs elektronischen Datenbanken zur Identifikation und Beurteilung von praktischen Instrumenten und Verfahren zur Erhöhung der Validität der Sozialmedizinischen Begutachtung geplant. Nach Festlegung der Suchstrategie (Publikationen der letzten 20 Jahre, Freitextsuche ergänzt durch Medical Subject Headings in deutscher und englischer

Sprache) wurden Schlagworte und deren Synonyme ausgearbeitet und die Recherche mittels des PICO-Schemas über OVID, Web of Science, CDSR und CENTRAL durchgeführt. In Teilstudie 2 wurde ein indikations- und trägerübergreifendes Peer Review-Verfahren als Instrument der Qualitätssicherung für die Sozialmedizinische Begutachtung evaluiert. Dazu wurde die Interrater-Reliabilität eines dazugehörigen Prüffragenkatalogs zum Review-Verfahrens überprüft. Im Rahmen der Untersuchung wurden insgesamt 260 anonymisierte sozialmedizinische Gutachten anhand von 23 Prüffragen von insgesamt 19 Peers bewertet, die im Vorfeld der Studie in einem Schulungsworkshop mit den Prüffragen vertraut gemacht wurden. Dabei wurden 240 Gutachten von je zwei Peers und 20 Gutachten von allen am Verfahren teilnehmenden Peers beurteilt. Die konsekutive Teilstudie 3 umfasste die Entwicklung und Evaluation eines strukturierten Peer-Schulungsprogramms für das trägerübergreifende Qualitätssicherungsverfahren. Nach der inhaltlichen und didaktischen Ausarbeitung der Schulung mittels der adaptierten Bloom'schen Taxonomie der kognitiven Lernziele wurde die Schulung formativ, mittels Fragebögen, und summativ, mittels der Bestimmung der Interrater-Reliabilität basierend auf drei im Abschluss der Schulung zu beurteilenden sozialmedizinischen Gutachten, evaluiert. Insgesamt wurden 40 Sozialmediziner aus allen Rentenversicherungsträgern geschult. Dreizehn der 40 teilnehmenden Sozialmediziner waren bereits am vorangegangenen Projekt zur Analyse der Interrater-Reliabilität beteiligt. Das Hauptbewertungskriterium der Teilstudien 2 und 3 war jeweils die Interrater-Reliabilität des sog. übergeordneten Kriteriums, welches die Nachvollziehbarkeit der gutachterlichen Argumentationskette innerhalb eines Gutachtens überprüft.

Teilstudie 1: Systematische Literaturrecherche. Durch die Datenbankrecherche wurden N = 17.079 Publikationen identifiziert. Nach Streichen von Duplikaten (n = 5.059) und dem Ausschluss von n = 11.916 Referenzen anhand von Titel und Abstract wurden n = 104 Referenzen zur Volltextsichtung ausgewählt, von denen wiederum n = 88 ausgeschlossen wurden, da sie die Selektionskriterien nicht erfüllten. Die verbliebenen 16 Studien untersuchten „allgemeine und indikationsspezifische Hinweise zur Begutachtung sowie Assessmentinstrumente“, „ICF und ICF Core Sets“ sowie „Besonderheiten der Begutachtung aus Perspektive der gutachterlich tätigen Ärzte“. Es wurden primär narrative Reviews und Beobachtungsstudien identifiziert (n = 10), welche die niedrigste Evidenzklasse aufweisen. Insgesamt erscheint die Evidenz stark limitiert, qualitativ hochwertige Studiendesigns (RCT, quasi-experimentelle Studien etc.) sind in der Sozialmedizin sehr selten zu identifizieren. Die gefundenen narrativen Reviews beschrieben zwar zentrale zu begutachtende Bestandteile bei der Begutachtung, jedoch war die Evidenz und damit die Verallgemeinerbarkeit dieser Publikationen begrenzt.

Teilstudie 2: Interrater-Reliabilität des Peer Review-Verfahrens. Die Interrater-Reliabilität des übergeordneten Kriteriums, basierend auf 325 Reviews, schwankte je nach herangezogenen Reliabilitätskoeffizienten (Cohens Kappa, Kendalls Tau, Spearmans rho, Kendalls W) zwischen 0,2 und 0,39. Insgesamt konnte eine sehr hohe Varianz bei der Beurteilung des übergeordneten Kriteriums festgestellt werden. Die Range der Peer-Übereinstimmung lag zwischen 87,5 % und 29,2 %. Als valider Evaluationskoeffizient wurde Kendalls W ($r_w = 0,39$) definiert. Die Interrater-Reliabilität des übergeordneten Kriteriums erreichte die im Allgemeinen als befriedigend angesehene Höhe von 0,7 nicht. Im direkten Vergleich mit nationalen und internationalen Studien zeigte die Reliabilität des Peers Reviews eine annähernd ähnliche Ausprägung. Dennoch überschritt die Interrater-Reliabilität von Kendalls W die kritische Orientierungsmarke von 0,4 nicht.

Teilstudie 3: Entwicklung und Evaluation einer Peer-Schulung. Der routinemäßige Einsatz einer standardisierten Peer-Schulung sollte die Reliabilität des Verfahrens weiter erhöhen und die Sozialmediziner auf die Rolle als Peer vorbereiten. Das ausgearbeitete zweieinhalbtägige Peer-Schulungsprogramm setzte sich aus insgesamt neun Lernzielen zusammen. Der zweite Teil der Schulung wurde in Kleingruppenarbeit von maximal zwölf Personen realisiert. Während der Schulung wurden Wissen, Kompetenzen und Fertigkeiten im Umgang mit dem Review-Verfahren gelehrt. Die formative Evaluation zeigte, dass sich die Peers im Anschluss der Schulung als relativ sicher im Umgang mit dem Prüffragenkatalog fühlten ($70 \% < x < 84 \%$). In der allgemeinen Bewertung der Schulung wurde diese als gelungen und empfehlenswert eingeschätzt. Für die standardisierte Peer-Schulung ergab sich eine Prozentübereinstimmung aller möglichen Rater-Paare in Höhe von 61 %. Dieser Wert unterschied sich zwar nominal, aber nicht signifikant von der erwarteten zufälligen Übereinstimmungsrate von 36 % ($\kappa_m = 0,39$, $p > 0,05$). Die Interrater-Reliabilität von Fleiss Kappa (κ_m) verfehlte erneut die kritische Orientierungsmarke von 0,4.

Schlussfolgerung. Die Verbesserung der Qualität sozialmedizinischer Gutachten und die Entwicklung eines Qualitätssicherungssystems für die Sozialmedizinische Begutachtung wurde seitens der DRV Bund kontinuierlich seit dem Jahr 2004 vorangetrieben. Zusammenfassend lässt sich feststellen, dass die Ergebnisse des vorliegenden Promotionsvorhabens wichtige Beiträge zur Optimierung der Gutachtenqualität beigetragen haben. Langfristig wird die systematische Bewertung der Gutachtenqualität durch Anwendung einheitlicher Qualitätskriterien sowie die regelmäßige Schulung der Peers die Gesamtqualität des Sozialmedizinischen Begutachtungsprozesses positiv beeinflussen. Es ist davon auszugehen, dass auch die Reliabilität mit Voranschreiten des Verfahrens durch die kontinuierliche Übung der Peers ansteigen wird. Auf der Suche nach höherer Evidenz sollte zukünftig ein Vorgehen gewählt werden, bei dem nur einzelne Begutachtungsschritte (z.B. Befunderhebung) mittels einer Literaturrecherche untersucht werden.

1 Theoretischer Teil

1.1 Problemstellung und Überblick

Unter Berücksichtigung des demographischen Wandels, der gekennzeichnet ist durch eine sukzessiv steigende Lebenserwartung und eine tendenziell sinkende Geburtenhäufigkeit, wird sich der Altersaufbau der Bevölkerung innerhalb der Bundesrepublik Deutschland grundlegend ändern (DESTATIS, 2015). Hochrechnungen gehen davon aus, dass im Jahr 2050 voraussichtlich mehr als 33 % der Bundesbürger 65 Jahre und älter sein werden (Peters, Pritzkeleit, Beske & Katalinic, 2010). Diese Entwicklung wird maßgeblich durch die Alterung der geburtenstarken Jahrgänge der 1950er- und 1960er-Jahre, den sogenannten Baby-Boomern, disponiert. Daraus resultiert, dass die Zahl der jüngeren Menschen im erwerbsfähigen Alter abnimmt und ältere Arbeitnehmer vermehrt aus dem Erwerbsleben ausscheiden oder kurz vor dem Ruhestand stehen (Gellert, Kesselmann & Wilke, 2018). Zur Aufrechterhaltung der Finanzierung der Sozialversicherungssysteme könnte zukünftig, als Folge dieser Entwicklung, die Lebensarbeitszeit auch über das 67. Lebensjahr hinaus verlängert werden (Bomsdorf, 2015; Buslei, 2017). Mit steigender Lebenserwartung und längerer Lebensarbeitszeit treten degenerative Prozesse und vormals eher latent ausgeprägte chronische Erkrankungen auch bei der noch arbeitenden Bevölkerung in den Vordergrund (Maaz, Winter & Kuhlmeier, 2007, S. 16). In Kombination mit einer Zunahme der subjektiven Arbeitsbelastung steigt daher im Alter auch die Zahl der Arbeitsunfähigkeitstage signifikant an (Gellert et al., 2018). Dies könnte insgesamt zu einem steigenden Bedarf an Leistungen zur Teilhabe sowie einer steigenden Anzahl an Antragstellungen zur Erwerbsminderungsrente führen (Bäcker, Bispinck & Hofemann, 2013, S.54; Dannenberg, Hofmann, Kaldybajewa & Kruse, 2010; Deck & Pohontsch, 2011; Nowossadeck, 2018). Bereits heute liegt das durchschnittliche Renteneintrittsalter mit 61,8 Jahren unter der allgemeinen Altersrentengrenze (Stand: 2016). Vor allem Renten infolge verminderter Erwerbsfähigkeit (EM-Rente) üben mit einem mittleren Renteneintrittsalter von 51,7 Jahren einen signifikanten Einfluss auf das durchschnittliche Rentenzugangsalter aus (Clemens, 2014; DRV, 2017a, S. 113). Es ist jedoch davon auszugehen, dass sich dieser Einfluss nach dem Jahr 2040 reduziert, da die Baby-Boomer den Arbeitsmarkt zu diesem Zeitpunkt fast vollständig verlassen haben werden (Gellert et al., 2018).

Infolge gesundheitlicher Einschränkungen haben Versicherte der Deutschen Rentenversicherung (DRV) unter bestimmten versicherungsrechtlichen und medizinischen Voraussetzungen (s. Kap. 1.2.2) einen Anspruch auf Leistungen zur Teilhabe oder EM-Rente (Götz, Roth & Chojetzki, 2011). Zur Absicherung des Risikos der Minderung der Erwerbsfähigkeit bedarf es einer sozialmedizinischen Sachaufklärung, die im Rahmen einer Sozialmedizinischen Begutachtung auf Grundlage der vorhandenen Aktenlage oder mit

persönlicher Untersuchung des Versicherten durchgeführt wird (Cibis, Paulus, Mai & Gehrke, 2011). Dabei können begutachtende Ärzte je nach Aufbau des einzelnen Rentenversicherungsträgers und benötigter Fachrichtung hauptamtlich interne oder auch externe Vertragsgutachter sein (Legner & Cibis, 2007). Das sozialmedizinische Gutachten wird durch den Sozialmedizinischen Dienst der DRV erstellt und dient als Grundlage für die Entscheidung über eine beantragte Sozialleistung im Rentenverfahren. Folglich ist bei einer Zahl von mehr als 800 hauptamtlich bei der DRV gutachterlich tätigen Sozialmedizinern und zusätzlich bundesweit über 1.000 auf Honorarbasis arbeitenden Gutachtern eine trägerübergreifende systematische Qualitätssicherung erforderlich, die bis dato noch nicht etabliert war (Ueberschär, 2008).

Die Entwicklung einer externen Qualitätssicherung für die Sozialmedizinische Begutachtung war bereits Teil der Handlungsempfehlungen der Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung (SOMEKO) aus dem Jahr 2004 und sollte sowohl die individuelle Begutachtungsqualität für den einzelnen Versicherten verbessern als auch die Vergleichbarkeit zwischen den Sozialmedizinischen Diensten der DRV-Träger gewährleisten (VDR, 2004). Da der sozialmedizinische Begutachtungsprozess im Ganzen als ein komplexes und individuelles Geschehen zwischen Gutachter und Versichertem gestaltet ist, wurden ärztliche Gutachten, deren formale und inhaltliche Qualität zentrale Merkmale der Ergebnisqualität sind, für die Qualitätssicherung herangezogen (Gehrke, 2008). Im Folgenden wurde daher für die trägerübergreifende Qualitätsprüfung sozialmedizinischer Gutachten ein Peer Review-Verfahren zugrunde gelegt. Das Peer Review ist ein weit verbreitetes Verfahren der externen Qualitätssicherung im Gesundheitswesen (Shaw, 2001), das in der DRV bei der Qualitätssicherung der medizinischen und psychosomatischen Rehabilitation bereits Anwendung findet (Farin et al., 2003; Glattacker & Jäckel, 2007). Im Allgemeinen beurteilt im Review-Prozess ein externer, erfahrener Peer, also ein gleichgestellter Fachkollege der jeweiligen Berufsgruppe, die Qualität von Behandlungsprozessen und -ergebnissen auf Basis definierter Bewertungskriterien. In der Qualitätssicherung der Sozialmedizinischen Begutachtung werden anonymisierte Gutachten anhand von Qualitätskriterien, die verschiedene qualitative Aspekte innerhalb eines Gutachtens widerspiegeln, durch erfahrene Sozialmediziner aus den Häusern der Rentenversicherungsträger bewertet (Gehrke, 2010; Strahl et al., 2016). Der Einsatz eines Peer Reviews als Bewertungsmethode erfordert im Vorfeld die Überprüfung der Interrater-Reliabilität (Beobachterübereinstimmung) als wichtigstes Gütekriterium (Goldman, 1994) sowie die regelmäßige Schulung der teilnehmenden Sozialmediziner (Strahl et al., 2018).

Neben der Empfehlung zur Entwicklung einer trägerübergreifenden Qualitätssicherung hat die SOMEKO-Kommission auch die politische Debatte für mehr Qualität im

Gesundheitswesen im Zusammenhang mit der Sozialmedizin und speziell mit der Sozialmedizinischen Begutachtung erörtert. Dadurch wurde die Forderung, effektiv, wirtschaftlich und evidenzbasiert zu arbeiten, an die Sozialmedizin herangetragen. Folge des Diskurses war die Forderung nach der wissenschaftlichen Einschätzung der Sozialmedizinischen Begutachtung einschließlich der dabei verwendeten diagnostischen Instrumente. Daraus ging die Entwicklung von Leitlinien für die Steuerung von Behandlungs- und Begutachtungsprozessen im Bereich der Rehabilitation und der EM-Renten hervor (VDR, 2004). Diese Begutachtungsleitlinien werden von der DRV sukzessiv entwickelt und allen Sozialmedizinern zur Verfügung gestellt. Abseits der entwickelten Leitlinien existiert bislang jedoch keine umfassende Literatur- und Datenbasis zur Evidenzbasierung auf dem Gebiet der Sozialmedizinischen Begutachtung. Im Sinne der Gleichbehandlung aller Versicherten sowie der außerordentlichen sozioökonomischen Relevanz der Thematik bedarf es im Bereich der Sozialmedizinischen Begutachtung eines Mindestmaßes an Standardisierung und Transparenz.

Die vorliegende Dissertation beschäftigt sich mit der Evidenzbasierung der Sozialmedizin und der Sozialmedizinischen Begutachtung. Die Daten für die Dissertationsarbeit stammen aus den zwei DRV Bund geförderten Drittmittelprojekten „Evaluation und Aufwandschätzung zum Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung (EvaBegut)“ und „Entwicklung eines Peer-Schulungsprogramms für das trägerübergreifende Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung (SchuP)“. Ein weiteres Drittmittelprojekt der DRV Bund „Systematische Literaturrecherche zur Sozialmedizinischen Begutachtung bei EM-Renten und Reha-Zugang (LitBegut)“ dient lediglich als Grundlage für eine neue Recherche in der hier vorliegenden Arbeit.

Im Zentrum der Dissertationsarbeit stehen die Durchführung einer **systematischen Literaturrecherche**, eine **Interrater-Reliabilitätsanalyse** sowie die Entwicklung und **Evaluation einer Peer-Schulung**. Aus der Literatur ist bekannt, dass aus individuellen Unterschieden zwischen den Peers eine niedrige Reliabilität resultieren kann. Daraus abgeleitet wird zur Kontrolle dieses Bias ein regelmäßiges Training für Anwender von Peer Review empfohlen (Harris & Bratzler, 2013; Smith, Atherly, Kane & Pacala, 1997; Tuijn, Janssens, Robben & van den Bergh, 2012). Die Untersuchung erfolgt primär auf Basis quantitativer Forschungsmethoden (Bortz & Lienert, 2008; Bortz, Lienert & Boehnke, 2008) sowie der Analyse von Freitexten. Im theoretischen Teil der Arbeit, dem **Kapitel 1**, wird eine umfassende Zusammenfassung über die wichtigsten Hintergründe zur DRV, EM-Rente, Sozialmedizinischen Begutachtung und Qualitätssicherung gegeben. Im sich anschließenden **Kapitel 2** wird ein Überblick über das Studiendesign und die in der Dissertationsarbeit genutzten methodischen Verfahren dargelegt. Unter **Kapitel 3 bis 5** werden die drei Teilstudien zur Beantwortung der Hauptfragestellung inklusive einer

studienspezifischen Diskussion beschrieben. Im abschließenden **Kapitel 6** erfolgt eine Zusammenfassung, einschließlich einer allgemeinen Diskussion, Handlungsempfehlungen für die Qualitätssicherung der Sozialmedizinischen Begutachtung und Implikationen für die Praxis und Forschung. Teilbereiche dieser Arbeit wurden bereits in wissenschaftlichen Zeitschriften publiziert:

- Strahl, A., Gerlich, C., Alpers, G.W., Ehrmann, K., Müller-Garnn, A., Gehrke, J. & Vogel, H. (2018). Development and evaluation of a standardized peer-training in the context of peer review for quality assurance in work capacity evaluation. *BMC Medical Education*, 18, 135.
- Strahl, A., Gerlich, C., Wolf, H.-D., Gehrke, J., Müller-Garnn, A. & Vogel, H. (2016). Qualitätssicherung in der Sozialmedizinischen Begutachtung durch Peer Review – ein Pilotprojekt der Deutschen Rentenversicherung. *Gesundheitswesen*, 78, 156-160.

Entsprechend der Promotionsordnung der Europa-Universität Flensburg wird mit diesen Publikationen völlig transparent umgegangen. Es erfolgte eine intensive redaktionelle Überarbeitung und inhaltliche Fortentwicklung. Entsprechend ist diese Dissertationsschrift keine zusammenfassende Replikation der Einzelpublikationen, sondern ein eigenständiger wissenschaftlicher Beitrag.

1.2 Die gesetzliche Rentenversicherung

Die weitestgehende Verankerung des Versicherungsprinzips im Art. 20 Abs. 1 des Grundgesetzes fixiert bis heute den Grundgedanken des Sozialstaates in der deutschen Verfassung. Dabei stellt die gesetzliche Rentenversicherung einen zentralen Versicherungszweig im sozialen Sicherungssystem nach Art. 20 des Deutschen Grundgesetzes dar (Horn & Schuchardt, 2015, S. 7). Ihre Aufgabe dient primär der Alterssicherung und bildet zusammen mit der betrieblichen und der privaten Altersvorsorge das klassische 3-Säulen-System der Alterssicherung. Dabei werden in Deutschland etwa 2/3 aller Ausgaben der Alterssicherung von der gesetzlichen Rentenversicherung übernommen (DRV, 2018a, S. 38). Die Rentenversicherung als Teil des Sozialversicherungssystems operiert auf Basis von fünf verbindlichen Grundprinzipien, um die soziale Absicherung für die Bürger der Bundesrepublik Deutschland zu gewährleisten (s. Tab. 1).

Tabelle 1: Grundprinzipien der Sozialversicherung (modifiziert nach Horn & Schuchardt, 2015, S. 5).

Prinzip der Versicherungspflicht	Es besteht eine Versicherungspflicht für alle Bürger mit Ausnahme von Beamten, viele Selbstständigen und Angehörigen verkammerter Berufe.
Prinzip der Beitragsfinanzierung	Das Budget der Sozialversicherung wird aus den Beiträgen der Bruttogehälter der Arbeitnehmer finanziert.
Prinzip der Solidarität	Die Sozialversicherung wird solidarisch von den Arbeitnehmerbeiträgen getragen, d.h. jeder Versicherte ist nicht primär für sich selbst, sondern gleichzeitig für alle anderen Mitglieder verantwortlich. Ausnahmen bilden die Unfallversicherung, knappschaftliche Rentenversicherungen, individuelle Gesundheitsleistungen und Riester-Renten.
Prinzip der Äquivalenz	Die Höhe von Geld- und Einkommensersatzleistungen berechnet sich aus der Höhe des zugrunde gelegten beitragspflichtigen Einkommens.
Prinzip der Selbstverwaltung	Die Sozialversicherungsträger sind organisatorisch und finanziell selbstständig.

Der § 1 des SGB VI beschreibt im Rahmen des Prinzips der Versicherungspflicht den Versichertenkreis der gesetzlichen Rentenversicherung. Pflichtversichert sind Arbeitnehmer, bestimmte Gruppen von Selbstständigen (z. B. Hebammen, Lehrer/Erzieher ohne versicherungspflichtige Tätigkeit, Seelotsen etc.) und sonstige Personenkreise (z. B. Wehrdienst, Bezug von Kranken-, Verletzten-, Versorgungskranken-, Übergang- oder Arbeitslosengeld).

Neben der kontinuierlichen Aufrechterhaltung der sozialen Sicherung bergen einige der o.g. Grundprinzipien auch gewisse Risikopotenziale. So wird das Prinzip der umlagefinanzierten Solidarität infolge des demografischen Wandels in seiner jetzigen Form voraussichtlich nicht aufrecht zu erhalten sein, da sich das Verhältnis von Beitragszahlern zu Leistungsempfängern prospektiv stärker in Richtung der Leistungsempfänger verändern wird (Gräßler, 2012, S. 50; Vogt, 2017, S. 167). Nach dem Prinzip der Beitragsfinanzierung müsste der Versicherungsbeitrag der Rentenversicherung im Jahr 2055 von aktuell 18,6 % auf etwa 28 % steigen, um eine gesicherte Altersrente für alle Mitglieder gewährleisten zu können (Vogt, 2017, S. 165). Da neben der Gewährung einer Basisrente eine zusätzliche Aufgabe der Rentenversicherung in der Absicherung von Lebensrisiken ihrer Mitglieder besteht, werden auch Leistungen zur Teilhabe nach § 4 SGB IX aus den Versicherungsbeiträgen finanziert. Entsprechend belegt und betreibt die Rentenversicherung eigene medizinisch-berufliche Versorgungseinrichtungen, um mögliche Renten wegen Erwerbsminderung oder Invalidität zu vermeiden (Welti, 2002).

Die demographische Alterung der Bevölkerung wirkt sich auch auf sog. Leistungen zur Teilhabe (s. Kap. 1.2.1) aus. Dieser Umstand macht sich durch ansteigende Fallzahlen beantragter Leistungen, insbesondere der medizinischen Rehabilitation bemerkbar. Das aus der Geschichte der Bundesrepublik Deutschlands entstandene gegliederte Sozialsystem umfasst verschiedene Sozialträger, die nach dem Verfahren der einheitlichen Risikoordnung Sozialleistungen für Versicherte übernehmen (Götz et al., 2011, S. 4). Danach wird dem Sozialträger diejenige Leistung zugeordnet, der das finanzielle Risiko ihres Scheiterns trägt. Entsprechend konzentrieren sich die Sozialleistungen der Rentenversicherung auf Menschen, die im Erwerbsleben stehen und das Ziel haben, die Erwerbsfähigkeit zu erhalten oder wiederherzustellen (Korsukéwitz & Irle, 2010). Daraus ableitend belegen aktuelle Entwicklungsanalysen, dass sich die Fallzahlen für medizinische Rehabilitationsleistungen bis zum Jahr 2040 zu Lasten der Rentenversicherung um 5,2 % verringern und die der gesetzlichen Krankenversicherung, als Rehabilitationsträger für Menschen älter als 65 Jahre, um 25,9 % steigen werden (Nowossadeck, 2018). Der Rückgang der Fallzahlen wird zwar keine Mehrkosten verursachen, dennoch wird das Reha-Budget aufgrund der „Schieflage“ innerhalb der Versichertenstruktur angepasst werden müssen.

An dieser Stelle sei darauf hingewiesen, dass auch gegenteilige Analysen vorliegen, in denen unter Berücksichtigung einer dynamischen Wirtschaftsentwicklung und einer jährlichen Erhöhung des Einkommens davon ausgegangen wird, dass die Beitragssteigerung durch ein höheres Realeinkommen kompensiert wird. Entsprechend würde somit ein Arbeitnehmer finanziell nicht in dem hohen Maße belastet werden wie in den Prognosen angenommen (Niemeier, 2018).

Struktur der gesetzlichen Rentenversicherung. Mit dem Inkrafttreten der Organisationsreform im Jahr 2005 nach dem „Gesetz zur Organisationsreform in der gesetzlichen Rentenversicherung“ (RVOrgG) vom 9.12.2004 wurde die aktuelle Struktur der heutigen DRV geschaffen. Nach dem Prinzip der Selbstverwaltung wurde die Umsetzung der Organisationsreform zur Erreichung der konkreten Reformziele (s. Tab. 2) durch die Rentenversicherungsträger eigenständig umgesetzt (Göbel, 2015).

Tabelle 2: Reformziele des Gesetzes zur Organisationsreform in der gesetzlichen Rentenversicherung (RVOrgG; modifiziert nach Göbel, 2015, S. 84).

Hauptziel: Verbesserung und Modernisierung der Verwaltungsstrukturen
Auftrennung der formalen Unterscheidung zwischen Arbeiter und Angestellten
Etablierung stabiler Rahmenbedingungen für alle Rentenversicherungsträger (RV-Träger)
Optimierung der internen Steuerung und Koordinierung
Reduktion des Verwaltungsaufwandes
Optimierung der internen Finanzstruktur
Reduktion der Anzahl an RV-Trägern in Deutschland
Weiterentwicklung interner Steuerungsinstrumente (Controlling und Benchmarking)

Im Zuge dieser Reform entstanden aus der Zusammenlegung einzelner RV-Träger insgesamt 14 Regionalträger. Ferner fusionierten die Sonderanstalten Bundesknappschaft, Bahnversicherung und Seekasse zur DRV Knappschaft-Bahn-See. Auf Bundesebene wurde die Bundesversicherungsanstalt für Angestellte und der VDR zur DRV Bund zusammengeführt (Klenk, 2005, S. 103). Die DRV Bund fungiert sowohl als überregionaler Träger als auch als Dachverband zur Bearbeitung von Grundsatz- und Querschnittsaufgaben.

1.2.1 Leistungen der gesetzlichen Rentenversicherung

Das Vorliegen einer Krankheit oder Behinderung hat Auswirkungen auf den Betroffenen selbst, seine Umwelt und die Gesellschaft. Vor allem chronische Erkrankungen sind die Hauptdeterminanten für Fehlzeiten und Arbeitsbeeinträchtigungen in Deutschland (Baase, 2007, S. 51; Mittag, Reese & Meffert, 2014; s. Kap. 1.3). Versicherte der DRV können als Folge einer gesundheitlichen Einschränkung unter bestimmten versicherungsrechtlichen und medizinischen Voraussetzungen verschiedene Sozialleistungen beantragen. Im Jahr 2016 wurden bei der DRV etwa 1,64 Millionen Anträge auf medizinische Rehabilitation, 450.000 Anträge auf Leistungen zur Teilhabe am Arbeitsleben und ca. 360.000 Erstanträge auf EM-Rente gestellt (DRV, 2017a, S. 7; 2018b, S. 9). Ferner kommen

Weitergewährungsanträge von Versicherten, die bereits eine befristete EM-Rente erhalten haben, als auch sozialmedizinische Beurteilungen im Rechtsmittelverfahren hinzu (Legner & Cibis, 2007).

Leistungen zur Teilhabe. Mit dem Inkrafttreten des SGB IX am 19. Juni 2001 hat der Gesetzgeber die rechtliche Grundlage der Leistungen zur Rehabilitation und Teilhabe für alle Rehabilitationsträger geschaffen (Welti, 2001). Mit dem SGB IX wurde auch der Begriff „Leistungen zur Teilhabe“ erstmals eingeführt. Diese lassen sich in Leistungen zur medizinischen Rehabilitation, Leistungen zur Teilhabe am Arbeitsleben, unterhaltssichernde und andere ergänzende Leistungen und Leistungen zur Teilhabe am Leben in der Gemeinschaft einteilen. Renten (Ausnahme Altersrente) und Leistungen zur Teilhabe sind grundsätzlich Antragsleistungen, die jedoch auch von Amts wegen erbracht werden können (Götz et al., 2011, S. 4). Grundsätzlich gilt, dass laut § 8 Abs. 2 SGB IX Leistungen zur Teilhabe der Gewährung einer EM-Rente vorzuziehen sind (Welti, 2002). Aktuelle Forschungsergebnisse in diesem Bereich weisen jedoch darauf hin, dass in vielen Fällen (ca. 40 % bis 50 %) über alle Diagnosegruppen hinweg eine EM-Rente ohne vorangegangenen Rehabilitationsleistung gewährt wird (Herget, Roski & Grande, 2017; Martin & Zollmann, 2013; Mittag et al., 2014).

Rentenarten. Das Leistungsspektrum der Rentenversicherung beinhaltet Renten wegen verminderter Erwerbsfähigkeit, Renten wegen Alters und Renten wegen Todes und bietet dadurch Absicherung vor den wirtschaftlichen Folgen einer Erwerbsminderung (Götz et al., 2011, S. 14). Die Abbildung 1 zeigt eine Übersicht über die unterschiedlichen Rentenarten.

Die beiden veralteten sozialrechtlichen Begrifflichkeiten „Rente wegen Berufsunfähigkeit“ und „Rente wegen Erwerbsunfähigkeit“ nach §§ 43 und 44 SGB VI (nach alten Rentenrecht bis 31.12.2000) gingen davon aus, dass bei Versicherten eine geminderte Leistungsfähigkeit vorliegen kann und eine allgemeine Erwerbstätigkeit erhalten geblieben ist. Entsprechend wurde eine Berufsunfähigkeitsrente dann gewährt, wenn eine mehr als 50%ige Minderung der Erwerbsfähigkeit im bisherigen Beruf vorlag. Im Gegensatz dazu erhielten Versicherte eine Rente wegen Erwerbsunfähigkeit, wenn diese „auf nicht absehbare Zeit außerstande waren, noch irgendeine Erwerbstätigkeit auf dem allgemeinen Arbeitsmarkt in gewisser Regelmäßigkeit auszuüben oder mehr als ein geringfügiges Arbeitsentgelt oder Arbeitseinkommen zu erzielen“ (Götz et al., 2011, S. 15).

Zum 01.01.2001 wurden die beiden Rentenarten vom Gesetzgeber durch eine teilweise bzw. volle EM-Rente ersetzt. Ein Bestandsschutz der Zahlungen wurde beibehalten, wenn Leistungsansprüche seitens des Versicherten vor dem 31.12.2000 gestellt wurden. Der Bezug einer Berufs- bzw. Erwerbsunfähigkeitsrente besteht bis zum Erreichen der Regelaltersgrenze (Götz et al., 2011, S. 15).

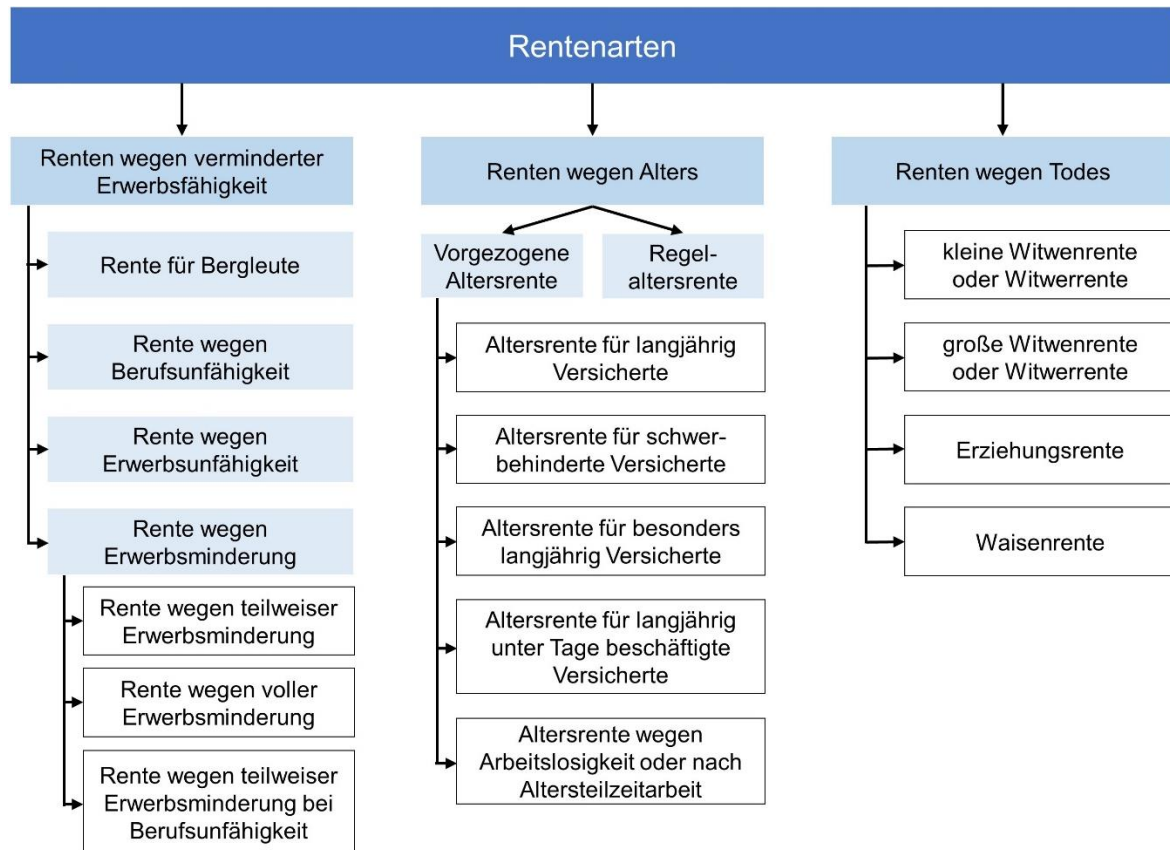


Abbildung 1: Überblick der Rentenarten (modifiziert nach Götz, Roth & Chojetzki, 2011, S. 14).

1.2.2 Rente wegen Erwerbsminderung

Der Anteil an EM-Renten (1.813.535 Fälle) gegenüber den Renten wegen Alters (18.130.131 Fälle) liegt im aktuellen Rentenbestand (Stand: 2016) bei 10,0 %. Signifikant höher liegt der Anteil der EM-Rente an allen Rentennewuzugängen. Im Jahr 2016 gab es in Deutschland insgesamt 766.821 Rentennewuzugänge (exklusive Renten wegen Todes), davon wurden 139.833 Renten wegen Erwerbsminderung gewährt. Damit liegt der Anteil der EM-Rente an allen Rentenzugängen bei 18,2 %. Frauen (71.170 Fälle) erhielten tendenziell häufiger einer EM-Rente als Männer (68.663 Fälle; DRV, 2017a, S. 48-53).

Nur etwa 40 % der EM-Rentennewuzugänge erfolgt aus zuvor sozialversicherungspflichtigen Beschäftigungsverhältnissen. Die Mehrheit der Versicherten, die eine EM-Rente beantragten, bezogen im Jahr vor dem Leistungsfall Arbeitslosengeld oder Grundsicherung (40 %), waren durch sonstige Versicherungsverhältnisse geschützt (10 %) oder erhielten gar keine (10 %) Leistungen (Mittag et al., 2014). Jährlich erheben etwa 360.000 Versicherte einen Anspruch auf EM-Rente, weil Sie durch Krankheit nicht mehr als sechs Stunden am Tag arbeiten können (DRV, 2018c, S. 5).

Die Gewährung dieser Rente richtet sich nach dem verbliebenen individuellen Restleistungsvermögen auf dem allgemeinen Arbeitsmarkt, nicht nach dem Grad der Leistungsminderung. Das in der Sozialmedizinischen Begutachtung festgestellte Restleistungsvermögen wird anschließend in ein dreistufiges quantitatives Stufenschema eingeordnet:

1. **Restleistungsvermögen von über 6 Stunden:** nicht erwerbsgemindert nach § 43 Abs. 3 SGB VI
2. **Restleistungsvermögen von 3 bis unter 6 Stunden:** teilweise erwerbsgemindert nach § 43 Abs. 1 SGB VI
3. **Restleistungsvermögen unter 3 Stunden:** voll erwerbsgemindert nach § 43 Abs. 2 SGB VI

Die Erwerbsminderung wird erst gewährt, wenn eine Verbesserung des Restleistungsvermögens auf nicht absehbare Zeit, d.h. mehr als sechs Monate, unter Ausschöpfung aller zur Verfügung stehenden therapeutischen Mittel, ausgeschlossen werden kann. Als Grenze für volle Erwerbsminderung wurde ein Leistungsvermögen von drei Stunden, äquivalent zur Geringfügigkeitsgrenze in der Arbeitslosenversicherung, gewählt, um Schnittstellenprobleme zwischen den beiden Trägern zu vermeiden (Götz et al., 2011, S. 16).

Bezugsvoraussetzung nach § 43 SGB VI. Allgemein gilt, dass eine EM-Rente nur dann bezogen werden kann, wenn bestimmte medizinische und versicherungsrechtliche Voraussetzungen erfüllt sind (Götz et al., 2011, S. 16; Schewe, 2017, S. 6).

Die medizinischen Voraussetzungen beschreiben das grundsätzliche Vorhandensein einer Krankheit oder Behinderung, welche das Leistungsvermögen auf dem allgemeinen Arbeitsmarkt auf unter sechs Stunden reduziert. Die Feststellung des Leistungsvermögens vollzieht sich nach Aktenlage, d.h. anhand vorliegender ärztlicher Unterlagen oder, bei unklarer Sachlage, durch die Beauftragung eines medizinischen Gutachtens im Rahmen der Sozialmedizinischen Begutachtung (DRV, 2018c, S. 6).

Neben den individuellen medizinischen Voraussetzungen werden zusätzlich versicherungsrechtliche Voraussetzungen seitens der DRV geprüft. Vor Eintritt der Erwerbsminderung muss der Versicherte mindesten fünf Jahre versichert gewesen sein. Innerhalb dieses Zeitraumes müssen vom Versicherten mindestens drei Jahre lang Pflichtbeiträge aus sozialversicherungspflichtigen Tätigkeiten eingezahlt worden sein (Götz et al., 2011, S. 16).

Eine EM-Rente wird nach § 102 Abs. 2 SGB VI grundsätzlich zunächst befristet für einen Zeitraum von maximal drei Jahre geleistet. Anschließend erfolgt eine erneute Prüfung durch den Sozialmedizinischen Dienst der RV-Träger.

Widerspruch und Klageverfahren. Ein EM-Rentenbescheid kann innerhalb eines Monats in einem formalen Widerspruchsverfahren neu geprüft werden. Überprüft wird die falsche Anwendung von Rechtsnormen, die ungenügende Ermittlung oder mangelhafte Bewertung von Tatsachen. Nach Bekanntgabe des Widerspruchsbescheides hat der Versicherte erneut die Möglichkeit in einem Sozialgerichtsverfahren die Entscheidung gerichtlich anzufechten. Zur Entlastung der Sozialgerichtsbarkeit ist eine Klage ohne ein vorangegangenes formales Widerspruchsverfahren nicht möglich (Götz et al., 2011, S. 22).

Aktuelle Entwicklungen und Trends. Im Bereich der EM-Renten zeigt sich in den letzten Jahren ein erheblicher Veränderungstrend bezüglich der zu Grunde gelegten medizinischen Indikationen als auch bezüglich des wirtschaftlichen Stellenwertes der Erwerbsminderung.

Die Häufigkeit der Diagnosen, die zu einer EM-Rente führen, zeigen einen kontinuierlichen Rückgang von Frühberentungen wegen Erkrankungen der Bewegungsorgane und des Herz-Kreislauf-Systems. Stattdessen ist ein gravierender Anstieg an EM-Renten aufgrund psychischer Störungen zu verzeichnen (Kaldyabajewa & Kruse, 2012; Korsukéwitz & Rehfeld, 2008). Psychische Störungen sind seit 2001 der häufigste Grund für die Bewilligung einer EM-Rente (DRV, 2014, S. 23) und machen aktuell 42,8 % aller bewilligten EM-Renten aus (Stand: 2016; DRV, 2017b). Abbildung 2 zeigt eine Übersicht über die Veränderung im Spektrum der Hauptdiagnosegruppen im Zeitverlauf.

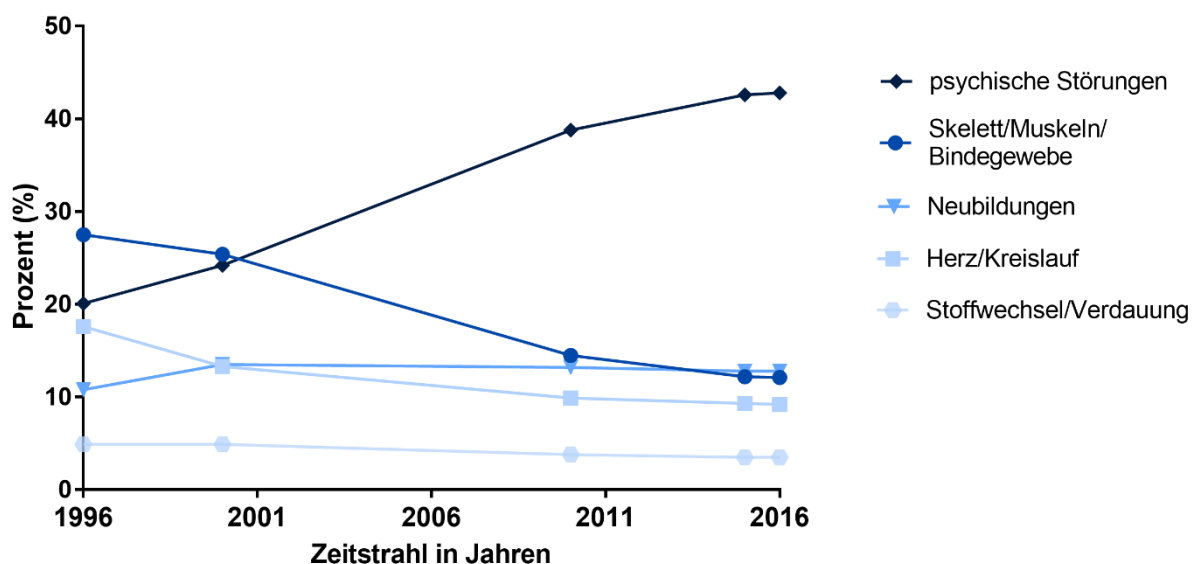


Abbildung 2: EM-Rentenzugänge nach ausgewählten Diagnosegruppen im Zeitverlauf (modifiziert nach DRV, 2017b, S. 2).

Es kann angenommen werden, dass der Wandel im Diagnosespektrum auf den zunehmenden medizinisch-technischen Fortschritt und einer gesellschaftlichen Verschiebung hin zum Dienstleistungssektor zurückzuführen ist. Vor allem die Wirksamkeit der Therapien und der Rehabilitation bei Herz-Kreislauferkrankungen führte voraussichtlich

zu dem sukzessiven Rückgang der EM-Rentanträge innerhalb dieser Diagnosegruppe (Korsukéwitz & Rehfeld, 2008). Durch die Verschiebung des Beschäftigungsschwerpunktes auf den Dienstleistungssektor sinken die körperlichen Belastungen am Arbeitsplatz. Dies geht einher mit sinkenden EM-Rentenfällen infolge von Erkrankungen des Skeletts, der Muskeln und des Bindegewebes. Gleichzeitig weisen Tätigkeiten im Dienstleistungsbereich jedoch verstärkt psychosoziale Belastungen durch z. B. hohen Termin- und Leistungsdruck, hohe Konzentrationsanforderungen oder qualifikatorische Überforderung auf. Vor allem ältere Beschäftigte sind diesen komplexen Arbeitsherausforderungen des vegetativen Regulationssystems verstärkt ausgesetzt (Frerichs, 2015; Hesse & Walden, 2014; Jansen & Müller, 2000). Der seit den letzten 20 Jahren mehr als 50%ige Anstieg psychischer Störungen, als Indikation einer EM-Rente, beruht insbesondere auf der Zunahme von Depressionen und Angststörungen (DRV, 2014, S. 24). Seit der medialen Destigmatisierung der sog. „Volkskrankheit Depression“ (Bramesfeld & Schwartz, 2007) sinkt die Scham vor psychischen Diagnosen auf Patientenseite und verbessert kontinuierlich die Wahrnehmung und Diagnostik seitens der behandelnden Ärzte. Ferner werden psychische Belastungen zunehmend auch als komorbide Störung auf der Grundlage einer somatischen Erkrankung diagnostiziert (Hesse & Walden, 2014). So erhöht beispielsweise das Vorliegen einer psychischen Komorbidität bei chronischen Rückenschmerzen das Risiko einer EM-Rente statistisch signifikant um das 1,4- bis 1,7-fache (Schmidt, Bernert & Spyra, 2014).

Darüber hinaus verändert sich seit mehreren Jahren auch der wirtschaftliche Stellenwert einer EM-Rente. Die durchschnittlichen EM-Rentenzahlungsbeträge sanken kontinuierlich und lagen bei voller Erwerbsminderung im Jahr 2016 bei 736 EUR pro Monat unter dem Grundsicherungsniveau. Der Grund für diesen Trend liegt in den veränderten Verhältnissen auf dem Arbeitsmarkt (Niedriglohnsektor, unregelmäßige Beschäftigung, Arbeitslosigkeit) begründet (Bäcker, 2013; DRV, 2017b). Ohne weitere private Vorsorgemaßnahmen droht bei schwerer Krankheit infolge des niedrigen EM-Rentenniveaus ein erhöhtes gesundheitsbedingte Armutsrisiko (Kemptner, 2014).

Die Gewährung einer EM-Rente ist von Seiten der DRV nicht als Dauerlösung für Versicherte konzipiert. Die anfängliche Befristung jeder bewilligten EM-Rente ist damit begründet, dass eine Rückkehr an den Arbeitsplatz durch geeignete Therapien und Rekonvaleszenz möglich sei. Allerdings wird nur ein geringer Anteil der Versicherten nach Erhalt einer EM-Rente auch wieder eingegliedert. Gründe hierfür könnten sein, dass bereits vor Gewährung der EM-Rente keiner versicherungspflichtigen Beschäftigung nachgegangen wurde, Versicherte bereits vor Antragstellung beruflich desintegriert waren oder sich der Gesundheitsstauts des Versicherten durch die progredient verlaufende chronische Grunderkrankung weiter verschlechtert hat (DRV, 2014, S. 25).

1.3 Chronische Erkrankungen

Bislang existiert in der Literatur keine einheitliche Definition für chronische Erkrankungen. Sie werden im Allgemeinen jedoch als „lang andauernde Krankheiten bezeichnet, die nicht vollständig geheilt werden können und eine andauernde oder wiederkehrend erhöhte Inanspruchnahme von Leistungen des Gesundheitssystems nach sich ziehen“ (RKI, 2011, S. 54). Dies führt für Betroffene zu zahlreichen Einschränkungen, wie permanente körperliche Beschwerden, Aktivitäts- und Funktionsstörungen sowie eine reduzierte berufliche Leistungsfähigkeit (Krämer & Bengel, 2016, S. 28; Vogel, Faller & Reusch, 2016, S. 350; Wolf-Kühn & Morfeld, 2016, S. 33).

In modernen Industrieländern wurden infektiöse und parasitäre Krankheiten als Ursache für Morbidität und Mortalität sukzessiv von chronischen Erkrankungen abgelöst. Diese stellen aktuell die größte Herausforderung für das öffentliche Gesundheitssystem dar. Zu den relevantesten Erkrankungen zählen koronare Herzkrankheiten, Krebs, Schlaganfall, Diabetes mellitus, muskuloskelettale Erkrankungen, chronische Lungenerkrankungen (z. B. COPD) sowie psychische Erkrankungen (Harris, 2010, S. 1). In Deutschland verursachen koronare Herzkrankheiten und Krebserkrankungen etwa zwei Drittel aller Todesfälle (Prütz et al., 2014, S. 113).

Per Definition führen chronische Krankheiten nicht unmittelbar zum Tode, verursachen jedoch über mehrere Jahre hinweg direkte und indirekte Krankheitskosten (Prütz et al., 2014, S. 113). Damit haben chronischen Erkrankungen eine erhebliche Relevanz als medizinische Indikation einer EM-Rente (Hagen, Himmelreicher, Kemptner & Lampert, 2011). In Anlehnung an die bisher entwickelten Rentenversicherungsleitlinien zur Sozialmedizinischen Begutachtung der Leistungsfähigkeit im Erwerbsleben werden im Folgenden die wichtigsten Erkrankungen kurz erörtert (vgl. DRV, 2018d). Hierbei wird neben den chronisch somatischen Erkrankungen ein besonderer Fokus auf die psychischen und chronisch psychischen Störungen als häufigste Diagnosekategorie für die Beantragung einer EM-Rente gelegt.

1.3.1 Häufige Diagnosen im Zusammenhang mit EM-Rente

Psychische Störungen. Psychische Störungen sind bereits seit einigen Jahren der häufigste Grund zur Beantragung einer EM-Rente. Mit etwa 45 % machen affektive Störungen gefolgt von neurotischen, Belastungs- und somatoformen Störungen (24 %) sowie Schizophrenie, schizotypische und wahnhaftige Störungen (14 %) die häufigsten EM-Rente begründenden Diagnosen aus (DRV, 2012). Während die Häufigkeit affektiver Störungen kontinuierlich zunimmt und somatoforme Störungen bzw. Suchterkrankungen

nur gering ansteigen, bleibt die Anzahl der Fälle von EM-Rente wegen schizophrener, schizotyper und wahnhafter Störungen stabil. Mit im Mittel 48,1 Jahre sind Versicherten mit psychischen Störungen bei Rentenbeginn jünger als EM-Rentenbezieher anderer Diagnosegruppen (DRV, 2014, S. 24).

Die 12-Monats-Prävalenz psychischer Störungen von Erwachsenen im Alter zwischen 18 und 79 Jahren liegt laut der repräsentativen Studie zur Gesundheit Erwachsener in Deutschland (DEGS1) aus den Jahren 2008 bis 2011 bei insgesamt 27,8 % (Jacobi et al., 2016). Frauen sind häufiger (33,3%) betroffen als Männer (22,0 %) und leiden häufiger (49,8 %) unter mehr als einer psychischen Diagnose. Mit 15,3 % stellen Angststörungen, als subsummierte Kategorie einzelner Störungen, die häufigsten psychischen Erkrankungen dar (Tab. 3). Anschließend folgen unipolaren Depressionen (7,7 %) und Störungen durch Alkohol- oder Medikamentenkonsum (5,7 %) (Jacobi et al., 2014).

Tabelle 3: 12-Monats-Prävalenzen psychischer Störungen in der erwachsenen Allgemeinbevölkerung (modifiziert nach Jacobi et al., 2014, S. 80).

Diagnose	Frauen	Männer	Gesamt
	Prozent (95 %-Konfidenzintervall)		
Spezifische Phobien	15,4 (13,8-17,2)	5,1 (4,2-6,2)	10,3 (9,3-11,3)
Unipolare Depression	10,6 (9,2-12,2)	4,8 (4,0-5,7)	7,7 (6,9-8,6)
Major Depression	8,4 (7,2-9,9)	3,4 (2,8-4,3)	6,0 (5,2-6,8)
Agoraphobie	5,6 (4,6-6,8)	2,3 (1,7-3,1)	4,0 (3,4-4,7)
Zwangsstörung	4,0 (3,1-5,1)	3,3 (2,6-4,3)	3,6 (3,1-4,4)
Alkoholabhängigkeit	1,6 (1,1-2,2)	4,4 (3,5-5,5)	3,0 (2,5-3,6)
Soziale Phobie	3,6 (2,7-4,8)	1,9 (1,4-2,5)	2,7 (2,2-3,4)
Posttraumatische Belastungsstörung	3,6 (2,8-4,6)	0,9 (0,6-1,5)	2,3 (1,8-2,8)
Generalisierte Angststörung	2,9 (2,2-4,0)	1,5 (1,1-2,2)	2,2 (1,8-2,8)
Panikstörung	2,8 (2,2-3,6)	1,2 (0,8-1,8)	2,0 (1,6-2,5)
Alkoholmissbrauch	0,4 (0,2-0,9)	3,1 (2,4-4,1)	1,8 (1,4-2,3)
Medikamentenmissbrauch	1,7 (1,1-2,4)	1,5 (1,0-2,2)	1,6 (1,2-2,0)
Bipolare Störung	1,7 (1,2-2,5)	1,3 (0,8-2,0)	1,5 (1,1-2,0)
Anorexia nervosa	1,1 (0,7-1,8)	0,3 (0,2-0,8)	0,7 (0,5-1,1)
Medikamentenabhängigkeit	0,7 (0,4-1,2)	0,3 (0,1-0,6)	0,5 (0,3-0,8)
Bulimia Nervosa	0,3 (0,2-0,5)	0,1 (0,0-0,3)	0,2 (0,1-0,3)
Binge-Eating-Störung	0,1 (0,0-0,4)	0,1 (0,0-0,2)	0,1 (0,0-0,2)

Die Lebenszeitprävalenz psychischer Störungen ist höher als die 12-Monatsprävalenz und liegt studienübergreifend bei etwa 50 % (Kessler et al., 2005; Meyer, Rumpf, Hapke, Dilling & John, 2000; Schmidt et al., 2013). Trotz dieser hohen Prävalenzraten in der

Allgemeinbevölkerung sind psychische Störungen aber per se keine chronischen Krankheiten, die eine EM-Rente nach sich ziehen. Erst wenn eine Störung chronifiziert vorliegt, haben Versicherte Probleme, ihr Leistungsvermögen auf dem allgemeinen Arbeitsmarkt zu erhalten oder zu verbessern. Die DEGS1-Studie konnte aufzeigen, dass trotz einer vorhandenen psychischen Diagnose weniger als die Hälfte der Betroffenen in psychotherapeutischer oder psychiatrischer Behandlung waren (Jacobi et al., 2014).

Überwiegend werden psychische Störungen ambulant und teilstationär versorgt. Bleibt eine Behandlung aus, können sich psychische Störungen dauerhaft manifestieren, was langfristig eine erhebliche Belastung des Sozialsystems nach sich zieht. Obgleich dieser Risiken sind chronische psychische Störung in der Wissenschaft und Forschung nicht ausreichend vertreten (Eikermann & Zacharias, 2003, S. 215). Wie bei der oben genannten Definition von chronischen Krankheiten ist die Definition und Operationalisierung von chronisch psychischen Erkrankungen „bis heute ungelöst [...]: Wann ist ein Patient als chronisch psychisch Kranker anzusehen? Heißt chronisch psychisch krank:

- Es besteht eine häufig wiederkehrende Symptomatik?
- Es existiert eine dauerhafte Symptomatik, womöglich mit Progredienz?
- Chronizität geht unabhängig von der Schwere der Störung in einer Krankheitsdauer von z. B. zumindest 2 Jahren auf?
- Es handelt sich überwiegend um jene Patienten, die unter dem Aspekt der Komorbidität zwei, drei oder noch mehr „Störungen“ aufweisen?
- Es finden sich bei den chronisch psychisch Kranken überwiegend jene, die sozial schlecht angepaßt leben [...]?
- Es sind Patienten, die infolge langjähriger Hospitalisierung und Behandlung psychische Behinderungen aufweisen?“ (Eikermann & Zacharias, 2003, S. 216).

Ein vielversprechender Definitionsversuch wurde bereits in den 1980er Jahren erarbeitet. Die Autoren verknüpften die drei Dimensionen „Schwere der Diagnose“, „Behinderung“ und „Dauer der Erkrankung“ miteinander, um den Personenkreis von chronisch psychisch kranken Menschen näher zu definieren. Nach dieser Definition leiden Patienten mit einer chronischen psychischen Störung an einer schweren und anhaltenden psychischen oder emotionalen Störungen, die ihre Funktionsfähigkeit in Bezug auf Aspekte des täglichen Lebens wie Selbstversorgung, Beziehungen, Arbeit oder Schule beeinträchtigen und die oft eine längere (Krankenhaus-)Versorgung erfordern (Goldman, Gattozzi & Taube, 1981). Die Definition eignet sich auch zur Beschreibung chronisch psychischer Störungen im Rahmen des EM-Rentenverfahrens, da sowohl Aspekte der notwendigen stetig fortgesetzten Behandlung, Einschränkungen der Aktivitäten im Alltag und eine Verminderung der Leistungsfähigkeit angesprochen werden (vgl. DRV, 2014, S. 14).

Neubildungen. Krebserkrankungen stellen unter den Indikationen für eine EM-Rente die zweithäufigste Diagnosegruppe dar (DRV, 2017b). Im Jahr 2016 gab es insgesamt 22.319 (Frauen: 11.379/Männer: 10.940) Rentenzugänge wegen verminderter Erwerbsfähigkeit (DRV, 2017a, S. 57). Insgesamt erkrankten in diesem Jahr 482.500 Menschen an Deutschland an Krebs, davon 229.900 Frauen und 252.600 Männer. Männer und Frauen unterscheiden sich nicht nur hinsichtlich der Erkrankungshäufigkeit, sondern auch hinsichtlich der Tumorlokalisation. Die häufigste Diagnose der Frauen ist Brustkrebs mit 71.640 Neuerkrankungen, während Männer mit 59.520 Neuerkrankungen am häufigsten an Prostatakrebs erkranken. Die insgesamt zweithäufigsten Neubildungen stellen Tumore am Dick- und Enddarm dar (Frauen: 27.210/Männer: 33.370), gefolgt von Tumoren der Lunge (Frauen: 18.810/Männer: 34.690). Insgesamt hat sich die Zahl der Neuerkrankungen seit den 1970er Jahren fast verdoppelt. Dazu haben neben dem demographischen Wandel auch ein industriell geprägter Lebensstilwandel sowie eine sukzessiv bessere medizinische Diagnostik geführt. Durch die gleichzeitige Verbesserung der medizinischen Therapieoptionen steigt auch die Zahl der mit einer Krebserkrankung lebenden Menschen und Langzeitüberlebenden (RKI, 2016, S. 21-22). Der Fortschritt in Chirurgie, Strahlentherapie und systemischer Therapie wird jedoch gleichzeitig von spezifischen Spätkomplikationen begleitet, die in Kombination das Risiko für körperliche Langzeitfolgen erhöhen können (Schilling & Arnold, 2012). Aus Tumorerkrankung und Therapie können sich daher Funktions- und Fähigkeitsstörungen entwickeln, die langfristig zu einer EM-Rente führen (König, 2011, S. 253). Das durchschnittliche Alter bei Beantragung einer EM-Rente durch die Folgen einer Neubildung liegt für Frauen bei 51,0 und für Männer bei 52,8 Jahren (DRV, 2011, S. 10).

Erkrankungen des Skeletts, der Muskeln und des Bindegewebes. Die derzeit dritthäufigste Diagnosegruppe als Indikation einer EM-Rente sind Erkrankungen des Muskel-Skelett-Systems und des Bindegewebes (DRV, 2017b). Demgegenüber stehen 359.527 Leistungen zur medizinischen Rehabilitation (36 % aller durchgeführten Rehabilitationsleistungen), die damit die häufigste Diagnosegruppe für die Beantragung einer Rehabilitationsmaßnahme darstellt. Im Jahr 2016 wurden insgesamt 21.111 EM-Rentennewuzugänge (Frauen: 11.139/Männer: 9.972) verzeichnet. Tabelle 4 beschreibt die Divergenz zwischen der Anzahl durchgeführter Leistungen und Rentennewuzugänge (DRV, 2017c, S. 4). Besonders hervorzuheben ist die hohe Anzahl an medizinischen Rehabilitationsleistungen aufgrund von sonstigen Krankheiten der Wirbelsäule und des Rückens. Die Diagnosekategorie fasst zervikale und sonstige Bandscheibenschäden, sonstige Krankheiten der Wirbelsäule und des Rückens sowie Rückenschmerzen zusammen. Unabhängig von Dauer und Intensität leiden 61,8 % (95 %-CI: 60,7 %-62,8 %) innerhalb von 12 Monaten an Rückenschmerzen (Neuhauser, Ellert & Ziese, 2005).

Tabelle 4: Erbrachte Leistungen der DRV bei Krankheiten des Muskel-Skelett-Systems und des Bindegewebes unterteilt nach Diagnosegruppen (Stand: 2016; modifiziert nach DRV, 2017c, S. 4-5).

Diagnose	Stationäre Leistungen zur medizinischen Rehabilitation	Leistungen zur Teilhabe am Arbeitsleben	EM-Renten-neuzugänge
	Absolute Fallzahlen		
Entzündliche Polyarthropathien	8.987	817	1.828
Arthrose	84.460	6.886	4.592
Sonstige Gelenkkrankheiten	12.543	13.043	574
Systemkrankheiten des Bindegewebes	2.275	167	676
Deformitäten der Wirbelsäule und des Rückens	11.246	1.520	866
Spondylopathien	24.771	1.744	2.165
<i>Sonstige Krankheiten der Wirbelsäule und des Rückens</i>	<i>179.043</i>	<i>34.776</i>	<i>7.733</i>
Sonstige Krankheiten des Weichteilgewebes	27.949	3.915	1.290
Veränderungen der Knochendichte und -struktur	1.061	141	333
Sonstige	7.192	1.761	1.054

Für chronische Rückenschmerzen, definiert als drei Monate und länger anhaltende fast tägliche Schmerzen, beträgt die 12-Monatsprävalenz 18,7 % (17,9 %-19,5 %) und trifft damit einen hohen Bevölkerungsanteil in Deutschland (Neuhauser et al., 2005). Angesichts dieser repräsentativen Ergebnisse erscheint es fraglich, ob die bisher gültige Annahme, dass eine Schmerzremission bei Rückenschmerzen in etwa 90 % Fälle innerhalb von vier Wochen erfolgt, aufrecht zu erhalten ist (DRV, 2009a, S. 5; Pfingsten & Hildebrandt, 2007, S. 407). Unterstützt wird diese Hypothese durch weitere Studienergebnisse, die auch eine hohe Spannweite an Remissionsraten berichten (Croft, Macfarlane, Papageorgiou, Thomas & Silman, 1998; Hoy, Brooks, Blyth & Buchbinder, 2010; Maher, Underwood & Buchbinder, 2017; Pengel, Herbert, Maher & Refshauge, 2003). Die Chronifizierung und Aufrechterhaltung der Schmerzen entsteht durch das Zusammenwirken zahlreicher Einflussfaktoren der somatischen, kognitiven, emotionalen und Verhaltensebene (Pfingsten, 2016, S. 58). Für Anträge auf EM-Rente sind vor allem chronische Schmerzen auf Basis einer Bandscheiben- und bandscheibenassoziierten Erkrankung von Bedeutung. Das durchschnittliche Alter bei Beantragung einer EM-Rente liegt für Frauen bei 51,6 und für Männer bei 52,9 Jahren (DRV, 2009a, S. 6).

Herz-/Kreislaferkrankungen. Erkrankungen des Herzens und des Kreislaufs stellen die vierthäufigste relevante Diagnosegruppe unter den Indikationen für eine EM-Rente dar (DRV, 2017b). Die ischämische oder koronare Herzkrankheit (KHK; Frauen: 618/Männer: 2.990) sowie Kardiomyopathien und Herzinsuffizienz (Frauen: 483/Männer: 1.926) sind die häufigsten Einzeldiagnosen in Bezug auf EM-Renten. Insgesamt wurden 8.698 EM-Renten (Frauen: 2.046/Männer: 6.652) infolge von Herz-/Kreislaferkrankungen gewährt (Schwaab & Franz, 2011, S. 291). Männer beantragen 3-mal so häufig Rentenleistungen als Frauen. Obwohl die Sterblichkeit seit den 1980er Jahren kontinuierlich sinkt, sind Herz-/Kreislaferkrankungen (vor allem die KHK) die häufigsten Todesursachen in Industrieländern. In der für die DRV relevanten Altersgruppe der 45- bis 65-jährigen Versicherten beträgt die Lebenszeitprävalenz einer KHK für Frauen 3,5 % und für Männer 8,6 %. Ab dem 65. Lebensjahr steigt diese stark an (RKI, 2011, S. 82-83). Männer weisen im Gegensatz zu Frauen eine im altersstandardisierten Durchschnitt 3,3-fach höhere Herzinfarktmorbidität auf (RKI, 2006, S. 11). Gleichzeitig weisen jedoch Frauen unter 50 Jahren eine höhere Letalität und eine vermehrte Komplikationsrate im Krankenhaus auf als erkrankte Männer (Solimene, 2010; Vaccarino et al., 2009; Vaccarino, Parsons, Every, Barron & Krumholz, 1999). Herz-/Kreislaferkrankungen beeinflussen die Körperstrukturen und -funktionen von Versicherten. Infolge dieser veränderten Körperstrukturen, wie beispielsweise einer Koronarstenose oder Herzinsuffizienz, können kardiale Durchblutungsstörungen, eine verminderte Pumpleistung sowie Herzrhythmusstörungen auftreten. Diese körperlichen Veränderungen haben direkten Einfluss auf die Erwerbsfähigkeit in Form einer verminderten Gehstrecke oder körperlichen Belastbarkeit (DRV, 2015, S. 9).

Erkrankungen des Stoffwechsels und der Verdauung. Erkrankungen des Stoffwechsel- und Verdauungssystems machen prozentual die kleinste Diagnosegruppe im Spektrum der EM-Renten neuzugänge aus (DRV, 2017b). Im Jahr 2016 wurden 6.172 Anträge (Frauen: 2.583/Männer: 3.589) auf Rente wegen verminderter Erwerbsfähigkeit bewilligt. Auch in dieser Diagnosekategorie ist ein kontinuierlicher Rückgang an EM-Renten und medizinischen Rehabilitationsleistungen zu verzeichnen. Von diesem Trend ausgenommen sind chronisch entzündliche Darmerkrankungen (CED) wie Morbus Crohn und Colitis ulcerosa sowie Diabetes mellitus (Typ 1 und 2), deren Anzahl unverändert hoch bleibt (Warm, Metzler & Glaser, 2018, S. 71). Die Prävalenz von CED liegt zwischen 0,5 % und 1,0 % in der Allgemeinbevölkerung mit ansteigenden Fallzahlen in Industrieländern (Burisch & Munkholm, 2015; Loddo & Romano, 2015). CED sind gekennzeichnet durch wiederholende Entzündungen des Gastrointestinaltrakts mit teilweise extraintestinalen Manifestationen in den Gelenken, an der Haut oder den Augen (Fischer, Rath & Neurath, 2018). Durch die anhaltende Entzündung des Darmtraktes kommt es zu Diarrhoen und

Inkontinenz, was zu Gewichtsabnahme, Schwäche sowie vermehrter Müdigkeit führt und langfristig die Erwerbsfähigkeit beeinflusst (Zillesen, Reichel & Oehler, 2011, S. 366).

Diabetes mellitus beschreibt eine chronische Stoffwechselstörung, bei der die Regulierung des Blutzuckerspiegels über Insulin gestört ist. Typ 1-Diabetes ist eine Autoimmunerkrankung, welche die Insulin-produzierenden Zellen zerstört und daher eine lebenslange Insulinsubstitution herbeiführt. Typ 2-Diabetes beschreibt eine Insulinresistenz der Rezeptoren in den Insulinzielgeweben, wie Leber, Muskel- und Fettgewebe. Die daraufhin einsetzende Überproduktion von Insulin führt langfristig zur Ermüdung und damit zu einem Sekretionsdefizit der Insulin-produzierenden Zellen (Hien, Böhm, Claudi-Böhm, Krämer & Kohlhas, 2013, S. 18). Die 12-Monats-Prävalenz eines Diabetes mellitus beträgt 7,7 %, wobei Frauen mit 7,0 % seltener betroffen sind als Männer mit 8,6% (Heidemann, Kuhnert, Born & Scheidt-Nave, 2017). Etwa 95% der Betroffenen sind an Typ 2-Diabetes erkrankt, was mit Übergewicht und mangelnder Bewegung assoziiert ist und meist ab dem 40. Lebensjahr auftritt (Jacobs & Rathmann, 2018, S. 9). Unbehandelt oder schlecht eingestellt führt das Fehlen von Insulin zu erhöhten Blutzuckerkonzentrationen, was Schäden an Blutgefäßen, Neuropathien, Nephropathien, Retinopathien und das diabetischen Fußsyndrom verursachen kann (Diehl, Knisel & Haupt, 2011, S. 271).

1.3.2 Entwicklung der Morbidität

Die Prävalenz chronischer Erkrankungen wird, vor dem Hintergrund der demographischen Entwicklung, als Folge des steigenden Durchschnittsalters der Bevölkerung zukünftig voraussichtlich weiter zunehmen. Damit erhöht sich auch die Wahrscheinlichkeit, gleichzeitig an mehreren chronischen Erkrankungen zu leiden und die Multimorbidität steigt (Prütz et al., 2014, S. 123). Aus soziologischer Sicht ist diese in der Literatur beschriebene Implikation nicht die einzige Möglichkeit, wie sich Gesundheit und Krankheit in einer alternden Gesellschaft zueinander verhalten werden. Einer alternativen Sichtweise liegt die Annahme zugrunde, dass Alter und Krankheit nicht zwangsläufig miteinander verknüpft sind. Entsprechend wird eine höhere Krankheitslast nicht automatisch durch ein höheres gesellschaftliches Durchschnittsalter verursacht. Im Folgenden werden zwei theoretische Szenarien vorgestellt, die versuchen, zukünftige krankheitsbezogene Entwicklungen zu skizzieren: die „Expansionsthese“ (alternativ: Medikalisierungsthese) und die „Kompressionsthese“ (Blüher & Kuhlmei, 2016, S. 315; siehe Abb. 3).

Expansion und Kompression von Morbidität. Die **Kompressionsthese** geht davon aus, dass die gesunde Lebenszeit und die mittlere Lebensdauer ansteigen werden. Morbidität und Mortalität verschieben sich somit in ein höheres Lebensalter. Infolgedessen ist eine intensive medizinische Betreuung erst in den letzten Lebensjahren erforderlich. „Kompression“ meint das Verschieben schwerer gesundheitlicher Beeinträchtigungen an

das Lebensende und wird nach dieser These durch einen allgemein hohen Gesundheitsstatus infolge besserer Lebensbedingungen und des kontinuierlichen medizinischen Fortschritts erreicht (Fries, 1980, 2003, 2005). Als Folge der komprimierten Krankheitsphase am Lebensende könnten die relativen und die absoluten Pro-Kopf-Ausgaben für das Sozialversicherungssystem langfristig sinken (Breyer, 2015). Im Gegensatz dazu geht die **Expansionsthese** von einer Zunahme chronischer Erkrankungen und Multimorbidität im Alter aus. Es wird beschrieben, dass im Rahmen der steigenden Lebenserwartung vermehrt medizinische Dienstleistungen in Anspruch genommen werden. Die Verlängerung der Zeit in Krankheit ist zurückzuführen auf den medizinisch-technischen Fortschritt, der es ermöglicht, Menschen trotz Krankheit höhere Altersstufen erreichen zu lassen. In der Folge entsteht durch die Summation manifestierter chronischer Erkrankungen im Alter eine erhöhte Multimorbidität (Verbrugge, 1984). Gesamtgesellschaftlich betrachtet nimmt dadurch die Morbidität im Verhältnis zur Gesamtbevölkerung zu, was die medizinischen Behandlungsausgaben sukzessiv ansteigen lässt (Breyer, 2015; Heigl, 2004).

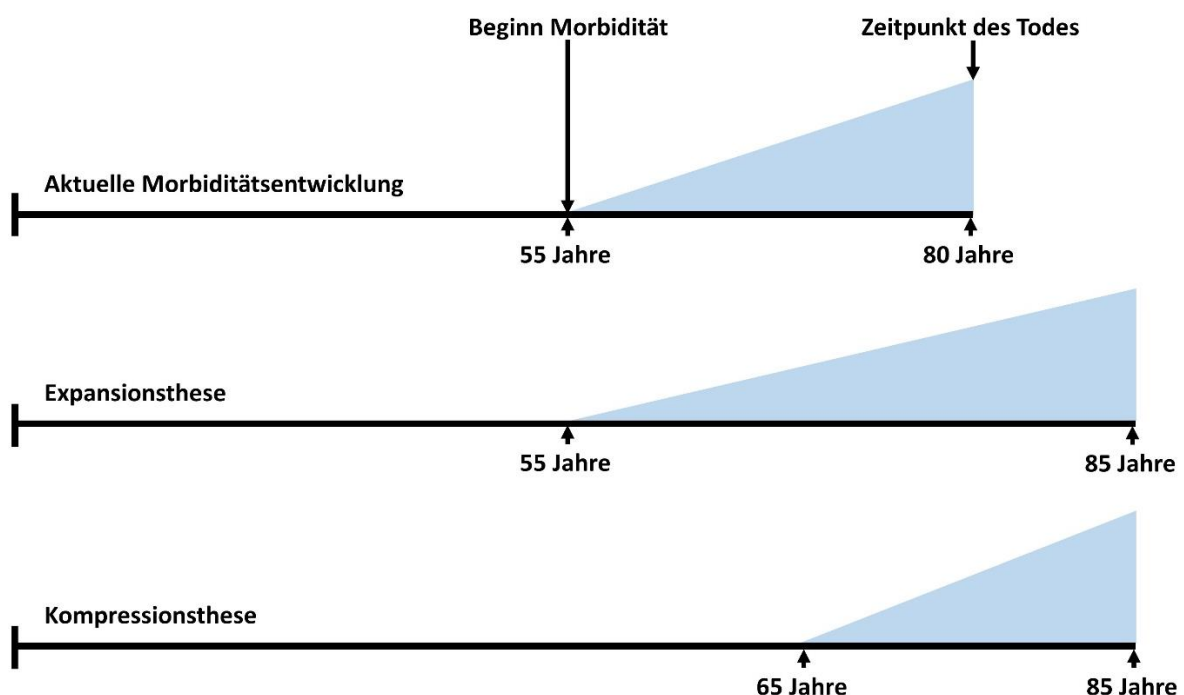


Abbildung 3: Szenarien für die Entwicklung von Morbidität und Langlebigkeit (modifiziert nach Fries, 2003, S. 456).

Derzeit existieren nur unzureichende Forschungserkenntnisse, um abschließend klären zu können, welche These die Morbiditätsentwicklung am wahrscheinlichsten voraussagt (Chatterji, Byles, Cutler, Seeman & Verdes, 2015). Tendenziell wird eher von der Zunahme der Lebenserwartung in Gesundheit (Kompression) ausgegangen (Kroll & Ziese, 2009, S. 110). Wahrscheinlich bedarf es jedoch einer Zusammenführung beider Konzepte, um Veränderungen in einer alternden Bevölkerung wirklichkeitsnah abzubilden.

Bi-Modalitätskonzept. Das Konzept der **Bi-Modalität** beschreibt den Versuch einer Synthese der Expansions- und Kompressionsthese. Der Entwurf basiert auf der Annahme, dass Menschen als Teil im gemeinschaftlichen Gefüge abhängig von verhaltens-, verhältnis- und versorgungsbezogenen Faktoren individuell und heterogen leben und altern. Die gewonnenen gesunden Lebensjahre sind entsprechend ungleich verteilt. Aus dieser Annahme resultiert, dass Individuen bzw. ganze Bevölkerungsgruppen existieren, die entsprechend der Kompressionsthese, ein längeres Leben in Gesundheit erreichen und andere Teile der Bevölkerung, entsprechend der Expansionsthese, im Alter gehäuft an chronischen Erkrankungen und Funktionseinschränkungen leiden (Blüher & Kuhlmei, 2016, S. 316).

Alle drei vorgestellten Thesen gehen im Kern von einer Verlängerung des Lebens durch den kontinuierlichen medizinisch-technischen Fortschritt aus. Die Entscheidung, ob chronische Erkrankungen und Morbidität komprimiert oder expandiert werden, unterliegt jedoch ganz individuellen Faktoren. Die Morbidität am Lebensende sollte also nicht nur dem Alterungsprozess per se zugeschrieben werden, sondern einem breiten Spektrum an persönlichen Risikofaktoren und Ressourcen. Dazu zählen auch persönliche Faktoren und Umweltfaktoren, wie „gesundheitsrelevante Verhaltensmuster, Lebensstile, sozio-ökonomische Lebenslagen und Zugangsmöglichkeiten zu medizinischer Versorgung“ (Blüher & Kuhlmei, 2016, S. 316). Jeder Versicherte hat selbstbestimmt die Möglichkeit, durch einen gesunden Lebensstil die Kompression von altersassoziierten Erkrankungen zu beeinflussen.

Der Sozialstaat kann diese Aufgabe durch geförderte Gesundheitsförderungs- und Präventionsprogramme unterstützen. So könnten die monetären Gesundheitsausgaben der Sozialversicherungen im Rahmen einer frühzeitigen salutogenetischen Ressourcenförderung nach Antonovsky reduziert werden (vgl. Antonovsky, 1997). Auf der Makroebene der politischen und institutionellen Entscheidungsträger ist allen beteiligten Akteuren der Stellenwert von Prävention und Gesundheitsförderung als Schlüsselkonzept für die Verhinderung von Multimorbidität bekannt (Pawils, Boettcher, Metzner, Plaumann & Walter, 2015). Im Jahr 2015 wurde daher das „Gesetz zur Stärkung der Gesundheitsförderung und der Prävention“ (Präventionsgesetz - PräVG) verabschiedet. Es verpflichtet die gesetzlichen Krankenkassen zur Gesundheitsförderung und Primärprävention, den Spitzenverband der gesetzlichen Krankenversicherung zur Berücksichtigung nationaler Gesundheitsziele und betrieblicher Gesundheitsförderung sowie die Pflegekassen zur Erbringung von Präventionsleistungen in voll- und teilstationären Pflegeeinrichtungen. Darüber hinaus soll Prävention in unterschiedlichen Lebenswelten (z. B. Kita, Schule, Betrieb) integriert werden, um alle Nutzer, unabhängig ihrer sozialen Herkunft, zu erreichen (Meierjürgen, Becker & Warnke, 2016).

Auch die DRV ist primärpräventiv tätig. Seit Inkrafttreten der Änderung des § 31 Abs. 1 Satz 1 Nr. 2 SGB VI am 01.01.2009 hat sie die Möglichkeit ambulante und stationäre medizinische Präventionsleistungen für Versicherte mit physisch und psychisch belastenden Berufen anzubieten. Vor dem Jahr 2009 war diese Versicherungsleistung ausschließlich im stationären Setting verfügbar. Bevor sich eine chronische Erkrankung manifestiert, soll Versicherten durch diese Maßnahme die Möglichkeit zur Regenerierung und zur Vorbeugung belastungsabhängiger Gesundheitsstörungen gegeben werden. Unter Berücksichtigung des Wandels der Arbeitswelt in Richtung des Informations- und Dienstleistungssektors fokussieren die angebotenen Präventionsmaßnahmen neben den klassischen Präventionsthemen wie schwere physische Arbeit, ungesundes Ernährungsverhalten, Bewegungsmangel und Nikotinkonsum, auch besonders psychische belastende Erwerbstätigkeiten, hohe Dauieranforderungen und Tätigkeiten im ständigen Stehen oder Sitzen. Langfristig besteht dadurch die Chance die berufliche Leistungsfähigkeit aufrechtzuerhalten und EM-Renten zu vermeiden (Moser, Fischer & Korsukéwitz, 2010).

1.3.3 Risiko- und Schutzfaktoren

Risikofaktoren sind pathogenetische Einflussfaktoren, die bei Vorhandensein die Auftretenswahrscheinlichkeit einer Erkrankung erhöhen. Dabei sind kausale Risikofaktoren nicht als deterministische Verursacher von Krankheiten zu verstehen, sondern tragen lediglich zur Erhöhung der Auftretenswahrscheinlichkeit einer chronischen Erkrankung im komplexen Prozess der multifaktoriellen Genese bei (Vogel et al., 2016, S. 342). Laut WHO sind Bluthochdruck, Tabak- und Alkoholkonsum, hohes Blutcholesterin, Übergewicht, zu geringer Obst- bzw. Gemüseverzehr und Bewegungsmangel die sieben führenden Risikofaktoren für die Entstehung chronischer Erkrankungen. Diese Risikofaktoren sind nicht getrennt voneinander zu betrachten. Ein Risikofaktor ist mindestens mit zwei der häufigsten chronischen Erkrankungen assoziiert und umgekehrt (WHO, 2006, S. 4).

Schutzfaktoren hingegen sind salutogenetische Einflussfaktoren, die bei Vorhandensein vor Krankheit schützen und die Gesundheit erhalten. Mehr als 50 % des bevölkerungsbezogenen Sterblichkeitsrisikos lässt sich auf veränderbare Lebensstilfaktoren zurückführen. Mit Nikotinverzicht, gesunder Ernährung, körperlicher Aktivität sowie mäßigem Alkoholkonsum werden nicht nur die Risikofaktoren Hypercholesterinämie und Übergewicht abgebaut, sondern auch das Risiko für koronare Herzkrankheit, Herzinfarkt, Schlaganfall, Diabetes mellitus Typ 2, chronische Rückenschmerzen, altersbedingte Stürze und Depression nachhaltig gesenkt (Vogel et al., 2016, S. 339).

Soziale Ungleichheit und Gesundheit. Das Vorhandensein der von der WHO definierten Risikofaktoren wird maßgeblich vom sozialen Status (Schulabschluss, Berufsstatus, Einkommen) beeinflusst. Folglich entsteht gesundheitliche Ungleichheit in Abhängigkeit vom sozialen Status (vgl. Mielck, 2012). Abbildung 4 beschreibt einen möglichen Erklärungsansatz dieser kausalen Ursachenzuschreibung:

Durch soziale Ungleichheiten in der Gesellschaft entstehen Unterschiede bezüglich gesundheitlicher Belastungen und Ressourcen. Somit ist ein niedriger sozio-ökonomischer Status assoziiert mit höheren Belastungen und niedrigeren Ressourcen.

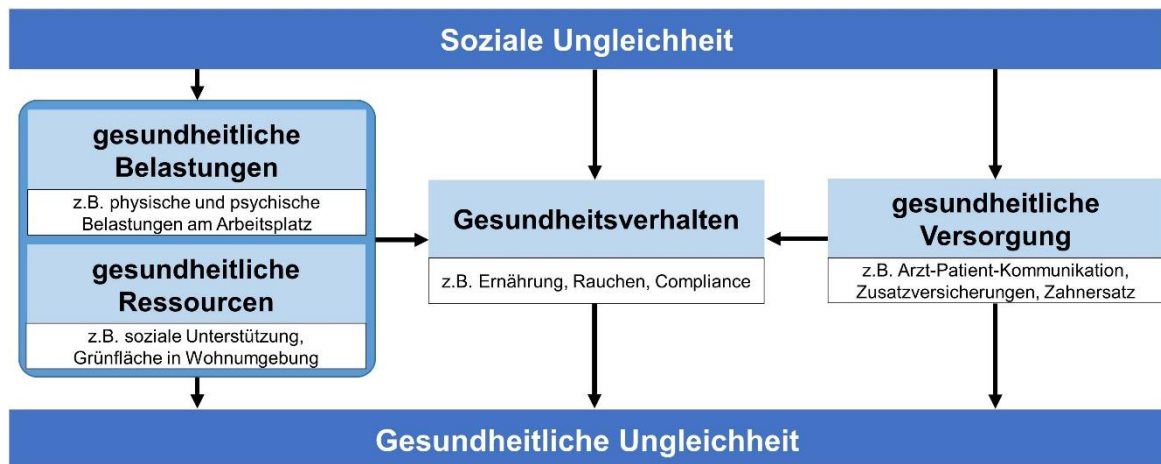


Abbildung 4: Modell zur Erklärung gesundheitlicher Ungleichheit (modifiziert nach Mielck, 2012, S. 136).

Belastungen und Ressourcen entstehen aus den Lebensverhältnissen jedes einzelnen Individuums und wirken sich direkt auf den Gesundheitszustand aus (z. B. durch schwere körperliche Arbeit). Zudem bestimmen die Lebensverhältnisse vornehmlich auch das **Gesundheitsverhalten** (Kalousova & Burgard, 2014), welches wiederum über das Vorhandensein von Risiko- und Schutzfaktoren Einfluss auf den Gesundheitszustand nimmt (z. B. durch ungesunde Ernährung und Rauchen). Als dritter Modellbaustein wird die **gesundheitliche Versorgung** selbst vom sozialen Status beeinflusst. Dies äußert sich in Form von Unterschieden beim Versorgungsangebot, der Inanspruchnahme und der Qualität von dargebotenen Gesundheitsleistungen (Mielck, 2012, S. 135). Zusammenfassend führt soziale Ungleichheit zu einer Ungleichverteilung von Erkrankungen und Risikofaktoren in der Gesellschaft, was sich in einer ungleichverteilten Morbidität und Lebenserwartung widerspiegelt (Lampert, Richter, Schneider, Spallek & Dragano, 2016). Der Zusammenhang zwischen sozialem Status und Morbidität überträgt sich schließlich auch auf das rehabilitative System und die Beantragung von EM-Renten (Fach, Bethge, Spanier & Richter, 2017; Hagen et al., 2011; Mika, 2013).

Entsprechend der Komplexität des Ineinandergreifens von Risiko- und Schutzfaktoren bei der Entstehung und Aufrechterhaltung von chronischen Erkrankungen bedarf es bei der

Sozialmedizinischen Begutachtung von Versicherten der DRV eines komplexen und ganzheitlichen Betrachtungsmodells. Um den funktionellen Gesundheitsstatus vor dem Hintergrund der individuellen Umwelt- und personbezogenen Faktoren zu bestimmen, hat die DRV das Modell der Internationalen Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) als Grundlage der Begutachtung definiert (Ostholt-Corsten & Schuntermann, 2011, S. 73).

1.4 Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF)

Die ganzheitliche Sozialmedizinische Begutachtung erfordert eine mehrdimensionale Betrachtungsweise von Krankheit und Behinderung entsprechend des biopsychosozialen Konzeptes der ICF (DRV, 2012, S. 26). Während in der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD) Erkrankungen anhand ihrer Symptome und Syndrome klassifiziert werden, beschreibt die zur ICD komplementäre ICF die Funktionsfähigkeit und Einschränkungen, die mit diesen Erkrankungen assoziiert sind (WHO, 2001, S. 4).

Die ICF ist ein von den Vereinten Nationen anerkanntes Mehrzweckklassifikationssystem, das für unterschiedliche Disziplinen und Anwendungsbereiche entwickelt wurde. Im Kern ist die ICF ein systematisches Verschlüsselungssystem, mit dem Komponenten der Funktionsfähigkeit und Behinderung sowie Komponenten der Kontextfaktoren klassifiziert werden können. Das Klassifikationssystem wurde mit dem Ziel entwickelt, eine gemeinsame Sprache zur Beschreibung des Gesundheitszustandes (d.h. Funktionsfähigkeit und Behinderung) und zur fachübergreifenden, allgemeingültigen Kommunikation zu etablieren. Dadurch können Forschungs- und Versorgungsdaten länderübergreifend und interinstitutionell verglichen und analysiert werden. Die ICF bietet für die verschiedenen Anwender vielfältige theoretisch Anwendungsgebiete (Tab. 5).

Tabelle 5: Anwendungsgebiete der ICF (modifiziert nach DIMDI/WHO, 2005, S. 11).

Statistisches Instrument	Zur Datenerhebung und -dokumentation in Bevölkerungsstudien
Forschungsinstrument	Zur Messung von Ergebnissen, Lebensqualität und Umweltfaktoren
Instrument der gesundheitlichen Versorgung	Zur Bedarfsbeurteilung, Behandlungsanpassung, berufsbezogenen Beurteilung, Rehabilitation und Ergebnisevaluation
Sozialpolitisches Instrument	Zur Planung der sozialen Sicherheit, für Entschädigungssysteme und Politikgestaltung und Umsetzung
Pädagogisches Instrument	Zur Entwicklung von Curricula, Schaffung von Problembewusstsein und Anstoß für soziales Handeln

Im Rahmen der Sozialmedizinischen Begutachtung der DRV wird die ICF als Instrument der gesundheitlichen Versorgung zur Beurteilung der berufsbezogenen Leistungsfähigkeit und als sozialpolitisches Instrument eingesetzt. So finden sich im SGB IX zur Rehabilitation und Teilhabe behinderter Menschen grundlegende Aspekte der ICF-Terminologie wieder, die damit auch sozialrechtlich verankert worden sind (Ostholt-Corsten & Schuntermann, 2011, S. 68).

1.4.1 Modell der ICF

Die Beschreibung der Funktionsfähigkeit im Konzept der ICF ist universell, bei Menschen aller Altersklassen anwendbar und unterscheidet in ihrer Anwendung nicht zwischen gesunden oder kranken Menschen (Meyer, Quittan & Rauch, 2017, S. 73). Funktionsfähigkeit und Behinderung bilden ein Kontinuum, welches sich aus den Dimensionen der funktionalen Gesundheit ableitet. Nach dem Konstrukt der ICF ist eine Person funktional gesund, wenn „ihre körperlichen Funktionen [...] und Körperstrukturen denen eines gesunden Menschen entsprechen [...], sie all das tut oder tun kann, was von einem Menschen ohne Gesundheitsproblem [...] erwartet wird [...], sie ihr Dasein in allen Lebensbereichen, die ihr wichtig sind, in der Weise und dem Umfang entfalten kann, wie es von einem Menschen ohne gesundheitsbedingte Beeinträchtigung der Körperfunktionen oder -strukturen oder der Aktivitäten erwartet wird“ (DIMDI/WHO, 2005, S. 4).

Dieser Definition entsprechend, besteht die ICF aus zwei Teilen mit jeweils zwei Komponenten. Der erste Teil „Funktionsfähigkeit und Behinderung“ beinhaltet die Komponenten (1) Körperfunktionen und -strukturen und (2) Aktivitäten und Partizipation (Teilhabe). Der zweite Teil „Kontextfaktoren“ umfasst die Komponenten (3) Umweltfaktoren und (4) personbezogene Faktoren (Abb. 5 und Tab. 6). Alle Komponenten stehen in sich beeinflussender Wechselwirkung zueinander. Die Komponenten werden in Domänen eingeteilt, die aus einzelnen Kategorien bestehen. Diese bilden die Einheiten (Items) der Klassifikation. Der Gesundheitszustand kann durch Auswahl eines oder mehrerer Codes aus der entsprechenden Kategorie dokumentiert werden. Für jede Domäne stellt die ICF eine spezifische Anzahl an Codes zur Verfügung. Ergänzt wird das Modell durch die Möglichkeit der quantitativen Beurteilung der Codes zur Beschreibung des Ausmaßes bzw. der Größe der Funktionsfähigkeit oder Behinderung auf einer 5-stufigen Skala von Problem „nicht vorhanden“ bis „voll ausgeprägt“ (DIMDI/WHO, 2005, S. 16).

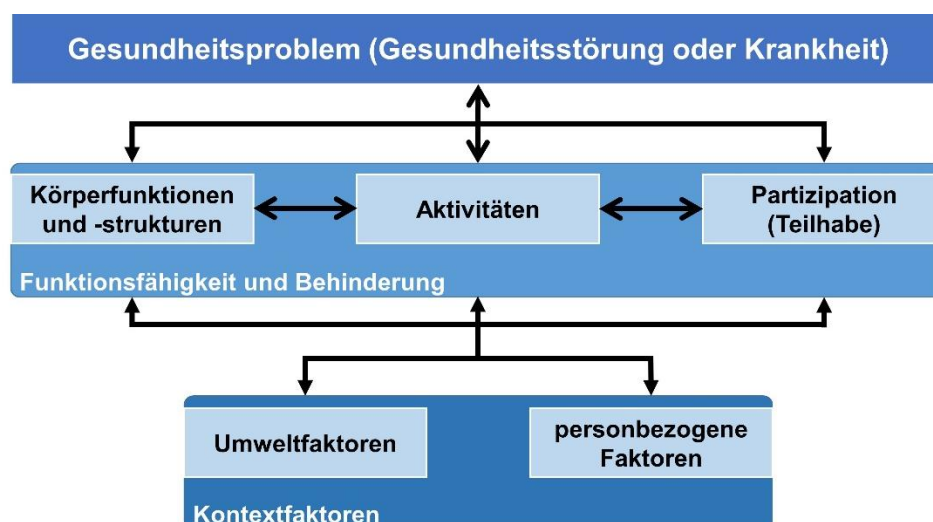


Abbildung 5: ICF-Modell der Funktionsfähigkeit und Behinderung mit Wechselwirkungen zwischen den Komponenten (modifiziert nach DIMDI/WHO, 2005, S. 23).

Tabelle 6: Definition der ICF Komponenten (modifiziert nach DIMDI/WHO, 2005, S. 17-19; Ostholt-Corsten & Schuntermann, 2011, S. 69).

Körperfunktionen	beschreiben die physiologischen und psychologischen Funktionen des Körpers
Körperstrukturen	beschreiben die einzelnen anatomischen Teile des Organismus (Organe, Gliedmaßen und ihre Bestandteile)
Aktivitäten	beschreibt die Durchführung von Aufgaben oder Handlungen
Partizipation (Teilhabe)	beschreibt die Integration in eine Lebenssituation oder einen Lebensbereich
Umweltfaktoren	beschreibt die materielle, soziale und einstellungsbezogene Umweltumgebung des Menschen
Personbezogene Faktoren	beschreibt persönliche Eigenschaften, Attribute: Alter, Geschlecht, Charakter, Lebensstil, Coping, sozialer Hintergrund, Bildung/Ausbildung, Beruf, Erfahrung, Prädispositionen

Zumeist verursacht eine Grunderkrankung (nach ICD) eine Einschränkung der funktionalen Gesundheit. Die Einschränkung manifestiert sich dann in den Komponenten Körperfunktionen und -strukturen, Aktivitäten oder Partizipation (Teilhabe). Die Kontextfaktoren können ergänzend sowohl positive als auch negative Auswirkungen auf die Funktionsfähigkeit haben (vgl. Kap. 1.3.3. Risiko- und Schutzfaktoren).

1.4.2 Anwendung der ICF in der Sozialmedizinischen Begutachtung

15 Jahre nach Einführung der ICF liegen für viele der ursprünglich intendierten Anwendungsgebiete (vgl. Tab. 5) konkrete Konzepte und Praxisumsetzungen vor (Cerniauskaite et al., 2011; Wenzel & Morfeld, 2016). In der Rehabilitation wird die ICF bereits seit einigen Jahren als Instrument der Behandlungssteuerung, Therapieplanung und Assessmentinstrument diskutiert und eingesetzt (vgl. z. B. Bülau, Kessemeier, Petermann, Bassler & Kobelt, 2016; Bürger & Deck, 2009; Ewert, Cieza & Stucki, 2002; Farin, 2008; Linden & Baron, 2005; Spies, Brütt, Freitag & Buchholz, 2015). Es wird jedoch deutlich, dass das biopsychosoziale Modell der ICF derzeit unsystematisch genutzt wird und eine übergreifende Koordination zur Implementierung vorhandener Ansätze weitgehend fehlt (Deutsche Vereinigung für Rehabilitation, 2015; Schliehe & Ewert, 2013; Wenzel & Morfeld, 2017).

Entsprechend der ubiquitären Relevanz für die ganzheitliche biopsychosoziale Beschreibung von Funktionsfähigkeit und Behinderung ist die ICF inzwischen konzeptionelle Grundlage für Anwendungsbereiche in der Sozialgesetzgebung (SGB IX) und im Sozialversicherungssystem (Arbeitslosenversicherung, Krankenversicherung,

Rentenversicherung, gesetzliche Unfallversicherung, Pflegeversicherung) geworden (BAR, 2016; Ewert, Freudenstein & Stucki, 2008; Schliehe & Ewert, 2013). Die DRV beschreibt in allen Sozialmedizinischen Leitlinien zur Begutachtung das biopsychosoziale Modell und die ICF als wesentliche Bestandteile zur Beurteilung der Beeinträchtigungen der Teilhabe.

Die Begutachtung des Leistungsvermögens ist ein komplexer Beurteilungsprozess, der vor dem Hintergrund der individuellen Persönlichkeit und des subjektiven Empfindens des Versicherten durchgeführt werden muss. Durch die Anwendung einer gemeinsamen Sprache, entsprechend der Codes in der ICF, könnten Gutachter die Beschreibung der Leistungsfähigkeit im sozialmedizinischen Gutachten weitestgehend standardisieren und harmonisieren (Kirschneck, Legner, Armbrust, Nowak & Cieza, 2015). Bei der Begutachtung haben Begriffe wie „Leistung“, „Leistungs-“ und „Funktionsfähigkeit“ eine zentrale Bedeutung. Durch Anwendung der ICF können die Wechselwirkungen zwischen den Anforderungen der Arbeitsumwelt (Umweltfaktoren) und der aktuellen Leistungsfähigkeit des Versicherten (Körperfunktionen- und -strukturen, personbezogene Faktoren, Aktivitäten und Teilhabe) standardisiert beschrieben und Diskrepanzen aufgedeckt werden (Ewert et al., 2008). Beim Versuch des Übertrags der ICF auf eine konkrete Begutachtungssituation bleibt jedoch offen, wie das Modell in die Gesamtbeurteilung des Leistungsvermögens integriert werden soll. Bei einer Gesamtzahl von 1.424 Codes der ICF bleibt unklar, welche Größen aus welchen ICF-Domänen im Begutachtungsprozess herangezogen werden können und sollen (Körner, 2005). Dies liegt vorrangig darin begründet, dass die ICF ein beschreibendes Modell darstellt und keine Klassifikation funktionaler Diagnosen vorsieht. Darüber hinaus ist die ICF per se auch kein in der Begutachtung einsetzbares Assessmentinstrument zur Messung der Funktionsfähigkeit und Behinderung (Ostholt-Corsten & Schuntermann, 2011, S. 71). Das Modell kann jedoch hinsichtlich der Sozialmedizinischen Begutachtung dem durchführenden Gutachter Hinweise geben, welche Domänen im Rahmen der Untersuchung zu erfassen sind (Körner, 2005).

Die DRV weist in ihren Leitlinien ausdrücklich darauf hin, dass die umfängliche ICF-Implementierung noch zukünftiger Entwicklungsschritte bedürfe. Zur Anpassung an den Kontext der sozialmedizinischen Praxis brauche es eine weitere Ausdifferenzierung und Operationalisierung der ICF-Items. Relevante sozialmedizinische als auch sozialrechtliche Terminologien wie Beruf, Leistungsvermögen in der bisherigen Tätigkeit, Leistungsvermögen auf dem allgemeinen Arbeitsmarkt, sozialmedizinische Prognose, quantitatives Leistungsvermögen oder die Fähigkeit, mehr als 500 Meter in unter 20 Minuten laufen können, sind derzeit mit der ICF nicht kodierbar. Folglich ist der Einsatz in der Begutachtung aktuell sehr gering ausgeprägt (DRV, 2017c, S. 7).

Core Set. Zur Erhöhung der Anwenderfreundlichkeit und Reduktion der 1424 ICF-Codes wurden sogenannte ICF Core Sets entwickelt. Die Sets verknüpfen Erkrankungen mit den dazu assoziierten wichtigsten ICF-Funktionskategorien. Core Sets werden mittels Expertenkonsens, Literaturrecherche und Delphi-Verfahren entwickelt. Ein ICF Core Set enthält eine Liste von Kategorien mit so wenigen Codes wie möglich, um praktisch anwendbar zu sein, aber so viele wie nötig, um ausreichend umfassend das typische Spektrum von Problemen und Funktionseinschränkungen bei Patienten mit einer bestimmten Erkrankung zu beschreiben. Im Anschluss kann für jede kodierte Kategorie der Schweregrad der Funktionseinschränkung beurteilt werden. Zu Beginn der Entwicklung der Core Sets im Jahr 2004 wurden zunächst für zwölf chronische Volkskrankheiten Sets entwickelt: generalisiertes Schmerzsyndrom, Rückenschmerzen, Arthrose, Osteoporose, rheumatoide Arthritis (RA), chronische ischämische Herzerkrankungen, Diabetes mellitus, Übergewicht, obstruktive Lungenerkrankungen, Brustkrebs, Depression und Schlaganfall (Cieza et al., 2004). Die Core Sets wurden sukzessiv evaluiert und für neue Indikationen weiterentwickelt.

Im Jahr 2008 wurde ein ICF Core Set speziell für die „Begutachtung von Beeinträchtigungen (disability evaluation) in der Sozialversicherung“ (EUMASS Core Set) in einem europaweiten formalen Entscheidungsfindungsprozess entwickelt und konsentiert. Ziel war es, ein Hilfsmittel für Sozialmediziner im Versicherungswesen zu generieren. Das Set soll bei der Entscheidungsfindung helfen, die Qualität der sozialmedizinischen Arbeit verbessern, die interprofessionelle Kommunikation fördern sowie eine Basis für die sozialmedizinische Forschung darstellen. Primär ist das Core Set für die Beurteilung von Versichertenansprüchen auf langfristige Leistungen, wie z. B. eine EM-Rente, vorgesehen. Aus 191 Vorschlägen für ICF-Kategorien wurden 20 für das Core Set ausgewählt. Fünf davon stammten aus der Domäne Körperfunktionen (Komponente b) und fünfzehn aus der Domäne Aktivität und Partizipation (Komponente d; Tab. 7). Bei der Entwicklung wurden noch anderweitige Kategorien wie spezifische Arbeitseinschränkungen, gesundheitsbezogene Einschränkungen, Zeitdimensionen und Persönlichkeitsfaktoren als mögliche Bestandteile des Core Sets vorgeschlagen. Es wurde jedoch kein Konsens gefunden, wie diese zu kodieren seien (Brage, Donceel & Falez, 2008). Ausschließlich in den krankheitsspezifischen ICF Core Sets werden Umweltfaktoren berücksichtigt, die einen wesentlichen Bestandteil bei der Beurteilung der Leistungsfähigkeit darstellen.

Ergänzend hierzu konnte eine Studie am Beispiel der Core Sets für das generalisierte Schmerzsyndrom und chronische Rückenschmerzen feststellen, dass der relevante Inhalt von medizinischen Leistungsbeurteilungen mithilfe der Kombination von vorhandenen Core Sets erfasst werden könne. Es konnten sowohl Aspekte der Funktionsfähigkeit als auch der Umweltfaktoren dargestellt werden. Dies sei jedoch nur durch die Erweiterung bestehender

1.4 Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF)

Core Sets, u.a. durch Kategorien der Core Sets für relevante Komorbiditäten wie Depression und Adipositas, möglich gewesen. Durch die höhere Anzahl von Codes bei der Kombination mehrerer ICF Core Sets haben sich jedoch die Wirkungsgrade gegenüber dem ursprünglichen indikationsspezifischen Core Set deutlich verringert (Schwegler et al., 2012).

Tabelle 7: EUMASS Core Set: ICF Core Set für die „Begutachtung von Beeinträchtigungen in der Sozialversicherung“ (modifiziert nach Brage, Donceel & Falez, 2008. S. 1395).

Code	Kapitel	Titel
b164	Mentale Funktionen	Höhere kognitive Funktionen
b280	Sinnesfunktionen und Schmerz	Schmerz
b455	Funktionen des kardiovaskulären, hämatologischen, Immun- und Atmungssystems	Funktionen der kardiorespiratorischen Belastbarkeit
b710	Neuromuskuloskelettale und bewegungsbezogene Funktionen	Funktionen der Gelenkbeweglichkeit
b730	Neuromuskuloskelettale und bewegungsbezogene Funktionen	Funktionen der Muskelkraft
d110	Lernen und Wissensanwendung	Zuschauen
d115	Lernen und Wissensanwendung	Zuhören
d155	Lernen und Wissensanwendung	Sich Fertigkeiten aneignen
d177	Lernen und Wissensanwendung	Entscheidungen treffen
d220	Allgemeine Aufgaben und Anforderungen	Mehrfachaufgaben übernehmen
d240	Allgemeine Aufgaben und Anforderungen	Mit Stress und anderen psychischen Anforderungen umgehen
d399	Kommunikation	Kommunikation, nicht näher bezeichnet
d410	Mobilität	Eine elementare Körperposition wechseln
d415	Mobilität	In einer Körperposition verbleiben
d430	Mobilität	Gegenstände anheben und tragen
d440	Mobilität	Feinmotorischer Handgebrauch
d445	Mobilität	Hand- und Armgebrauch
d450	Mobilität	Gehen
d470	Mobilität	Transportmittel benutzen
d720	Interpersonelle Interaktionen und Beziehungen	Komplexe interpersonelle Interaktionen

Zusammenfassend sind bei der Begutachtung immer die einzigartige individuelle Erfahrung und Persönlichkeit des Antragstellers zu berücksichtigen, um die Arbeitsfähigkeit umfassend beurteilen zu können. Das EUMASS Core Sets hat in dieser Hinsicht noch

weiteren Optimierungsbedarf. Es wird empfohlen, dieses Core Set immer in Kombination mit indikationsspezifischen Core Sets bei der Begutachtung von Erwerbsminderung einzusetzen (Schwegler, Selb, Escorpizo & Stucki, 2015, S. 423).

Am Beispiel der ICF Core Sets für chronische Rückenschmerzen und des generalisierten Schmerzsyndroms versuchte die DRV im Jahr 2008 in einem Forschungsprojekt zu überprüfen, ob krankheitsspezifische Core Sets die Standardisierung von Gutachten unterstützen können. Dazu wurden 600 ärztliche Gutachten in die Sprache der ICF überführt. Mit der sogenannten Linkage-Methode wurden Gutachteninhalte den einzelnen ICF Kategorien der Core Sets retrospektiv zugeordnet (Kirschneck et al., 2008). Die Ergebnisse dieser Studie zeigen, dass grundsätzlich ein Übertrag der Inhalte in die Sprache der ICF möglich ist. Die Gutachten beider Indikationen bilden aber rund 40 % der 1.424 ICF-Kategorien ab. Nur ein geringer Anteil der ICF-Kategorien konnte in mindestens 50 % der Gutachten kodiert werden. Ergänzend zeigen die Ergebnisse, dass 17 der 20 Kategorien des EUMASS Core Sets in allen Gutachten mindestens einmal genannt wurden. Grundsätzlich sind sozialmedizinische Gutachten nicht einheitlich und variieren infolge der individuellen Fallkonstellation in der Begutachtungssituation. Auch die Autoren der Studie kommen zu dem Schluss, dass krankheitsspezifische Core Sets erweitert und mit anderen ICF-Kategorien kombiniert werden müssen, um die Beschreibung der Leistungsfähigkeit im sozialmedizinischen Gutachten adäquat abbilden zu können. Zur Beschreibung des Leistungsvermögens bei chronischen Rückenschmerzen wären 77 ICF-Kategorien und für das generalisierte Schmerzsyndrom 80 ICF-Kategorien notwendig. Die Differenzierung gestattet es den Gutachtern, Schädigung, Aktivitäten und Teilhabe sowie Umweltfaktoren separat zu beschreiben und in der Epikrise zusammenzuführen (Kirschneck et al., 2015).

Trotz der Anwendungsmöglichkeiten der ICF trägt diese nicht zwangsläufig zu einer Qualitätsverbesserung der Gutachten bei. Durch die Beschreibung der Kategorien kann der Aufbau eines Gutachtens weitgehend standardisiert werden, jedoch können gutachterliche Schlussfolgerungen über die Leistungsfähigkeit aus den erhobenen Befunden dadurch weiterhin nicht gesteuert werden (Schliehe & Ewert, 2013).

1.5 Sozialmedizinische Begutachtung

In der Sozialmedizinischen Begutachtung wird die medizinische Voraussetzung des Versicherten für das Beziehen einer beantragten Sozialleistung geprüft. In der komplexen Beurteilung des Restleistungsvermögens werden die Befunde zum Gesundheitszustand, die Beeinflussbarkeit der Leistungsfähigkeit durch eine Rehabilitationsmaßnahme und Aussagen zur weiteren Prognose der Grunderkrankung miteinander verbunden. Hierzu bedarf es eines gutachterlichen, sozialmedizinischen, neutralen Sachverständigen (Hesse & Gebauer, 2011). Zentraler Punkt der Begutachtung ist die Gegenüberstellung der aktuellen Leistungsfähigkeit (definiert über den Gesundheitsstatus des Versicherten) mit den Anforderungen des allgemeinen Arbeitsmarktes. Die Dimensionen der zu berücksichtigenden Parameter finden sich als Bestandteil im Bi-Modalitätskonzeptes, im Modell zur Erklärung gesundheitlicher Ungleichheit sowie in den Risiko- und Schutzfaktoren chronischer Erkrankungen wieder. Mit Etablierung der ICF wurde das komplexe Interaktionsgeschehen zwischen Funktionsfähigkeit und Kontextfaktoren nochmals theoretisch fundiert und terminologisch gesetzlich verankert. Der Sozialmedizinischen Begutachtung ist damit eine breite theoretische Basis hinterlegt.

Das sozialmedizinische Gutachten, erstellt durch den Sozialmedizinischen Dienst der DRV, dient als Grundlage für die Entscheidung über eine beantragte Sozialleistung im Reha- und Rentenverfahren. Die Begutachtung ist Ausgangspunkt für die Entscheidung über das Vorliegen einer vollen oder teilweisen Erwerbsminderung oder einer Reha-Indikation. Zur Absicherung des Risikos der Minderung der Erwerbsfähigkeit bedarf es einer sozialmedizinischen Sachaufklärung, die im Rahmen der sozialmedizinischen Begutachtung durchgeführt wird (Cibis et al., 2011, S. 78).

Die Begutachtung kann auf Grundlage der vorhandenen Aktenlage oder mit persönlicher Untersuchung des Versicherten durchgeführt werden. Dabei können begutachtende Ärzte je nach Aufbau des Regionalträgers der RV hauptamtlich interne oder auch externe Vertragsgutachter sein (Legner & Cibis, 2007). Im laufenden Antragsverfahren wird dem Gutachten des Sozialmedizinischen Dienstes der DRV eine wesentliche Entscheidungsgrundlage zugesprochen. Die endgültige Entscheidung darüber, ob und welche Leistungen erbracht werden, fällt im Anschluss des Begutachtungsprozesses jedoch ausschließlich die Verwaltung (VDR, 2001, S. 14).

Die Aufgaben der Verwaltung und die der Ärzte werden innerhalb der DRV deutlich differenziert. Die Arbeitsgegenstand der Ärzte ist die Feststellung der *Leistungsfähigkeit* des Versicherten im Erwerbsleben. Die Verwaltung stellt auf dieser Basis den Bezug zur *Erwerbsfähigkeit* des Versicherten und dessen möglichen Leistungsansprüchen her (Cibis et al., 2011, S. 80).

1.5.1 Rahmenbedingungen der Begutachtung

Sozialmedizin ist ein eigenständiges Fachgebiet der Medizin, das die multifaktoriellen Wechselwirkungen zwischen Gesundheit und Krankheit analysiert und beschreibt. Die Beschreibung wird zusätzlich durch die Betrachtung von Risikofaktoren und protektiven Faktoren ergänzt. Das Fachgebiet beschäftigt „sich wissenschaftlich und praktisch mit der Gesundheit der Bevölkerung und ihren Determinanten, der Struktur und dem Management des Gesundheitswesens und der anderen sozialen Sicherungssysteme sowie den Wirkungen und Kosten der gesundheitlichen Versorgung“ (Groner, 2012, S. 7).

Gutachter. Um als Gutachter im Rahmen der Sozialmedizinischen Begutachtung tätig zu sein, sollten Ärzte möglichst eine Weiterbildung im Fach Sozialmedizin vorweisen. Gemäß der ärztlichen Weiterbildungsordnung besteht nach einer abgeschlossenen Facharztanerkennung die Möglichkeit, diese Zusatzbezeichnung zu erwerben (Hesse & Gebauer, 2011; VDR, 2004, S. 115). Derzeit gibt es in Deutschland 9.019 Ärzte mit der Zusatzbezeichnung Sozialmedizin, von denen 5.537 in ärztlicher Tätigkeit sind (GBE-Bund, 2018a). In den meisten Fällen sind diese Ärzte auch als sozialmedizinische Gutachter im Medizinischen Dienst der Krankenversicherung, in der RV, im Ärztlichen Dienst der Bundesagentur für Arbeit, in der Versorgungsverwaltung sowie in Rehabilitationseinrichtungen tätig.

Die Aufgabe des sozialmedizinischen Gutachters ist die Erhebung des Gesundheitszustandes durch eine ausführliche Anamneseerhebung sowie einer dem Sachverhalt angemessenen und erforderlichen Diagnostik in Form einer klinischen Untersuchungen, ergänzt durch weitere apparativ-technischen und laborchemischen Erhebungsmethoden (Fritze & Fritze, 2008, S. 1-5). Zu den zur Verfügung stehenden Untersuchungsverfahren zählen:

- psychometrische und psychiatrische Untersuchungen
- Blutabnahme
- Elektrokardiogramm (EKG)
- Belastungs-EKG
- Elektroenzephalographie (EEG)
- Lungenfunktionsdiagnostik (LuFu)
- Ultraschall
- Röntgenuntersuchungen
- Computertomographie (CT)
- Magnetresonanztomographie (MRT)

In der Begutachtung sind Röntgen- und CT-Untersuchungen eigentlich nur bedingt zumutbare Verfahren. Sie sind aber als duldungspflichtig eingestuft, da sie zwar einerseits eine größere Belastung, andererseits im Einzelfall aber keine bedeutsame Gefährdung für den Versicherten darstellen (Schneider, 2011, S. 40). Insgesamt sollte eine Begutachtung den Grundsätzen der Wirtschaftlichkeit und Sparsamkeit folgen (VDR, 2004, S. 155).

Gutachterlich praktizierende Ärzte müssen ihre Tätigkeit nach bestem Wissen und Gewissen ausüben und dürfen keinen Weisungen von nicht-Ärzten folgen. Aus diesem Grundsatz resultiert die Unabhängigkeit eines Gutachter während seiner Tätigkeit (Nüchtern et al., 2015). Die Unabhängigkeit und Unparteilichkeit in der Begutachtung sind Grundvoraussetzung für die Gleichbehandlung aller Versicherten und Akzeptanz der Beurteilung (Alex, 2005). Eine valide und unabhängige Leistungsbeurteilung allein ist für die Akzeptanz des Verfahrens bei allen Akteuren (Verwaltung, Versicherter, Gerichte) jedoch nicht zwangsläufig ausreichend. Das Beurteilungsergebnis muss innerhalb des Gutachtens auch transparent und nachvollziehbar formuliert sein, um einen „Beitrag zum sozialen Frieden“ zu leisten (Cibis et al., 2011, S. 96). Ein Gutachter trifft keine Entscheidungen. Stattdessen gibt er dem Auftraggeber (Sozialversicherung, private Versicherung, Sozialgericht) sachverständige Antworten auf die Frage des ursprünglichen Gutachtauftrags. Ist der Gutachter befangen, obliegt es ihm, den Gutachtauftrag zurückgewiesen. Prinzipiell ist die ärztliche Schweigepflicht auch für einen sozialmedizinischen Gutachter gültig. Von dieser ist der durchführende Sozialmediziner bei der persönlichen Untersuchung im Rahmen der sozialmedizinischen Sachaufklärung aufgrund einer entsprechenden Erklärung des Versicherten in den Antragsformularen entbunden (VDR, 2001, S. 18). Obgleich im komplexen Geschehen der Sozialmedizinischen Begutachtung eine Gleichbehandlung aller Versicherten gewährleistet sein soll, ist „klinische Intuition“ eine traditionell menschliche Antwort auf komplexe Gesundheitsprobleme. So können gutachterliche Erfahrung und ärztliche Intuition einen wesentlichen Einfluss auf die Gesamtbeurteilung des Versichertenfalls haben. Es ist nicht auszuschließen, dass verschiedene Gutachter durch ihre persönlichen intuitionsgestützten Beurteilungen zu divergierenden Einschätzungen über die Leistungsfähigkeit eines Versicherten kommen. Folglich benötigen Gutachter neben der medizinischen Fachlichkeit auch ausgeprägte Interaktionskompetenzen und die Fähigkeit zur Selbstreflexion. Diese Softskills helfen dem Sozialmediziner zu entscheiden, ob die geschilderten Beschwerden mit den befundeten Funktionseinschränkungen übereinstimmen (Körner, 2005). Ferner ist der persönliche Begutachtungsprozess immer durch die Interaktion zweier Individuen gekennzeichnet, welcher zumeist subjektiven Einflüssen wie Vorannahmen, Werthaltungen, Rollenerwartungen, Deutungsmuster und Übertragungsprozessen unterliegt (Cibis et al., 2011, S. 98; Harding, 2004; Simon, 2004).

Um die Interaktion zwischen Versichertem und Gutachter weitestgehend zu objektivieren, sollten daher die von der DRV publizierten indikationsspezifischen Leitlinien zur sozialmedizinischen Beurteilung der Leistungsfähigkeit bei der Begutachtung herangezogen werden. Die Nutzung dieser Kriterien fördert die Gleichbehandlung aller Versicherten und die korrekte Beurteilung des Einzelfalls (Cibis et al., 2011, S. 99).

Antragsteller. Zur Gewährung einer EM-Rente ist ein Antragsverfahren notwendig, welches aktiv vom Versicherten eingeleitet wird. Mit der Beantragung müssen sich Versicherte nach § 62 SGB I auf Verlangen des zuständigen RV-Trägers einer ärztlichen oder psychiatrischen Untersuchung unterziehen, wenn dies der Entscheidungsfindung über die bestragte Leistung dienlich ist (Cibis et al., 2011, S. 99). Während der Sozialmedizinischen Begutachtung obliegt der Antragsteller einer gewissen Mitwirkungspflicht nach §§ 60 bis 67 SGB I. Wird dieser Mitwirkungspflicht nicht nachgekommen, sodass die Aufklärung des EM-Rentenanliegens nicht vollständig abgeschlossen werden kann, ist die beantragte Leistung seitens der RV bis auf Weiteres zu versagen (VDR, 2001, S. 20).

Für Versicherte hat die Begutachtungssituation eine besondere Bedeutung. Diese warten längere Zeit auf ihren i. d. R. einmaligen Begutachtungstermin bei der RV und müssen ihre körperlichen oder psychischen Schädigungen in der Begutachtungssituation ad hoc deutlich beschreiben. Der Gutachter wird die geschilderten Symptome als gegeben betrachten und in der persönlichen Untersuchung deren Vorhandensein oder Nichtvorhandensein beweisen (Fritze & Fritze, 2008, S. 5). Wie im Kapitel 1.3.1 beschrieben, werden Anträge auf EM-Rente immer häufiger mit psychischen Erkrankungen begründet. Für Gutachter resultiert daraus das Erfordernis, immer häufiger nicht direkt beobachtbare und nur annäherungsweise objektivierbare innere Vorgänge beurteilen zu müssen. Entsprechend besteht bei einem EM-Rentenantrag auf Basis einer psychischen Erkrankung verstärkt die Möglichkeit, durch nichtauthentische Kommunikation seitens des Versicherten eine modifizierte Selbstdarstellungen vorzutragen, um die beantragte EM-Rente zu erhalten (Heubrock & Petermann, 1998; Schmidt, Lanquillon & Ullmann, 2011; Walter, Geissler, Petermann & Kobelt, 2015; Walter, Petermann & Kobelt, 2012). Neben einem verantwortungsvollen Umgang und Verständnis für Ängste und Sorgen der Versicherten sollte der beurteilende Gutachter dennoch eine Beschwerdevalidierung im Rahmen der klinischen Untersuchung und mittels eines Beschwerdevalidierungstests in Betracht ziehen (Walter et al., 2012). In diesen Situationen ist es wichtig, dem zu begutachtenden Versicherten vorurteilsfrei gegenüberzustehen und das anschließende Gutachten neutral zu verfassen, da nach § 25 SGB X jeder Versicherte ein Recht auf Akteneinsicht hat und das Gutachten einsehen kann (VDR, 2001, S. 17).

1.5.2 Das ärztliche Gutachten

Regelhaft dauern das persönliche Gespräch sowie die körperliche Untersuchung des Versicherten insgesamt etwa 60 Minuten. Rentenbegutachtungen wegen psychischer Störung können länger ausfallen. Im Anschluss oder bereits während der Untersuchung wird das sozialmedizinische Gutachten durch den begutachtenden Sachverständigen diktiert (Hesse & Gebauer, 2011). Im sozialmedizinischen Gutachten werden anhand der Anamnese, Befunde, Diagnosen und Epikrise das positive und negative Leistungsbild sowie das Leistungsvermögen (qualitativ und quantitativ) des Versicherten abgeleitet. Die DRV veröffentlichte im Jahr 2001 den Leitfaden „Das ärztliche Gutachten für die gesetzliche Rentenversicherung“ (VDR, 2001), in dem die Gliederung und das Anforderungsprofil eines Gutachtens definiert wurden (Tab. 8). Durch einheitliche Rahmenbedingungen sowie eines einheitlichen Formularsatzes (vgl. *Anhang 1*) sollte die Transparenz und Vergleichbarkeit aller sozialmedizinischen Gutachten optimiert werden (Cibis et al., 2011, S. 89). Aufgabe des Sozialmediziners ist es, das Gutachten anhand der Gliederungsvorgaben übersichtlich aufzubauen. Die im Gutachten verwendete Ausdrucksweise muss klar und auch für nicht-Mediziner verständlich abgefasst sein. Darüber hinaus sind stereotype Formulierungen sowie die unkritische Nutzung von Textbausteinen zu vermeiden, um dem individuellen fallspezifischen Gutachten gerecht zu werden (VDR, 2001, S. 22).

Tabelle 8: Gliederung und Anforderungsprofil des ärztlichen Gutachtens für die gesetzliche Rentenversicherung (modifiziert nach VDR, 2001, S. 21).

Gliederungspunkte	Verbindliche Gliederungsabschnitte
Anamnese	• Anamnese medizinischer Daten • Familienanamnese • Eigenanamnese • Jetzige Beschwerden • Therapie • Behandelnde Ärzte • Biographische Anamnese (nur bei psychiatrischen Gutachten) • Arbeits- und Sozialanamnese • Berufs- und Arbeitsanamnese • Umstände der Antragstellung, bisherige Sozialleistungen • Familiäre Situation
Untersuchungsbefunde	• Klinische Untersuchungsbefunde • Medizinisch-technische Untersuchung
Diagnosen	
Epikrise	
Sozialmedizinische Leistungsbeurteilung	• Individuelles positives und negatives Leistungsbild • Letzte berufliche Tätigkeit • Leistungsfall • Zeitliche Befristung der EM-Rente (Zeitrente)

Anamnese. Die Anamnese ist zentraler Bestandteil des medizinischen Gutachtens (VDR, 2001, S. 22). Sie sollte immer vor dem Hintergrund der Fragestellung des Gutachtens durchgeführt werden. So werden bei Anträgen auf EM-Rente insbesondere das Restleistungsvermögen im Erwerbsleben und prognostische Aspekte hinsichtlich der Dauer einer Erwerbsminderung geprüft. Nur durch eine ausführliche Berufs- und Sozialanamnese kann der medizinische Sachverständige diese Aspekte sozialmedizinisch fundiert in die Beurteilung einfließen lassen. Festgestellte physische oder psychische Störungen können in Beziehung zur beruflichen Tätigkeit und zum sozialen Umfeld beurteilt werden. Unter Einbezug der ICF (vgl. Kap. 1.4) lässt sich die Anamnese unter Zuhilfenahme des biopsychosozialen Modells strukturiert erheben. Dadurch können bereits an dieser Stelle der Begutachtung Funktions-, Aktivitäts- und Teilhabestörungen identifiziert werden, die der Gutachter im weiteren Begutachtungsverlauf evaluieren kann (Cibis et al., 2011, S. 90).

Die biographische Anamnese ist ausschließlich bei psychiatrischen Gutachten obligat und umfasst „Besonderheiten der frühkindlichen Entwicklung, der sozialen Herkunft, des späteren Lebensweges im Hinblick auf Familie, Partnerschaft und Beruf, so dass Zusammenhänge von Lebensgeschichte und Erkrankungsbeginn, -art und -verlauf zu erkennen sind“ (VDR, 2001, S. 24).

Untersuchungsbefunde. In der klinischen und medizinisch-technischen Untersuchung werden die Körperfunktionen untersucht und beschrieben. Bei der Befundung müssen sowohl ältere Vorbefunde als auch aktuelle Informationen aus der Anamnese berücksichtigt und erhoben werden. Befundete Funktionseinschränkungen werden immer mit Angabe von Maßangaben und Referenzwerten dargestellt (VDR, 2001, S. 24). Die Untersuchung bildet die Grundlage der sozialmedizinischen Beurteilung des Leistungsvermögens im Erwerbsleben. Die ICF bietet auch an dieser Stelle der Begutachtung wieder die Möglichkeit zur Systematisierung. Potenzielle Funktionseinschränkungen können durch Einbezug der ICF den einzelnen Komponenten des Modells zugeordnet werden. Durch Betrachtung der Wechselwirkungen kann die funktionelle Problematik des Versicherten ganzheitlich erläutert werden (Cibis et al., 2011, S. 92).

Diagnosen. Im Gutachten sind Diagnosen mit ihren korrespondierenden ICD Codes, nach Schweregrad geordnet, gesondert als Funktionsdiagnosen aufzulisten. Das sind medizinische Diagnosen, die um die Beschreibung des zugehörigen Funktionszustandes ergänzt wurden (z. B. LWS-Syndrom mit geringer Funktionseinbuße ohne Zeichen einer Nervenwurzelentzündung) (DRV, 2009b, S. 43). Die drei sozialmedizinisch relevantesten Diagnosen müssen nochmals auf dem Schlussblatt des einheitlichen Formularsatzes aufgenommen werden (VDR, 2001, S. 26). Für mehr Verständlichkeit innerhalb eines Gutachtens sollen auch die Funktionsdiagnosen für Nichtmediziner vom begutachtenden Sozialmediziner verständlich formuliert werden (Cibis et al., 2011, S. 93).

Epikrise. Die Epikrise bezeichnet die zusammenfassende Darstellung der vorliegenden Erkrankungen des Versicherten inklusive der damit assoziierten klinischen Auswirkungen. Sie bildet die Grundlage für die sozialmedizinische Leistungsbeurteilung. Die Epikrise umfasst die Beschreibung des Krankheitsverlaufes unter der kritischen Würdigung der Anamnese, Vorbefunde, erhobener eigener Befunde, der bisherigen Therapie und mögliche therapeutischer und rehabilitativer Optionen. In ihr werden auch Diskrepanzen zwischen erhobenen Befunden und geäußerten Beschwerden sowie zwischen Vorbefunden und vom Gutachter selbst erhobenen Befunden beschrieben. Abschließend trifft der Gutachter Aussagen über die Prognose und zum weiteren Verlauf der Erkrankung unter Beachtung der bisherigen und zukünftigen Therapieoptionen, einschließlich medizinischen Rehabilitationsmaßnahmen und Leistungen zur Teilhabe am Arbeitsleben (VDR, 2001, S. 26).

Wurden Anamnese und Untersuchungsbefunde bereits Unterzuhilfenahme der ICF systematisch erhoben, können die einzelnen Komponenten des Modells in der Epikrise miteinander verknüpft werden, um den funktionalen Gesundheitsstatus zu beschreiben. Die Nutzung der ICF Terminologie erlaubt in der Epikrise die systematische „Ableitung der jeweiligen krankheitsbezogenen Befunde auf die Ebenen der Körperstrukturen und -funktionen sowie Aktivitäten (Teilhabe) unter Berücksichtigung der gesundheitlich relevanten Kontextfaktoren“ (Cibis et al., 2011, S. 94).

Sozialmedizinische Leistungsbeurteilung. Die abschließende sozialmedizinische Leistungsbeurteilung beschreibt das aktuelle Fähigkeitsprofil des Versicherten in Bezug auf die Anforderungen der letzten Tätigkeit und des allgemeinen Arbeitsmarktes. Die Leistungsbeurteilung muss im laufenden Fließtext des Gutachtens und auf dem Gutachtenformularsatzes (Schlussblatt) erläutert werden (VDR, 2001, S. 27).

Die Leistungsbeurteilung steht am Ende der gutachterlichen Argumentationskette. Ausgehend von der Fragestellung des Gutachtens („Ist eine EM-Rente indiziert?“) leitet sich die Sozialmedizinische Leistungsbeurteilung aus der nachvollziehbaren und transparenten Verknüpfung der Informationssammlung und -bewertung ab (Abb. 6)

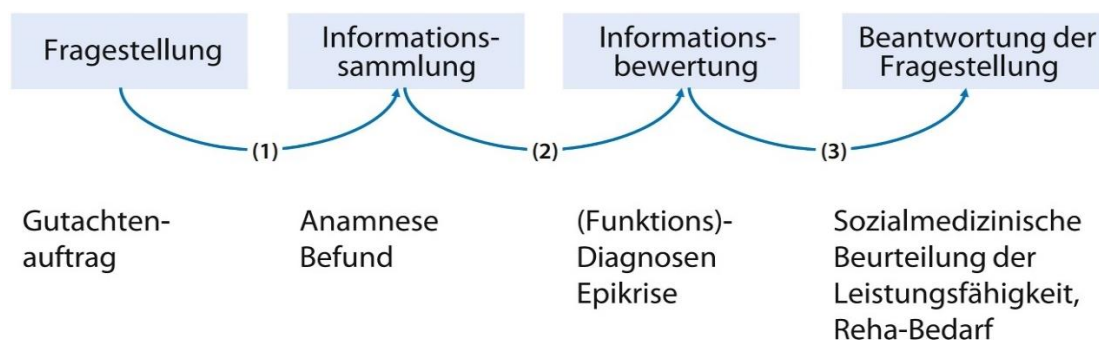


Abbildung 6: Gutachterliche Argumentationskette innerhalb eines sozialmedizinischen Gutachtens (entnommen aus Cibis et al., 2011, S. 111).

Inhaltlich umfasst die Leistungsbeurteilung das qualitative, individuelle positive und negative Leistungsbild sowie das sich daraus ableitende quantitative Leistungsbild (Cibis et al., 2011, S. 94).

Das qualitative Leistungsbild sollte sich direkt aus den in der Epikrise beschriebenen Beeinträchtigungen von Funktionen und Aktivitäten ableiten lassen. Es beschreibt die zumutbare körperliche Arbeitsschwere, Arbeitshaltung und Arbeitsorganisation, die ein Versicherter aufweist (positives Leistungsvermögen) und welche Fähigkeiten krankheitsbedingt nicht mehr ausgeübt werden können (negatives Leistungsvermögen). Hinsichtlich des negativen Leistungsbildes sind nur dann Angaben im Gutachten zu machen, wenn Einschränkungen bezüglich der geistigen/psychischen Belastbarkeit, Sinnesorgane, Bewegungs- und Haltungsapparat sowie Gefährdungs- und Belastungsfaktoren festgestellt wurden (VDR, 2001, S. 37).

In der anschließenden quantitativen Beurteilung erfolgt die Bewertung des Restleistungsvermögens in den Kategorien „über 6 Stunden“, „3 bis unter 6 Stunden“ oder „unter 3 Stunden“ hinsichtlich der letzten Tätigkeit und auf dem allgemeinen Arbeitsmarkt. Qualitative Einschränkungen haben nur dann einen Einfluss auf die quantitative Leistungsfähigkeit, wenn günstig wirkenden Kompensationsmöglichkeiten auf Ebene der Kontextfaktoren im Sinne des ICF-Modells fehlen. Es liegen folglich keine Leistungseinschränkungen vor, wenn zentrale Funktions- und/oder Aktivitätsstörungen kompensiert werden können (Cibis et al., 2011, S. 95).

1.5.3 Evidenz in der Sozialmedizinischen Begutachtung

Die medizinischen Standards der Begutachtung schützen nicht unvermeidlich vor subjektiven Verzerrungen bei der Bestimmung des Leistungsvermögens (vgl. Kap. 1.5.1). Im Einzelfall müssen Gutachter auch Sachverhalte beurteilen, die kein standardisiertes und objektives Korrelat aufweisen. In diesen Fällen darf die subjektive, auf Erfahrungen beruhende Meinung des Gutachters ausschlaggebend für die Leistungsbeurteilung sein (Cibis et al., 2011, S. 100). Anhand solcher komplexer Gutachtenkonstellationen stellt sich die Frage, wie jenseits der sozialmedizinischen Leitlinien während der Begutachtung evidenzbasierte Entscheidungen getroffen werden können, um eine Gleichbehandlung aller Versicherten zu gewährleisten.

Evidenz-basierte Medizin (EbM) beschreibt den gewissenhaften, ausdrücklichen und vernünftigen Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung für individuelle Fallkonstellationen. Hierbei werden individuelle klinische Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung kombiniert (Sackett, Rosenberg, Gray, Haynes & Richardson, 1996). Die Anwendung von EbM soll den Gebrauch nicht-wirksamer und schädigender

klinischer Maßnahmen reduzieren, wirksame Maßnahmen zur Maximierung der Lebensqualität und -quantität fördern und damit insgesamt zu verbesserten medizinischen Entscheidungsprozessen führen (Sackett, 1997). In der Sozialmedizinischen Begutachtung wurde mit dem Abschlussbericht der SOMEKO seit 2004 die allgemeine Debatte über Evidenzbasierung und Qualität im Gesundheitswesen auch in der Sozialmedizin aufgegriffen und diskutiert. In den folgenden Jahren wurde daraufhin die Entwicklung von Begutachtungsleitlinien im Bereich der Rehabilitation und der EM-Renten stetig forciert (VDR, 2004). An der Erstellung der Leitlinien sind Ärzte verschiedener Fachrichtungen der DRV unter Einbeziehung der fachlichen Kompetenz leitender Ärzte im Rahmen eines Überarbeitungs- und Konsensfindungsprozesses beteiligt. Die Leitlinien sind Beurteilungshilfen zu sozialmedizinischen Fragestellungen und dienen als Hilfe zur Durchführung fundierter sozialmedizinischer Begutachtungen nach einheitlichen und nachvollziehbaren Maßstäben. Sie enthalten Aussagen zur Beurteilung der Leistungsfähigkeit im Erwerbsleben (als Ausgangspunkt für die anschließend zu treffenden Festlegungen zur verminderten Erwerbsfähigkeit bzw. Erwerbsminderung) oder/und zur Rehabilitationsbedürftigkeit (Ueberschär, 2008). Eine weitere Forderung des SOMEKO-Berichts stellt die wissenschaftliche Einschätzung der aktuellen Sozialmedizinischen Begutachtung einschließlich der dabei verwendeten diagnostischen Instrumente dar.

Systematische Literaturrecherche zur sozialmedizinischen Begutachtung. Ein erster Schritt zur Einschätzung der Evidenz der Sozialmedizinischen Begutachtung wurde im Jahr 2014 im Rahmen des DRV Bund geförderten Forschungsprojektes „Systematische Literaturrecherche zur Sozialmedizinischen Begutachtung bei EM-Rente und Reha-Zugang (LitBegut)“ unternommen. Bis dato existierte keine systematische Literaturübersicht zur Sozialmedizinischen Begutachtung bei Anträgen auf Reha-Leistungen bzw. Anträgen auf EM-Rente. Für die wissenschaftliche Fundierung und Weiterentwicklung der sozialmedizinischen Tätigkeiten und insbesondere auch für die regelmäßige Fortentwicklung der Sozialmedizinischen Begutachtungsempfehlungen und Leitlinien ist es jedoch erforderlich, auf entsprechende evidenzbasierte Informationen zurückgreifen zu können (Strahl & Vogel, 2016, S. 3).

Die Recherche umfasste nationale und internationale Publikationen in deutscher und englischer Sprache. Eingeschlossen wurden Artikel, in denen begutachtungsrelevante Aspekte mit Bezug auf die Beurteilung der Leistungsfähigkeit, den Zugang zur Rehabilitation oder den Zugang zur EM-Rente untersucht wurden. Die Literaturrecherche erfolgte in den Datenbanken PUBMED/MEDLINE, EMBASE, PsycINFO, Psynex, Cochrane Database of Systematic Reviews und Cochrane Central Register of Controlled Trials für den Zeitraum zwischen 1993 und 2015 (Strahl & Vogel, 2016, S. 8).

Es wurden die Suchtermini „Sozialmedizinische Begutachtung“, „Funktionsfähigkeit“, „Erwerbsminderungsrente“ und „Rehabilitationszugang“ für die Recherche herangezogen. Da infolge der Komplexität des Begutachtungsgeschehens nur wenig Literatur mit höherer Evidenzstufe vermutet wurde, ist die Recherche mit einer sensitiven Suchstrategie durchgeführt worden. Somit wurden sowohl Original- und systematische Übersichtsarbeiten als auch narrative Reviews in einbezogen wurden. Von insgesamt 2.874 gefundenen Dokumenten, verblieben nach Titel-, Abstract- und ggf. Volltextsichtung 52 Dokumente (Strahl & Vogel, 2016, S. 14). Zur Bewertung der Evidenz wurde das für systematische Reviews verbreitete Bewertungsschema nach Shekelle et al. (Shekelle, Woolf, Eccles & Grimshaw, 1999) verwendet, welches eine Einteilung der Literatur in die Evidenzgrade Ia bis IV erlaubt. Hierunter versteht man die hierarchische Anordnung von Studientypen entsprechend ihrer methodischen Charakteristika zur Beurteilung der Aussagekraft (Ia: höchstes Evidenzlevel; IV: niedrigstes Evidenzlevel). Die Bewertung von Leitlinien folgte der Stufenklassifikation der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) von der Kategorie S1 (niedrigstes Evidenzlevel: Handlungsempfehlung von Experten) bis S3 (höchstes Evidenzlevel: evidenz- und konsensbasierte Leitlinien) (Kopp, 2010).

Die 52 relevanten Publikationen lassen sich thematisch in **drei Kategorien** einteilen:

- A) Assessmentinstrumente zur Feststellung der Funktionsfähigkeit sowie des Reha- und Rentenbedarfs (n = 18),
- B) Allgemeine Vorgehensweisen und Hinweise zur Begutachtungspraxis (n = 22),
- C) Leitlinien, die in Fachzeitschriften publiziert wurden (n = 12).

Die Publikationen aus der **Kategorie A** befassen sich mit der Entwicklung und/oder Evaluation von Assessmentinstrumenten zur Einschätzung der Funktions- und Arbeitsfähigkeit, des Unterstützungs- und/oder des Rehabilitationsbedarfs. In einem systematischen Review werden die psychometrischen Eigenschaften von Assessmentinstrumenten zur Bewertung der funktionellen Einschränkungen von Menschen, die eine EM-Rente beantragen, untersucht. Die Studien wurden überwiegend für spezifische Indikationen (n = 13), zum Teil aber auch indikationsübergreifend (n = 5) konzipiert. 15 der insgesamt 18 Publikationen zeigten aufgrund des Studiendesigns die Evidenzklasse III (n = 12) oder IV (n = 3) und weisen damit eine eher niedriges Evidenzniveau auf. Darüber hinaus wurden eine quasi-experimentelle Studie, eine randomisiert kontrollierte Studie (RCT) und ein systematisches Review identifiziert.

Die Publikationen der Kategorie **B zeigen** hinsichtlich ihres Themenspektrums eine erhebliche Heterogenität. Die gefundenen Studien lassen sich in die vier Themenbereiche (1) allgemeine oder indikationsspezifische Hinweise zur Begutachtung und zum

Begutachtungsgeschehen, (2) Implementierung der Internationalen Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) in den Begutachtungsprozess, (3) Identifikation von Reha-Bedarf und (4) Validität und Reliabilität von Gutachten und Gutachtenergebnissen einteilen. Insgesamt 21 der 22 Publikationen dieser Kategorie wurden aufgrund des Studiendesigns der Evidenzklasse III (n = 4) oder IV (n = 17) zugeordnet. Lediglich eine quasi-experimentelle Studie (IIb) konnte identifiziert werden.

Kategorie C umfasst die sozialmedizinischen Leitlinien. Aufgrund der methodischen Umsetzung der Recherche wurden ausschließlich in Fachzeitschriften publizierte Leitlinien identifiziert. Es konnten Begutachtungshinweise zu den Krankheitsbildern koronare Herzkrankheit, Bandscheiben und bandscheibenassoziierte Erkrankungen, COPD, CED, Mammakarzinom, dermatologische Erkrankungen, chronische Schmerzen und chronische nicht maligne Leber- und Gallenwegserkrankungen identifiziert werden. Fast alle Leitlinien wurden als S1-Leitlinien, d.h. basierend auf Expertenkonsens, entwickelt. Lediglich die Leitlinie für die Begutachtung von Schmerzen, die nicht speziell für die Sozialmedizinischen Begutachtung erstellt wurde, erzielt den Evidenzgrad IV, da es sich bei dieser Publikation um ein narratives Review einer anderenorts veröffentlichten Begutachtungsleitlinien handelt.

Die Untersuchung zeigt, dass die Evidenzbasierung in der Sozialmedizinischen Begutachtungsliteratur häufig auf Expertenkonsens „limitiert“ ist. RCTs wurden selten identifiziert. Ein hoher Anteil der eingeschlossenen Studien sind narrative Reviews oder Beobachtungsstudien mit einem niedrigen Evidenzniveau. Eine Besonderheit im komplexen Prozess der Sozialmedizinischen Begutachtung ist, dass höhere Evidenzlevel nicht in allen Themenbereichen (v.a. zur Begutachtungspraxis) erreichbar sind. Im wissenschaftlichen Kontext der Grundlagen der Sozialmedizinischen Begutachtung besteht zwar ein erhebliches Entwicklungspotential, im Allgemeinen ist der pragmatische Übertrag von Erkenntnissen der EbM auf die Sozialmedizinische Begutachtung jedoch zu hinterfragen. Häufig basieren evidenzbasierte Empfehlungen auf Grundlage von Studien, die in anderen Gesundheitssystemen und unter anderen Rahmenbedingungen als der Begutachtungssituation durchgeführt wurden. Darüber hinaus werden für höherwertige RCTs verbindliche Ein- und Ausschlusskriterien definiert, die das multimorbide Patientenkollektiv der Sozialmedizinischen Begutachtung nur selten berücksichtigen (Cibis et al., 2011, S. 108). Als Limitation sei zusätzlich zu erwähnen, dass aufgrund der sensitiven Suche eine hohe Zahl an falsch-positiven Treffern angezeigt wurde. Es ist nicht auszuschließen, dass bestimmte indikationsspezifische Literaturstellen unentdeckt blieben, da eine primär indikationsunspezifische Suche durchgeführt wurde (vgl. Strahl, Rose, Brüggemann & Vogel, 2016a, 2016b; Strahl & Vogel, 2016).

1.6 Qualitätssicherung in der Sozialmedizinischen Begutachtung

Die Sozialmedizinische Beurteilung ist, wie bereits dargelegt, ein komplexer und individueller Bewertungsprozess, der durch verschiedene Störfaktoren einer gewissen Beurteilungsvarianz unterliegt. In einer Studie mit 22 Psychiatern konnte die Varianz in der Beurteilung exemplarisch dargestellt werden. Anhand einer standardisierten Kasuistik (bestehend aus Anamnese, Befund, Entlassungsdiagnosen einer psychiatrischen Fachklinik, Behandlungsdiagnosen eines Psychiaters und Hausarztes sowie einer biografischen Anamnese und eines Video-Interviews) wurden alle Teilnehmer gebeten eine sozialmedizinische Leistungsbeurteilung durchzuführen. Sowohl die vergebenen Hauptdiagnosen als auch die Leistungsbeurteilung zeigten eine hohe Heterogenität (<3 Std.: 27 % / 3 bis 6 Std.: 36 % / >6 Std.: 37 %). In acht Fällen wäre ein EM-Rentantrag voraussichtlich abgelehnt worden, in vier Fällen wäre eine volle EM-Rente und 10 Fällen teilweise EM-Rente zugewiesen worden (Dickmann & Broocks, 2007). Auch weitere Studien legen dar, dass die Übereinstimmung von Experten bezüglich der Beurteilung der beruflichen Leistungsfähigkeit häufig limitiert ist (Clark et al., 1988; Elder, Symington & Symington, 1994; Ingravallo et al., 2008).

In einem aktuellen systematischen Literaturreview aus dem Jahr 2017 von Barth und Kollegen (2017) wurden 23 Studien zur Übereinstimmung von medizinischem Sachverständigen bei der Beurteilung von Leistungsansprüchen (z. B. Rente) bei Arbeitsunfähigkeit untersucht. Die identifizierten Studien wurden sowohl im Versicherungs- (n = 16 Studien) als auch im Forschungskontext (n = 7 Studien) erstellt. Die Reliabilität im Versicherungskontext lag im durchschnittlich bei 0,45 (Median; Range 0,86 bis -0,10) und war damit bedeutend niedriger als die Reliabilität von Studien, die im Forschungsumfeld durchgeführt wurden. Hier lag die mittlere Interrater-Reliabilität bei 0,76 (Range 0,91 bis 0,53). In den meisten Studien erreichten medizinische Experten bei der Beurteilung der Arbeitsfähigkeit nur eine geringe bis mittlere Übereinstimmung hinsichtlich der zu beurteilenden Outcome-Parameter, wie funktionaler Arbeitsfähigkeit, beruflicher Funktionsfähigkeit, Arbeitszeitverkürzung, Frühpensionierung oder EM-Rente. Die aktuelle Forschung über die Zuverlässigkeit der gutachterlichen Beurteilung der Arbeitsunfähigkeit ist derzeit begrenzt, deutet jedoch auf erhebliche Unterschiede zwischen den medizinischen Sachverständigen hin (Barth et al., 2017).

Linden beschrieb bereits im Jahr 2007, dass es kaum wissenschaftliche Grundlagen der Entscheidungsfindung und -begründung gebe. Der Grund dafür sei, dass für jede gutachterliche Entscheidung, die in einem multifaktoriellen Bedingungsraum des Versicherten getroffen wird, eigentlich keine uneindeutigen Schlüsse zulässig seien. Er

vergleicht die sozialmedizinische Schlussfolgerung mit einem Richterurteil, bei dem gilt: „Auf offener See und vor Gericht ist man in Gottes Hand“ (Linden, 2007, S. 379).

Darüber hinaus ist die Beurteilung der Erwerbsunfähigkeit ein Prognoseurteil. Die Gutachteraussagen über die Prognose und zum weiteren Verlauf der Erkrankung sind Wahrscheinlichkeitsaussagen, die je nach Fallkonstellation stark variieren können. Dementsprechend würde auch der verstärkte Einsatz von reliablen Fremdbeurteilungsskalen in der Begutachtung nur bedingt hilfreich sein, da ein Punktwert meist nur einen Informationsaspekt der Untersuchung darstellt und kein gutachterliches Urteil begründet (Linden, 2007). Das Kontinuum zwischen der auf der einen Seite komplexen Begutachtung und auf der anderen Seite Forderung zur Gleichbehandlung aller Versicherten veranschaulicht, wie zentral die Implementierung einer einheitlichen, systematischen und trägerübergreifenden Qualitätssicherung ist (Cibis et al., 2011, S. 108).

1.6.1 Definition und Begriffsbestimmung

Das aus dem lateinischen Sprachraum stammende Wort „qualitas“ bedeutet Beschaffenheit oder Eigenschaft eines Gegenstandes. Seinen Ausgangspunkt hat Qualitätssicherung im Produktions- und Dienstleistungsbereich. Qualität beschreibt hier die Güte eines Gegenstandes oder einer Dienstleistung. Allen Gütern und Dienstleistungen ist eine festgelegte Anforderung bzw. eine zu erreichende Norm hinterlegt (Wirtz, 2018a).

Qualität im Gesundheitswesen kann auf den unterschiedlichen Ebenen Ergebnis, Prozess und Struktur gemessen werden. Die Ergebnisqualität beschreibt die Effizienz bzw. die Zweckmäßigkeit eines Gegenstandes, die Prozessqualität alle Maßnahmen, die im Verlauf von Produktions- oder Dienstleistungsprozessen durchgeführt werden und die Strukturqualität die Ausstattung (materiell/immateriell) oder den Kontext eines Gegenstandes oder Arbeitsplatzes (vgl. Donabedian, 2005). Infolge der a priori definierten Anforderungen, kann deren Güte in einem Qualitätssicherungsverfahren überprüft werden.

Qualitätsmanagement und -sicherung. Die Grundlagen, Terminologien und Anforderungen von Qualitätsmanagementsystemen sind in den DIN EN ISO 9000ff-Normen definiert (Cibis et al., 2011, S. 101). Der gesamte Prozess der Qualitätsprüfung wird durch das Qualitätsmanagement zentral gesteuert. Darunter werden alle Handlungen und Maßnahmen verstanden, die zur Wahrung der Qualität beitragen. Das oberste Ziel besteht in der Sicherstellung von Struktur- und Prozessqualitätsmerkmalen, welche direkt mit einer hohen Ergebnisqualität vergesellschaftet sind. Einige der zentralen Grundmerkmale sind Kunden- und Patientenorientierung, Mitarbeiterpartizipation, kontinuierliche Verbesserung sowie sachliche, daten- und informationsbasierte Entscheidungsfindung. Qualitätsmanagement sollte immer an der höchsten

Führungshierarchie angegliedert sein, um im Rahmen von „top-down Prozessen“ die Qualitätskultur und -weiterentwicklung sowie Maßnahmen zur Qualitätssicherung im Betrieb umzusetzen (Wirtz, 2018b).

Nach Definition der Deutschen Instituts für Normen ist Qualitätssicherung ein Teil des Qualitätsmanagements. Wurde vor 1994 unter Qualitätssicherung noch die Gesamtheit des Qualitätsmanagements verstanden, so ist sie nach internationaler Vereinbarung heute nur noch ein Teil davon (Cibis et al., 2011, S. 101). Kernelemente der Qualitätssicherung sind die systematische Erfassung und Prüfung von Qualitätsindikatoren, um potenzielle Defizite zu identifizieren und diese anschließend durch qualitätsoptimierende Maßnahmen und zu beseitigen (Wirtz, 2018c). Qualitätssicherung und Qualitätsmanagement greifen direkt ineinander und sind zwei voneinander untrennbare Konzepte. Da Qualität nicht direkt gemessen werden kann, muss das zu untersuchende Konstrukt durch geeignete Indikatoren operationalisiert werden. Diese stellvertretenden Merkmale werden im allgemeinen als Qualitätskriterien bezeichnet (Blumenstock, 2011).

Externe Qualitätssicherung. Neben dem Bestreben, durch interne Entwicklungsprozesse die Qualität sukzessiv zu verbessern, kann Qualität auch durch einen externen Vergleich optimiert werden. Hierbei werden verschiedenen Leistungen oder Einrichtungen miteinander verglichen. Im Gesundheitswesen wurde in Deutschland zunächst in den 1970er Jahren ein freiwilliges externes Qualitätssicherungsverfahren entwickelt. Ab den 1990er Jahren wurden dann Krankenhäuser gesetzlich verpflichtet, an externen Qualitätssicherungsprogrammen teilzunehmen. Mit Inkrafttreten von § 135a SGB V (Verpflichtung der Leistungserbringer zur Qualitätssicherung) besteht seit dem Jahr 2000 für alle Leistungserbringer (Vertragsärzte, medizinische Versorgungszentren, Krankenhäuser, Erbringer von Vorsorgeleistungen und Rehabilitationsmaßnahmen) eine Pflicht zur Teilnahme an externen Qualitätssicherungsprogrammen. Um die externe Qualitätssicherung neutral durchzuführen, wurde im Jahr 2007 mit Inkrafttreten von § 137a SGB V (Institut für Qualitätssicherung und Transparenz im Gesundheitswesen) ein fachlich unabhängiges, wissenschaftliches Institut für Qualitätssicherung und Transparenz gegründet. Das IQTIG-Institut wurde damit beauftragt, die ambulante und stationäre einrichtungsübergreifende Qualität zu messen und darzustellen. Um das medizinische Versorgungsgeschehen realitätsnah abzubilden, sollte mit dieser Maßnahme die Qualitätssicherung möglichst von einer sektorspezifischen (stationären) zu einer sektorenübergreifenden Qualitätssicherung umgebaut werden (Stengel, Bauwens & Ekkernkamp, 2007; Willms, Bramesfeld, Pottkämper, Broge & Szecsenyi, 2013).

Bei der externen Qualitätssicherung werden Qualitätskriterien verschiedener Einrichtungen gleichen Typs erhoben, um diese in einer Art Benchmarking miteinander zu vergleichen. Im Rahmen von (internen) Qualitätsmanagementprozessen sollen anschließend die als

defizitär identifizierten Bereiche optimiert werden. Die Ergebnisse der externen Qualitätssicherung dienen daher den einzelnen Einrichtungen als Impulsgeber zur klinikeigenen Qualitätsverbesserung (Egner, Gerwin & Buschmann-Steinhage, 2006).

1.6.2 Maßnahmen zur Qualitätssicherung in der Begutachtung

Die Forderung nach Qualitätssicherung im Gesundheitswesen wurden auch in der Sozialmedizin aufgenommen (Cibis et al., 2011, S. 105). Ausgangspunkt der strukturierten Qualitätsbemühungen war der Abschlussbericht der SOMEKO-Kommission im Jahr 2004. Da die Gesetzgebung ausschließlich medizinische Leistungserbringer zur Qualitätssicherung verpflichtete, war die Sozialmedizinische Begutachtung von diesen gesetzlichen Vorgaben nicht betroffen. Während medizinische Rehabilitationsmaßnahmen seit langem über evaluierte Qualitätssicherungsprogramme verfügten (vgl. Farin & Jäckel, 2011; Klosterhuis et al., 2010), wurden zur Umsetzungen der SOMEKO-Empfehlungen unterschiedliche Projektgruppen eingesetzt. Seit 2005 wurde ein umfassendes und einheitliches Qualitätssicherungskonzept in der Sozialmedizinischen Begutachtung durch eine dieser Projektgruppen erarbeitet (Legner & Cibis, 2007).

Bei mehr als 1.800 gutachterlich tätigen Sozialmedizinern und unter Beachtung der langfristigen finanziellen Konsequenzen für das Sozialversicherungssystem bei Gewährung einer EM-Rente, besteht ein begründeter volkswirtschaftlicher Anspruch auf ein Qualitätssicherungsverfahren im Bereich der Sozialmedizinischen Begutachtung. Die Schwierigkeit in der Umsetzung eines einheitlichen Verfahrens in der DRV liegt in deren Struktur begründet. Die Sozialmedizinische Sachaufklärung wird in jedem Regionalträger personell verschieden umgesetzt. Während einige Regionalträger bereits aus Gründen der Qualitätssicherung ausschließlich auf eigene, sozialmedizinisch trainierte Ärzte zurückgreifen, nutzen andere Regionalträger für Gutachten mit körperlicher Untersuchung teilweise oder ausschließlich externe Gutachter. Obgleich diese daraufhin sogenannten „externen Gutachten“ vor der Übergabe an die Verwaltung durch den auftraggebenden DRV Träger von hauseigenen Prüfarzten kontrolliert und korrigiert werden, kann deren Qualität stark schwanken. Ziel der Qualitätssicherung ist es, trotz der Unterschiede in den Abläufen und Strukturen, eine möglichst einheitliche Ergebnisqualität (bezogen auf die Gutachten) zu erreichen (Ueberschär, 2008). Derzeit wird die Qualität in der Sozialmedizinischen Begutachtung durch folgende Maßnahmen gewährleistet:

- DRV Schrift: „Verbindliche Hinweise zur Erstellung ärztlicher Gutachten für die gesetzliche Rentenversicherung“ (siehe Kap. 1.5.2)
- Leitlinien zur Sozialmedizinischen Begutachtung
- Qualitätszirkel

- Bundesweite sozialmedizinische Fortbildungsmaßnahmen
- Hausinterne Qualitätssicherungsprogramme
- DRV-Schrift: Sozialmedizinisches Glossar (Cibis et al., 2011, S. 105; Ueberschär, 2008).

Die Maßnahmen zur Qualitätssicherung verbessern die Standards der Begutachtung, können aber nicht garantieren, dass diese von den gutachterlich tätigen Sozialmedizinern auch angewendet werden. Vor allem niedergelassene oder extern angestellte Ärzte haben häufig nicht die Zeit, an Qualitätszirkeln oder bundesweiten sozialmedizinischen Fortbildungsmaßnahmen teilzunehmen bzw. sich stringent über die neuesten Leitlinien zu informieren. Ein trägerübergreifendes einheitliches Qualitätssicherungskonzept könnte solche Problemlagen aufdecken und Ärzte gezielt mit Schulungen unterstützen. Bisher werden Leitlinien jedoch als das wichtigste, schnellste und flächendeckendste Instrument für die Dissemination von Informationen angesehen (Cibis et al., 2011, S. 105).

Leitlinien. Leitlinien sind systematisch entwickelte, wissenschaftlich begründete und praxisorientierte Handlungsempfehlungen, die vor dem Hintergrund des gegenwärtigen Erkenntnisstandes eine angemessene ärztliche Vorgehensweise bei speziellen gesundheitlichen Problemen gewährleisten. Sie dienen sowohl dem Wissensmanagement, der Implementierung von EbM als auch der Qualitätssicherung (Cibis et al., 2011, S. 105; Selbmann & Kopp, 2006). Für den Bereich der Sozialmedizin werden Leitlinien zur sozialmedizinischen Leistungsbeurteilung bei EM-Rente und Leitlinien für den Rehabilitationszugang vorgehalten. Derzeit [Stand: Nov. 2018] befinden sich folgende Leitlinien zur Leistungsbeurteilung zum Abruf auf der Homepage der DRV (DRV, 2018d):

- Krankheiten des Atmungssystems (COPD, Asthma bronchiale)
- Gastroenterologische und Stoffwechselkrankheiten einschließlich Adipositas (CED, chronische nicht-malignen Leber- und Gallenwegserkrankungen)
- Herz-Kreislauf-Krankheiten (KHK, arterielle Gefäßkrankheiten)
- Bandscheiben- und bandscheibenassoziierten Erkrankungen
- Neubildungen (Mamma-Karzinom)
- Neurologische Krankheiten
- Psychische und Verhaltensstörungen einschließlich Abhängigkeitserkrankungen

Qualitätszirkel. Viele RV-Träger nutzen Qualitätszirkel, um im interkollegialen Diskurs über neueste Entwicklungen in der Qualitätssicherung (z. B. neue Leitlinien) zu sprechen. Darüber hinaus werden häufig in Form von kollegialer Supervision eigene Gutachten auf Plausibilität, Nachvollziehbarkeit und Neutralität geprüft (Ueberschär, 2008).

Sozialmedizinische Fortbildungsmaßnahmen. Die DRV Bund organisiert regelmäßige Fortbildungsseminare zu krankheitsfolgebezogenen Fragen der Leistungsbeurteilung für

sozialmedizinisch tätige Ärzte. Diese Angebote sind für interne und externe Sozialmediziner geöffnet. Ebenfalls stehen die Veranstaltungen für alle Berufsgruppen (Psychologen, Sozialarbeiter, Rehabilitationsfachberater, Juristen, Sozialrichter) offen, die im Bereich der Sozialmedizin eingebunden sind (Ueberschär, 2008).

Hausinterne Qualitätssicherungsprogramme. Viele RV-Träger haben eigene Qualitätssicherungsprogramme und eigene Konzepte zum Qualitätsmanagement. Im SOMEKO-Abschlussbericht wurde eine Analyse der damalig aktuellen Qualitätssicherungsbemühungen vorgelegt. Es zeigte sich, dass lediglich in 15 von 26 befragten Versicherungsanstalten Maßnahmen zur Qualitätssicherung durchgeführt wurden. Die Maßnahmen reichten von der Überprüfung von Gutachten durch wechselnd zusammengesetzte Prüfgruppen über unsystematisches prüfärztlichen Gegenlesen von Gutachten bis zur schriftlichen und telefonischen Gutachterbetreuung (VDR, 2004, S. 231).

Sozialmedizinisches Glossar. Das Glossar enthält 258 der wichtigsten sozialmedizinischen und sozialrechtlichen Begrifflichkeiten mit kurzer Erklärung. Es berücksichtigt den besonderen Bedarf der für die DRV tätigen Sozialmediziner und dient der einheitlichen Verwendung sozialmedizinischer Terminologien. Die Qualität wird durch eine verbesserte Verständlichkeit und Transparenz der Gutachten erhöht. Eine einheitliche Sprache innerhalb der Gutachten ermöglicht auch deren Vergleichbarkeit und bildet dadurch die Voraussetzung für eine trägerübergreifende Qualitätssicherung (DRV, 2009b, S. 5).

Trägerübergreifende Qualitätssicherung. Im Jahr 2007 wurde der Abschlussbericht der Projektgruppe zur Erarbeitung des einheitlichen Qualitätssicherungskonzeptes in der Sozialmedizinischen Begutachtung vorgelegt. In diesem Bericht steht die Ergebnisqualität von Gutachten als Produkte des Begutachtungsprozesses im Fokus der Qualitätssicherungsbemühungen. Das beschriebene Konzept der Projektgruppe bezieht sich vorerst ausschließlich auf Gutachten mit persönlicher Untersuchung. Langfristig sollen jedoch Qualitätssicherungskonzepte für alle Gutachten erarbeitet werden.

Der vorgeschlagenen trägerübergreifenden Qualitätssicherung liegt als Methode ein Peer Review-Verfahren zu Grunde. Hierbei werden Gutachten anhand von definierten Qualitätskriterien, die verschiedene qualitative Aspekte eines Gutachtens widerspiegeln, durch erfahrene Sozialmediziner aus den Häusern der RV-Träger bewertet (Legner & Cibis, 2007; Ueberschär, 2008).

1.6.3 Peer Review-Verfahren

Qualität im Gesundheitswesen kann auf unterschiedliche Weise überprüft werden; z. B. durch direkte Beobachtung, prospektive Datensammlungen oder die Nutzung von

Routinedaten. Neben diesen Möglichkeiten ist das Review von Fallberichten (Chart-Review) ein weit verbreitetes Verfahren der externen Qualitätssicherung im Gesundheitswesen (Lilford et al., 2007; Shaw, 2001). Im Allgemeinen beurteilt im Peer Review-Prozess ein externer, erfahrener Peer, also ein gleichgestellter Fachkollege der jeweiligen Berufsgruppe, die Qualität von Behandlungsprozessen und -ergebnissen auf Basis definierter Bewertungskriterien. In der internationalen Literatur finden sich empirische Belege für die Effektivität von Peer Review-Verfahren, wenn sie in Verbindung mit systematischen Rückmeldungen der Ergebnisse eingesetzt werden; eindeutige Erkenntnisse für die direkte Beeinflussung medizinischer Outcome-Parameter über die Qualitätssicherung hinaus fehlen jedoch bisher (Glattacker & Jäckel, 2007; Khan & Ollenschläger, 2014).

Das Verfahren wird vorrangig in der Akutmedizin eingesetzt, ist jedoch auch auf andere medizinische Bereiche übertragbar. Die Qualitätsbewertungen werden auf Grundlage von Patientenakten oder Behandlungsunterlagen abgegeben, können aber auch infolge von Klinikbegehungen und Visitationen erstellt werden (Farin et al., 2003; Jäckel & Farin, 2004). In der Literatur ist die Grenze vom Peer Review zum „Audit“ oder zur „Visitation“ fließend, so verwendet Shaw (2000) den Begriff Peer Review als übergeordneten Begriff für „Visitationen“.

Der Einsatz eines Peer Reviews als Bewertungsmethode erfordert die Überprüfung der Beurteilerübereinstimmung (Interrater-Reliabilität) als wichtigstes Gütekriterium (Goldman, 1994; Wirtz & Caspar, 2002). Zuverlässige Qualitätsvergleiche basieren auf einer präzisen, vom einzelnen Peer in höchstmöglichem Maße unabhängigen Bewertung. Mit einer hohen Reliabilität sollte die Qualitätsbeurteilung unabhängig vom beurteilenden Peer sein.

Für das Peer Review-Verfahren lassen sich drei Formen identifizieren: unstrukturiertes implizites, strukturiertes implizites sowie explizites Review (Broder et al., 2002; Hofer et al., 2004).

Unstrukturiertes implizites Review. Bei dieser Form des Peer Reviews entscheidet der Peer selbst darüber, welche Elemente des Behandlungsprozesses er als relevant für seine Qualitätsbeurteilung erachtet. Implizite Reviews basieren auf impliziten Qualitätsurteilen der Peers, die die Qualität der Behandlung auf Basis medizinischer Behandlungsberichte oder ähnlichem einschätzen. Die Reproduzierbarkeit und Interrater-Reliabilität dieser Reviews variieren stark nach Art der Review-Objekte sowie nach Erfahrung und Schulung der Peers (Broder et al., 2002, S. 2).

Strukturiertes implizites Review. Hierbei handelt es sich um eine Weiterentwicklung des unstrukturierten impliziten Reviews. Beim strukturierten impliziten Review werden Datenquellen und Prüffragen, die ein Peer beurteilen muss, standardisiert. Des Weiteren

sind für die Peer-Beurteilung Qualitätskriterien definiert, die eine Orientierungshilfe zur Bewertung darstellen. So wird der Peer konsequent auf potenzielle, wichtige Qualitätsaspekte hingewiesen, die einer Bewertung bedürfen (Hayward, McMahon & Bernard, 1993).

Explizites Review. Diese Form des Peer Reviews charakterisiert sich durch das ausschließliche Beurteilen von Schlüsselementen und nicht durch die als Qualitätssicherungsmethode im Gesundheitswesen globale Bewertung des gesamten Behandlungsprozesses bzw. -ergebnisses. Die zu beurteilenden Schlüsselemente werden a priori festgelegt und bedürfen keiner weiteren Interpretation seitens des Peers. Das explizite Review basiert dementsprechend auf externen Standards für die Einschätzung der Qualität, die es dem Peer erlauben, die Beurteilungsobjekte checklistenartig gezielt nach spezifischen Informationen zu durchsuchen (Broder et al., 2002, S. 2).

Peer Review im Versicherungswesen. Das Peer Review hat vor allem im Versicherungswesen der Vereinigten Staaten einen zentralen Stellenwert. Im akutmedizinischen Bereich ist das Review die häufigste eingesetzte Methode der Qualitätssicherung (Goldman, 1994). Versicherungsgesellschaften wie MEDICARE und MEDICAID nutzen flächendeckend das Peer Review-Verfahren, um die Qualität im Gesundheitswesen zu überprüfen. MEDICARE begann bereits 1972 mit der Gründung der Professional Standards Review Organizations (PSROs). Schließlich gingen 1982 aus den PSROs die Peer Review Organizations (PROs) hervor, welche die Qualität von 25% aller Behandlungsverläufe von MEDICARE-Versicherten zufällig mittels eines Peer Reviews evaluieren. Das Review wurde ausschließlich als Chart-Audit (Berichts-/Tabellen-Review) durchgeführt. Peers waren Ärzte mit der entsprechenden Facharztausbildung. Ab 1995 wurden die Reviews nicht mehr randomisiert vorgenommen, sondern lediglich die Fälle untersucht, in denen Versicherte Beschwerden über ihre medizinische Versorgung gemeldet hatten (vgl. Broder et al., 2002). Ab dem Jahr 2000 intensivierte MEDICARE zunehmend explizite Maßnahmen zur Erhebung der Prozessqualität. Dennoch gibt es noch immer einen nahezu universellen Einsatz von implizitem Peer Review für Krankenhäuser und verschiedene Ärzteguppen, da dies die einzige Methode zur zuverlässigen Beurteilung leitliniengerechter Standardversorgung bei Rechtsstreitigkeiten darstellt (Hofer et al., 2004).

In Deutschland ist das Peer Review-Verfahren seit 1999 routinemäßig als fester Bestandteil der externen Qualitätssicherung des klinikvergleichenden Informationssystems der Rentenversicherung in den Rehabilitationssektor integriert (Egner, Gerwin & Schliehe, 2002). Die gesetzliche Verankerung der Qualitätssicherung in der Rehabilitation förderte, mit Änderung des SGB V und durch das Inkrafttreten des § 20 SGB IX im Jahr 2002, die Teilnahme der Kliniken an externen Qualitätssicherungsprogrammen nachhaltig (Farin,

Engel, Dimou & Jäckel, 2004). Im Peer Review-Verfahren der medizinischen und psychosomatischen Rehabilitation werden anonymisierte Fallunterlagen hinsichtlich der dokumentierten Behandlungsprozesse und -ergebnisse durch erfahrene und speziell geschulte Rehabilitationsmediziner beurteilt.

Dazu werden 20 zufällig ausgewählte Reha-Entlassungsberichte der teilnehmenden Rehabilitationskliniken anhand einer Checkliste, in der 52 qualitätsrelevante Prozessmerkmale definiert sind, hinsichtlich der Ausprägung von Qualitätsmängeln beurteilt. Die Ergebnisse des Qualitätssicherungsverfahrens in der Rehabilitation, die sich im klassischen Sinne vollständig an der Struktur-, Prozess- und Ergebnisqualität orientieren (vgl. Donabedian, 2005), werden regelmäßig einrichtungsvergleichend dargestellt und den teilnehmenden Einrichtungen zurückgemeldet (Benchmarking). Das Verfahren unterstützt damit die Qualitätsorientierung, -entwicklung und das interne Qualitätsmanagement der medizinischen Rehabilitationseinrichtungen (Korsukéwitz & Irle, 2010).

Neben der DRV bediente sich vorübergehend auch die gesetzliche Krankenversicherung (GKV) des Peer Review-Verfahrens zur Qualitätssicherung der medizinischen Rehabilitation (QS-Reha-Verfahren). Nach § 111 oder § 111a SGB V ist eine Teilnahme am QS-Reha-Verfahren für alle Rehabilitationseinrichtungen mit einem Versorgungsvertrag gemäß des § 137d SGB V (Qualitätssicherung bei der ambulanten und stationären Vorsorge oder Rehabilitation) gesetzlich verpflichtend. Ähnlich wie im Qualitätssicherungsverfahren der DRV erfasst das QS-Reha-Verfahren die drei zentralen Qualitätsdimensionen: Struktur-, Prozess- und Ergebnisqualität. Das Peer Review als Bestandteil der Qualitätssicherung der Prozessqualität basierte im QS-Reha-Verfahren auf der Bewertung von medizinischen Behandlungsdokumenten durch Peers (Bassler, Nosper, Follert, Böwering & Polak, 2007). Ist eine Rehabilitationseinrichtung bereits am Qualitätssicherungsverfahren der DRV beteiligt, so ist diese nicht mehr zur Teilnahme am QS-Reha Verfahren verpflichtet. Für jede Rehabilitationseinrichtung ist nur erforderlich, an einem der beiden Verfahren teilzunehmen (Farin, Jäckel, Schalaster & AQMS, 2009). Seit 2002 lag eine einheitlich konsentrierte Version des Peer Review-Verfahrens vor, infolgedessen die Qualitätssicherungsprogramme der DRV und der GKV vereinheitlicht werden konnten. Dieser Schritt wurde bereits 1999 in einer gemeinsamen Erklärung der beiden Sozialversicherungsträger bekannt gegeben (Farin et al., 2003; Kawski & Koch, 2004). Seit 2011 wird die Prozessqualität durch die GKV nicht mehr mittels des Peer Reviews erhoben. Das Verfahren wurde zugunsten von erweiterten Einrichtungs- und Patientenfragebögen mit prozessorientierten Kriterien substituiert (Farin & Jäckel, 2011).

Reliabilität. Die meisten Studien zur Interrater-Reliabilität des Peer Reviews liegen für den Gesundheitssektor aus dem Bereich der Akutmedizin vor. Lange galt das implizite Review als Goldstandard der externen Qualitätssicherung. Im Gegensatz zum unstrukturierten

Review wurde vom strukturierten impliziten Review eine ausreichende (0,41-0,6) bis gute (0,61-0,8) Reliabilität des Kappa-Koeffizienten erwartet (Broder et al., 2002, S. 3). Jedoch wurde in mehreren Studien zur Messung der Qualität der medizinischen Versorgung lediglich eine Reliabilität in Höhe von 0,2 bis 0,3 berichtet (Hofer et al., 2004). Diese Ergebnisse bestätigen Goldmans (1994) Resultat einer Meta-Analyse zu Peer-Bewertungen der Versorgungsqualität im Gesundheitswesen. In diese Meta-Analyse wurden 21 unabhängige Reviews aus 13 Studien berücksichtigt. Im Mittel zeigte sich für das Übereinstimmungsmaß Kappa einen Reliabilitätskennwert in Höhe von 0,31.

In einer randomisierten Studie wurde die Güte von impliziten Reviews mit auf Patientenakten basierenden semi-strukturierten Interviews (Visitation) verglichen. Die Interrater-Reliabilität der impliziten Reviews ergab moderate Kennwerte in Höhe von 0,44 bis 0,56. Deutlich wurde zudem, dass im Kontext der Versorgungsqualität nicht alle Informationen mittels eines Peer Reviews erhoben werden können. Die Reviews wiesen gegenüber den Visitationen eine niedrigere Sensitivität und Spezifität auf. Die Klärung von klinischen Praxisaspekten mit Hilfe eines Interviews vor Ort war verbunden mit einer positiveren Bewertung des Reviews und weniger Problemen für die Peers im Beurteilungsprozess (Goulet, Jacques, Gagnon, Racette & Sieber, 2007).

In einer systematischen Literaturanalyse zur Qualität in der Akutmedizin wurden 26 Publikationen gesichtet, bei denen die Interrater-Reliabilität in einem Range zwischen 0,32 bis 0,7 lag (Lilford et al., 2007). Die Autoren stellten fest, dass der in den Studien genutzte Kappa-Koeffizient für explizite Reviews höher ausfällt als bei impliziten Reviews (vgl. Broder, 2002). Zudem wurde in der Diskussion darauf verwiesen, dass Kappa ursprünglich für das niedrigste Skalenniveau entwickelt wurde und heute bessere Koeffizienten zur Abbildung der Reliabilität zur Verfügung stehen.

Smith und Kollegen (1997) verglichen die Reliabilität von Peer Review-Verfahren zur Messung der Ergebnis- und der Prozessqualität. Die Autoren konnten zeigen, dass die Reliabilität der Ergebnisqualität höher ausfällt als die der Prozessqualität. Sie beschreiben darüber hinaus Varianzquellen, die zu einer niedrigen Reliabilität führen können. So hat ein systematischer individueller Peer-Bias sowie ein systematischer Bias, bezogen auf das Training der Peers, negative Auswirkung auf die Ausprägung der Reliabilität.

Auch die in Deutschland durchgeführten Studien zur Reliabilität von Peer Review-Verfahren kommen zu ähnlichen heterogenen Ergebnissen. So liegen die in einem Pilotprojekt der gesetzlichen Unfallkassen berechneten Koeffizienten, je nach medizinischer Indikation, zwischen 0,3 und 0,5 (Neuderth, 2004). Auch die etablierte Qualitätssicherung in der medizinischen Rehabilitation durch das Peer Review der Reha-Entlassungsberichte zeigt schwankende Reliabilitätskennwerte. Bei der Beurteilung der sozialmedizinischen

Stellungnahmen innerhalb der Reha-Entlassungsberichte schwankt die Interrater-Reliabilität je nach medizinischer Grunderkrankung zwischen 0,15 in der Pneumologie und 0,6 in der Onkologie/Gynäkologie (Farin et al., 2003).

1.7 Peer Review in der Sozialmedizinischen Begutachtung

Die Studienlage zum Peer Review als Qualitätssicherungsinstrument zeigt, dass Peer-Übereinstimmungen limitiert sind und einer Optimierung bedürfen. Es ist davon auszugehen, dass in der Routine erhobene Reliabilitätskennwerte noch niedriger ausfallen als solche aus Forschungskontexten, da Peers in diesen motivierter arbeiten als in implementierten Qualitätssicherungsprogrammen (Goldman, 1994). Obwohl die Reliabilität begrenzt zu sein scheint, ist der klinische Nutzen des Peer Reviews im stationären Sektor als auch in der Rehabilitation belegbar. Mit Einführung des Verfahrens zeigen sich kontinuierliche Qualitätsverbesserungen, die sich möglicherweise in einer erhöhten Behandlungssicherheit und in verbesserten Patientenoutcomes widerspiegeln (Baumgarten & Klosterhuis, 2007; Edwards, 2011; Krahwinkel et al., 2011). Die DRV Bund übertrug die langjährigen Erfahrungen aus dem Reha-Qualitätssicherungsverfahren der Reha-Entlassungsberichte in die Qualitätsbemühungen der Sozialmedizin.

Grundlage des Peer Review-Verfahrens in der sozialmedizinischen Sachaufklärung ist die Beurteilung von sozialmedizinischen Gutachten, insbesondere die Bewertung der Ergebnisqualität. Ziel des Verfahrens ist die überregionale vergleichbare Gutachtenprüfung durch systematische Qualitätsbeurteilungen unter Anwendung einheitlicher Beurteilungskriterien. Das Konzept basiert auf der Sichtung von Gutachten ohne Beziehung zu den zugrunde liegenden Akten. Es eignet sich somit ausschließlich für sozialmedizinische Gutachten, deren Qualität anhand eines Kriterienkatalogs auch ohne Beziehung zur Akte zu beurteilen ist. (DRV, 2009c, S. 3-7). Im Peer Review werden Erstgutachten mit persönlicher Untersuchung von Versicherten beurteilt, die in den letzten zwölf Monate vor Antragstellung keine Rehabilitationsmaßnahme in Anspruch nahmen. Gutachten mit Auslandsberührung und Gutachten im Widerspruchsverfahren werden zunächst nicht geprüft.

Die einheitlichen Qualitätskriterien sind in einem Prüffragenkatalog aufgelistet. Die Entwicklung und Manualisierung des Prüffragenkatalogs wurde vom November 2009 bis Oktober 2010 von der DRV Bund in Kooperation mit der Universität Würzburg im Forschungsprojekt „Manualisierung eines Prüffragenkatalogs zur Qualitätssicherung sozialmedizinischer Gutachten (QSGUT)“ umgesetzt. Ziel des Projektes war es, ein Manual bestehend aus einem Prüffragenkatalog und einem Bewertungsschema zu erstellen, welches im Peer-Review eingesetzt werden kann (vgl. Löffler, Gerlich, Meng & Vogel, 2010). Das Verfahren ist als strukturiert-implizites Review konzipiert, da die Prüffragen und Antwortkategorien standardisiert verfasst und dezidierte Qualitätskriterien definiert wurden (vgl. Kapitel 1.6.3).

1.7.1 Manual zum Peer Review-Verfahren

Die Gesamtheit der Qualitätskriterien bildet die überprüfbare Ergebnisqualität der Sozialmedizinischen Begutachtung ab. Zusammengenommen definieren diese Kriterien indirekt die Anforderungen, die an ein sozialmedizinisches Gutachten gestellt werden. Die Qualitätskriterien untergliedern sich in sechs Einzelkriterien und einem übergeordneten Kriterium. Die Kriterien werden wiederum über definierte, standardisierte Prüffragen operationalisiert (vgl. Abb. 7). Die Anzahl als auch der Umfang der einzelnen Prüffragen variieren innerhalb der Einzelkriterien (DRV, 2009c, S. 8; Legner & Cibis, 2007).



Abbildung 7: Operationalisierung der Ergebnisqualität (modifiziert nach DRV, 2009c, S. 8).

Mit dem Abschlussbericht des Forschungsprojektes QSGUT wurde eine Entwurfsversion der Prüffragen, Einzelkriterien und des übergeordneten Kriteriums in einem Konsensverfahren gemeinsam mit Experten der DRV abgestimmt. Insgesamt enthielt der Prüffragenkatalog 23 Prüffragen, um Gutachten im Peer Review-Verfahren zu beurteilen. Im QSGUT-Projekt wurde zudem für jede Prüffrage eine Beurteilungssystematik ausgearbeitet und das Manual in einem Pre-Test evaluiert (vgl. Kap. 7.1.2). Der Prüffragenkatalog und dessen hinterlegte Beurteilungssystematik wurden im sogenannten „Manual zum Peer Review-Verfahren“ niedergelegt (vgl. Löffler et al., 2010).

Übergeordnetes Kriterium. Das Übergeordnete Kriterium „Nachvollziehbarkeit des Gutachtens“ beurteilt keine solitären Qualitätsaspekte, sondern das gesamte Gutachten als Ganzes. Es wird über lediglich eine Prüffrage im Manual definiert. Die Nachvollziehbarkeit wird durch die Beurteilung der gutachteninternen Argumentationskette (vgl. Abb. 6) geprüft. Unter der Argumentationskette wird die logische inhaltliche Verlinkung der einzelnen Begutachtungsschritte verstanden. Mit der Sozialmedizinischen Beurteilung des Leistungsvermögens werden gutachterliche Aussagen zum quantitativen und qualitativen Leistungsvermögen, zu Beginn und Dauer der Leistungsminderung sowie zum Reha-Bedarf getätigt. Im übergeordneten Kriterium wird beurteilt, inwieweit die getroffenen Aussagen zum Leistungsvermögen ein nachvollziehbares Ergebnis der Informationssammlung und -bewertung darstellen. Mit dem Gesamturteil zur Nachvollziehbarkeit wird auch die Neutralität des Gutachters geprüft (DRV, 2013, S. 6-7).

Das übergeordnete Kriterium hat die Funktion eines K.O.-Kriteriums. Werden im Review gravierende Mängel bei der Nachvollziehbarkeit aufgedeckt, so ist das Gutachten als „durchgefallen“ zu bewerten und wäre im Rentenverfahren eigentlich nicht reliabel verwertbar (Cibis et al., 2011, S. 110). Da es sich bei dem Verfahren jedoch um eine retrospektive Gutachtenbeurteilung handelt, wurde die Entscheidung über eine EM-Rente bereits gefällt.

Eine Schwierigkeit bei der Beurteilung dieses Qualitätskriterium liegt darin begründet, dass im Peer Review nicht die Richtigkeit der Argumentationskette geprüft wird, sondern ausschließlich die Nachvollziehbarkeit. Da die prüfenden Peers selbst erfahrene Sozialmediziner sind, ist es möglich, dass diese aufgrund der Informationen im Gutachten zu einer anderen Auffassung hinsichtlich der Leistungsbeurteilung kommen. In diesen Fällen kann die im Gutachten zusammengetragene Argumentationskette jedoch trotzdem nachvollziehbar sein (DRV, 2009c, S. 11).

Das übergeordnete Kriterium prüft auf einer drei-stufigen Ordinalskala, ob Brüche in der Argumentationskette vorliegen. Für die Beurteilung ist es von Relevanz, ob diese Brüche von einem Sozialmediziner überbrückt werden können. Die drei Ausprägungen der ordinalen Beurteilungsskala werden mittels einer Ampelbewertung mit den Farben grün, gelb und rot beurteilt. Treten keine nennenswerten Brüche der Argumentationskette auf, so ist das Gutachten mit „grün“ zu bewerten. Die Bewertung „gelb“ wird vergeben, wenn Brüche in der Argumentationskette vorhanden sind, diese jedoch durch eine Neubewertung der im Gutachten enthaltenen Informationen von einem Sozialmediziner überbrückt werden können. In diesem Fall ist das Gutachtenergebnis trotz der Brüche noch nachzuvollziehen. Liegen Brüche in der Argumentationskette vor, die auch vom Sozialmediziner nicht mehr überbrückt werden können, wird die Nachvollziehbarkeit mit „rot“ bewertet. Auch gravierende Verstöße gegen das Neutralitätsgebot führen zu einer „roten“ Beurteilung der Ampelbewertung (Cibis et al., 2011, S. 111; DRV, 2013, S. 7).

Einzelkriterien. Die Einzelkriterien werden durch eine vorgegebene Anzahl von Prüffragen operationalisiert und überprüfen spezifische Qualitätsaspekte des sozialmedizinischen Gutachtens. Aus der unterschiedlichen Anzahl an Prüffragen je Einzelkriterium ist keine Gewichtung hinsichtlich der Wichtigkeit der Kriterien ableitbar. Die untersuchten Kriterien „Formale Gestaltung“, „Verständlichkeit“, „Transparenz“, „Vollständigkeit“, „Medizinisch-wissenschaftliche Grundlagen“ und „Wirtschaftlichkeit“ sind spezifische Mängelkategorien, bei denen, im Gegensatz zum übergeordneten Kriterium, gravierende Mängel die allgemeine Verwertbarkeit des Gutachtens nicht grundsätzlich in Frage stellen. Ein dokumentierter Mangel in den Einzelkriterien kann konkrete Hinweise für das Qualitätsmanagement liefern (DRV, 2013, S. 5). Die Kriterien und deren zentrale Prüfgegenstände werden in Tabelle 9 näher erläutert.

Tabelle 9: Einzelkriterien zur Erfassung der Qualität von sozialmedizinischen Gutachten (modifiziert nach Cibis et al., 2011. S. 111).

Einzelkriterium	Prüfinhalte des Peer Review-Verfahrens
Formale Gestaltung	• Einhaltung der formalen Anforderungen an das Gutachten gemäß der Gliederungsvorgabe der DRV-Schrift Band 21 • Verwendung des einheitlichen Formularsatzes bestehend aus Deckblatt und Schlussblatt
Verständlichkeit	• Sprachliche Darstellung • Umgang mit medizinischer Fachsprache • Umgang mit sozialrechtlichen Terminologien
Transparenz	• Beschreibung der verwendeten Verfahren, Methoden und Begutachtungsinstrumente • Beschreibung der Kriterien zur Beurteilung der Befunde • Beschreibung der Herkunft anamnestischer Informationen und Befunde
Vollständigkeit	• Inhaltlich vollständige Darstellung von Anamnese, Befunden, Diagnosen und Epikrise • Vollständige Beantwortung der sozialmedizinischen Fragestellungen des Gutachtens
Medizinisch-wissenschaftliche Grundlagen	• Einhaltung des anerkannten Standards medizinischer Erkenntnisse • Berücksichtigung der Begutachtungsliteratur der DRV
Wirtschaftlichkeit	• Eignung und Notwendigkeit der eingesetzten Verfahren, Methoden und Begutachtungsinstrumente

Mängel in den Prüffragen der Einzelkriterien werden auf einer vierstufigen Ordinalskala beurteilt. Das Vier-Kategorien-Bewertungssystem verfügt über die Abstufungen „keine Mängel“, „leichte Mängel“, „deutliche Mängel“ und „gravierende Mängel“. Die Besonderheit des Verfahrens liegt darin begründet, dass für jede Prüffrage die Mängelbewertung je nach Frageninhalt verschiedentlich ausgestaltet ist. Jeder Prüffrage ist demnach eine eigene Mängelabstufung hinterlegt (Tab. 10).

Tabelle 10: Operationalisierung der Mängelkategorien am Beispiel der Prüffrage B1: „Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich?“ (entnommen aus DRV, 2013. S. 15).

Mangelkategorie	Deskription der Mängelkategorien
Leichte Mängel	Das Gutachten enthält unnötige Längen. ODER Im Gutachten werden unnötig komplizierte Formulierungen verwendet.
Deutliche Mängel	Durch die sprachliche Ausdrucksweise sind Aussagen vereinzelt nicht eindeutig bzw. nicht verständlich.
Gravierende Mängel	Durch die sprachliche Ausdrucksweise sind Aussagen überwiegend nicht eindeutig bzw. nicht verständlich.

Durch die hinterlegten Mängel im Vier-Kategorien-Bewertungssystem erhält der Peer im strukturierten, impliziten Review-Verfahren der Sozialmedizinischen Begutachtung konkrete Hinweise darüber, welche Qualitätsbereiche durch die Prüffrage abgedeckt werden. Die vorgegebene Bewertungsmatrix ist im Peer Review immer anzuwenden. In begründeten Ausnahmefällen kann jedoch von einer Mängelzuweisung abgesehen werden. Dies ist beispielsweise der Fall, wenn durch besondere Gutachtenkonstellationen (z. B. verkürzte Begutachtung im Spätstadium eines Carcinoms, bei Pflegebedürftigkeit/bettlägerigen Personen, Zusatzbegutachtung zu einem vorliegenden Hauptgutachten) zu beurteilende Gliederungspunkte im Gutachten nicht oder nur verkürzt dargestellt wurden (DRV, 2013, S. 8).

Dokumentation der Mängel. Alle dokumentierten Qualitätsdefizite sind im Peer Review der Sozialmedizinischen Begutachtung zusätzlich zur Mängelbewertung in aussagekräftigen, stichwortartigen Freitexten zu beschreiben. Wurden keine Mängel kodiert, ist eine weitere Dokumentation nicht notwendig (DRV, 2013, S. 7).

1.7.2 Explorative Analyse der Interrater-Reliabilität

Während der Konsentierung des Prüffragenkatalogs im Rahmen des Projektes QSGUT wurde eine vorangegangene Entwurfsversion des oben beschriebenen Manuals in einem explorativen Peer Review-Verfahren getestet. Am Praxistest nahmen zwölf Sozialmediziner teil. Die Peers kamen aus elf Dienststellen der DRV und wiesen die Fachrichtungen Chirurgie (n = 2), Orthopädie (n = 1), Innere Medizin (n = 3), Allgemeinmedizin (n = 3) und Neurologie/Psychiatrie (n = 3) auf. Beurteilungsbasis war eine systematische Auswahl von 24 anonymisierten Gutachten im Erstantragsverfahren auf EM-Rente. Eine Stratifizierung der Gutachten erfolgte nach Fachgebiet, „Ampelfarbe“ sowie ggf. dem Gutachtenergebnis (Leistungsbeurteilung). Jeder Peer bewertete sechs Gutachten anhand der Erprobungsversion des Manuals. Das Review sollte ausschließlich der formativen Evaluation dienen und war noch nicht für die Analyse der Interrater-Reliabilität konzipiert.

Berechnet wurde das Interrater-Reliabilitätsübereinstimmungsmaß Fleiss Kappa hinsichtlich des übergeordneten Kriteriums. Dieser Koeffizient bestimmt die Urteilsübereinstimmung für mehrere Rater, wenn dieselben Objekte von gleich vielen, aber nicht denselben Ratern beurteilt werden. Im Übereinstimmungsmaß (κ_m -Koeffizient von Fleiss) wird berücksichtigt, dass eine gewisse Urteilsübereinstimmung zufällig zustande kommen kann. Die resultierende Übereinstimmungswahrscheinlichkeit wird auf eine signifikante Abweichung von der Zufallsübereinstimmungsrate geprüft (Löffler et al., 2010, S. 11).

Die prozentuale Übereinstimmung in der Gruppe Innere Medizin/Allgemeinmedizin ergab eine Konkordanz in Höhe von 41 %. Somit würden zwei zufällig ausgewählte und voneinander unabhängig urteilende Peers bei Beurteilung der Gutachten zu 41 % in ihren Urteilen übereinstimmen. Die Werte für die Beurteilerübereinstimmung unterscheiden sich nominal, aber nicht signifikant von der erwarteten zufälligen Übereinstimmungsrate, die bei 34% ($\kappa_m = 0,11$, $p > 0,05$) lag. Die Übereinstimmung der Peers im Bereich Neurologie/ Psychiatrie fällt mit 33 % geringer aus als die zu erwartende zufällige Übereinstimmung von 36% ($\kappa_m = -0,04$, $p > 0,05$). Eine mittlere Übereinstimmung der Peers kann im Bereich Orthopädie/Chirurgie erzielt werden. Diese liegt mit 67 % signifikant ($\kappa_m = 0,48$, $p < 0,05$) über der Zufallsübereinstimmung von 36 % (Löffler et al., 2010, S. 16).

Nach Ansicht der Autoren weisen die eher geringen Reliabilitätsmaße darauf hin, dass eine hohe Objektivität und Reliabilität im Peer Review nur erzielt werden kann, wenn die Qualitätskriterien möglichst beobachtungsnah durch Prüffragen operationalisiert werden, auf einem differenzierten Beurteilungsschema basieren und eine intensive Schulung und Einarbeitung der Peers stattfindet (Löffler et al., 2010, S. 18).

Nach dem Review wurden die Operationalisierungen der Prüffragen sowie das Beurteilungsschema in Zusammenarbeit mit der DRV Bund überarbeitet. Das daraus hervorgehende Manual zum Peer Review-Verfahren zur Qualitätssicherung in der Sozialmedizinischen Begutachtung wurde dann in einem weiteren Schritt psychometrisch evaluiert. Die Ergebnisse dieser Evaluation werden im vorliegenden Untersuchungsvorhaben dargestellt.

1.8 Zusammenfassung und Relevanz des Forschungsvorhabens

Chronische Erkrankungen stellen einen erheblichen volkswirtschaftlichen Kostenfaktor zu Lasten der GKV und DRV dar (Maaz et al., 2007, S. 12). Der progrediente Verlauf dieser Erkrankungen führt neben der regelmäßigen Inanspruchnahme von Leistungen des Gesundheitssystems auch häufig zu einer reduzierten beruflichen Leistungsfähigkeit (Krämer & Bengel, 2016, S. 28; Vogel, Faller & Reusch, 2016, S. 350; Wolf-Kühn & Morfeld, 2016, S. 33). Bis heute kann nicht abschließend geklärt werden, wie sich der demographische Wandel langfristig auf den Verlauf und die Entwicklung von chronischen Erkrankungen auswirken wird (Chatterji et al., 2015). Jedoch gilt als gesichert, dass das Risiko, an einer chronischen Krankheit zu leiden, mit steigendem Alter zunimmt (Peters et al., 2010). Es liegen jedoch keine hinreichenden Forschungsergebnisse vor, um valide abschätzen zu können, ob eine längere Lebenszeit mit mehr Zeit in Gesundheit oder in Krankheit assoziiert ist. Als gut belegt gilt hingegen der Einfluss des sozialen Status auf die Inzidenz und Prävalenz von (chronischen) Erkrankungen. Ein niedriger sozio-ökonomischer Status ist häufig vergesellschaftet mit höheren Belastungen, niedrigeren Ressourcen und vermehrten Risikofaktoren, was zur Manifestation einer chronischen Erkrankung führen kann (Lampert et al., 2016; Mielck, 2012).

Ist die Minderung der Leistungsfähigkeit nicht kompensierbar, besteht für den Versicherten die Möglichkeit, eine EM-Rente zu beantragen (Cibis et al., 2011, S. 99). In der Sozialmedizinischen Sachaufklärung wird nach Prüfung der medizinischen und versicherungsrechtlichen Voraussetzungen über die Gewährung einer EM-Rente entschieden. Die Überprüfung der persönlichen medizinischen Voraussetzungen erfolgt im Rahmen der Sozialmedizinischen Begutachtung mit persönlicher Untersuchung oder nach Aktenlage (Cibis et al., 2011, S. 78). Dabei ist kennzeichnend, dass sich das Diagnosespektrum bei EM-Rentennewuzugänge in den letzten 20 Jahren elementar gewandelt hat. Während 1996 am häufigsten Erkrankungen des Skeletts, der Muskeln und des Bindegewebes für den Eintritt in die Erwerbsminderung verantwortlich waren, sind mit 42,8% aktuell psychische Erkrankungen die häufigste Indikation für eine EM-Rente (DRV, 2017b).

Der Wandel der medizinischen Indikationen stellt gutachterlich tätige Sozialmediziner vor die Schwierigkeit, zunehmend nicht direkt beobachtbare, intrapsychische Vorgänge objektivieren zu müssen. Im Allgemeinen ist die Begutachtung als komplexer und individueller Beurteilungsprozess definiert, der verschiedenen Störfaktoren unterliegen kann. Diese Unsicherheiten und Varianzen führen dazu, dass die Replizierbarkeit von Leistungsbeurteilungen bzw. die Übereinstimmung von medizinischen Sachverständigen limitiert ist. Diese Limitationen beziehen sich sowohl auf die Diagnosen als auch die

endgültige Leistungsbeurteilung (Clark et al., 1988; Dickmann & Broocks, 2007; Elder et al., 1994; Ingravallo et al., 2008). Da es kaum wissenschaftliche Grundlagen für die gutachterliche Entscheidungsfindung und -begründung gibt, bleibt die Validität von Gutachten bis auf weiteres begrenzt. Die indikationsübergreifende Literaturrecherche zur Evidenzbasierung in der Sozialmedizin zeigte, dass Studien mit höherer Evidenz in der Sozialmedizin selten zu finden sind. Häufig ist die Studienevidenz aus dem Bereich der Sozialmedizinischen Begutachtung auf den Expertenkonsens gering, was nicht zuletzt der Komplexität der Begutachtung geschuldet ist (Strahl et al., 2016a, 2016b; Strahl & Vogel, 2016).

Um dennoch die Gleichbehandlung aller Versicherten zu gewährleisten, besteht die Notwendigkeit, umfangreiche interne und externe Qualitätssicherungsmaßnahmen für Gutachten durchzuführen. Neben der Ausarbeitung von verbindlichen Hinweisen zur Erstellung von Gutachten, Leitlinien zur Begutachtung und eines sozialmedizinischen Glossars, werden Qualitätszirkel, Fortbildungsprogramme und hausinterne Qualitätssicherungsprogramme zur Qualitätssicherung eingesetzt (Ueberschär, 2008). Seit 2004 arbeitete die DRV Bund auch an einer trägerübergreifenden Qualitätssicherung auf Basis eines Peer Review-Verfahrens. Im Peer Review beurteilt ein gleichgestellter Fachkollege die Qualität eines Gegenstandes oder einer Dienstleistung anhand von a priori definierten Qualitätskriterien (Legner & Cibis, 2007).

Im Peer Review der Sozialmedizinischen Begutachtung sollen sozialmedizinische Gutachten anhand eines Prüffragenkatalogs durch Sozialmediziner beurteilt werden. Das wichtigste zu prüfende psychometrisches Gütekriterium ist dabei die Reliabilität. Eine hohe Reliabilität gewährleistet zuverlässige Qualitätsbeurteilungen, auch wenn die Bewertung von nur einem Peer durchgeführt wird (Shaw, 2001, 2000; Wirtz & Caspar, 2002). Der dargestellte Forschungsstand zum Peer Review beschreibt eine national als auch international sehr hohe Variabilität der Reliabilität (Broder et al., 2002; Farin et al., 2003; Goldman, 1994; Lilford et al., 2007; Strahl et al., 2016). Trotz der Limitierungen lassen sich nach Einsatz des Peer Reviews Qualitätsverbesserungen in den geprüften Qualitätskriterien feststellen (Baumgarten & Klosterhuis, 2007; Edwards, 2011; Krahwinkel et al., 2011). Der durch die DRV Bund und der Universität Würzburg entwickelte Prüffragenkatalog zum Peer Review ist ein neu entwickeltes Verfahren zur trägerübergreifenden Qualitätssicherung. Durch Rückmeldung der Review-Ergebnisse an die entsprechenden RV-Träger soll die Qualität der sozialmedizinischen Gutachten standardisiert gemessen und langfristig verbessert werden. Um den reliablen Einsatz des Verfahrens in der Routine zu gewährleisten, ergibt sich zuvor die Notwendigkeit zur Evaluation des Peer Review-Verfahrens.

Die Darstellung des Begutachtungsgeschehens und der Qualitätssicherung in der Sozialmedizin zeigt einen konkreten Forschungsbedarf hinsichtlich zweier Bereiche:

1. Es existieren, trotz der durchgeführten systematischen Literaturrecherche, keine eindeutigen Erkenntnisse über die Evidenzlage der Sozialmedizinischen Begutachtung, anhand derer die Begutachtung und Leitlinienentwicklung normierter durchgeführt werden könnte.
2. Es existiert keine wissenschaftlich evaluierte und trägerübergreifende Qualitätssicherung der Sozialmedizinischen Begutachtung.

Das Ziel dieser Forschungsarbeit ist dementsprechend die Identifikation weiterer evidenzbasierter Publikationen. Dazu wird eine ergänzende systematischen Literaturrecherche zur Evidenz in der Sozialmedizinischen Begutachtung angefertigt. Aufbauend auf das Forschungsprojekt „LitBegut“ (vgl. Kap. 1.5.3) wird eine aktualisierte und indikationsspezifische Recherche durchgeführt. Mit der Recherche könnten Publikationen mit höherer Evidenz identifiziert werden, um das Begutachtungsgeschehen valider und transparenter gestaltet zu können. Eine höhere Standardisierung und Validität von Gutachten hätte auch direkte Auswirkungen auf die Qualitätssicherung. In einem Pilotprojekt der DRV Bund soll im Folgenden die Reliabilität des Peer Reviews zur trägerübergreifenden Qualitätssicherung der Sozialmedizinischen Begutachtung untersucht werden. Dazu wird im Rahmen des Projektes ein Peer Review durchgeführt, bei dem die zu beurteilenden Gutachten von mehreren Sozialmediziner*innen gleichzeitig bewertet werden, um die Interrater-Reliabilität des Verfahrens zu bestimmen. Aus der Literatur ist bekannt, dass die Reliabilität eines Peer Reviews maßgeblich durch den Peer selbst (Peer-Bias) beeinflusst werden kann (Tuijn et al., 2012). Diese Variabilität des Peer Bias wird überwiegend durch die individuellen Beurteilermaßstäbe und Persönlichkeitsmerkmale des Peers sowie durch das Training des Peers im Umgang mit dem Verfahren determiniert. Daher ist es für die Implementierung des Peer Reviews in die Routine notwendig, eine standardisierte Schulung für alle teilnehmenden Peers zu entwerfen und zu evaluieren. Diese Peer-Schulungen können nicht nur als initiales Training für den Peer, sondern auch als Element der regelmäßigen Fort- und Weiterbildung für alle am Peer Review teilnehmenden Sozialmediziner*innen dienen. Insgesamt stellt das vorliegende Untersuchungsvorhaben ein zentrales Element in der Weiterentwicklung der Sozialmedizin und Begutachtung dar.

Von den Ergebnissen der vorliegenden Forschungsarbeit können sowohl die Versicherten, die sozialmedizinischen Gutachter sowie die gesamte Solidargemeinschaft profitieren. Die Komplexität des sozialmedizinischen Begutachtungsgeschehens stellt insgesamt sehr hohe Anforderungen an die Qualität des Begutachtungsprozesses. Vor dem Hintergrund des hohen Leidensdrucks der Versicherten erwarten diese bei Anträgen auf

Erwerbsminderungsrente eine schnelle, unparteiische und korrekte Sozialmedizinische Begutachtung (VDR, 2004, S. 3). Die systematische Recherche zur Evidenz in der Sozialmedizin könnte einen wesentlichen Beitrag dazu leisten, die zunehmende Bedeutung der fortschreitenden medizinischen Entwicklungen in die Abläufe der Begutachtung zu integrieren und aus den Ergebnissen einheitliche evidenzbasierte Beurteilungskriterien abzuleiten. Zusätzlich werden durch den Einsatz externer Qualitätssicherungsverfahren die Begutachtungsprozesse und -ergebnisse transparenter und nachvollziehbarer gestaltet. Letztlich profitiert jeder Versicherte im Antragsverfahren durch die optimierten Anforderungen an die sozialmedizinischen Gutachten, da durch diese Maßnahmen eine zunehmende Gleichbehandlung aller Versicherten gewährleistet werden kann. Darüber hinaus können evidenzbasierte Beurteilungskriterien zu verbindlichen Bewertungsmaßstäben der gutachterlich tätigen Sozialmediziner führen und damit die fachliche Unabhängigkeit und Unparteilichkeit stärker fördern. Durch die Steuerung des ärztlichen Handelns in Form von evidenten Begutachtungsleitlinien kann schlussendlich auch der in der Begutachtung vorhandene Peer Bias minimiert werden. Insgesamt wirken sich alle Maßnahmen positiv auf die interne Validität der Begutachtung und der sozialmedizinischen Gutachten aus.

Abschließend kann davon ausgegangen werden, dass durch die Einführung einer routinemäßigen trägerübergreifenden Qualitätssicherung auch die „Wirtschaftlichkeit“ der Begutachtung erhöht wird. Durch die externe Gutachtenüberprüfung können unnötige Doppel- bzw. Mehrfachuntersuchungen aufgedeckt werden. Langfristig werden durch diese Maßnahme vermeidbare Kosten für die DRV eingespart. Insgesamt wird die sachgerechte sozialmedizinische Leistungsbeurteilung weiterentwickelt und gefördert und dadurch auch die Anzahl von möglichen Widerspruchsverfahren reduziert.

2 Aufbau des empirischen Teils

Der Aufbau der vorliegenden Forschungsarbeit orientiert sich am Rahmenmodell des Public Health Action Cycle, der ursprünglich für die Planung und Umsetzung von gesundheitsbezogenen Interventionen konzipiert wurde (Rosenbrock, 1995; siehe Abb. 8). Das Stufenmodell wird auch als Ausgangspunkt des systematischen Qualitätsmanagements in der Gesundheitsförderung und Prävention genutzt (Rosenbrock, 2004).

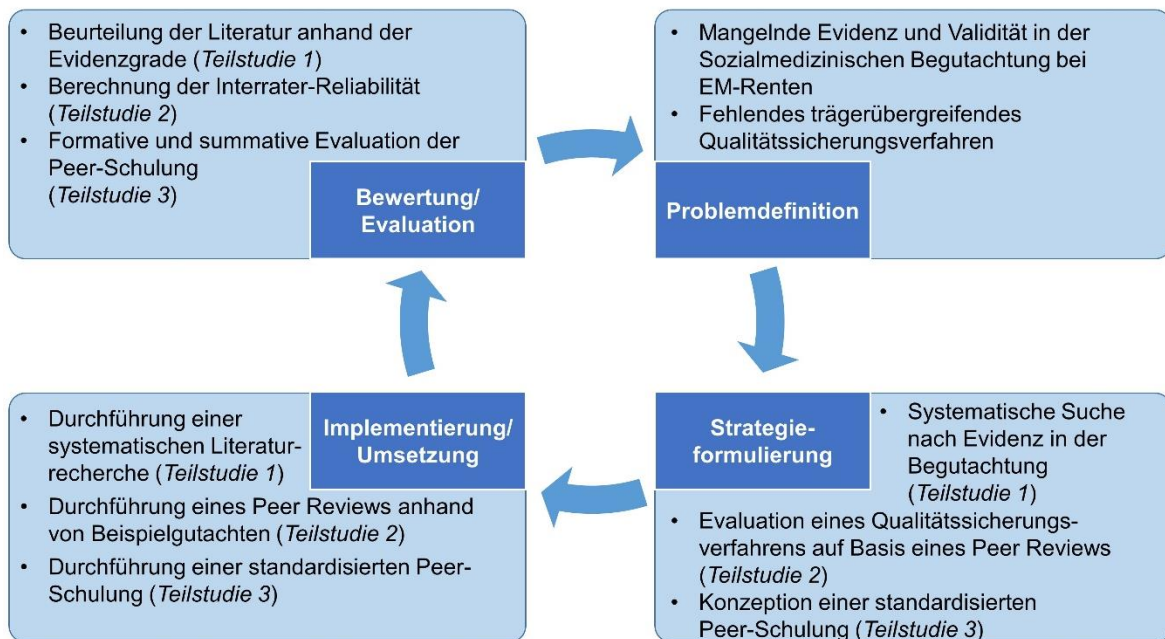


Abbildung 8: Struktur der Dissertation nach dem Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).

Im Modell wird ausgehend von einer spezifischen Problemdefinition eine evidenz- und theoriegestützte Handlungsstrategie zur Auflösung der Problemlage abgeleitet. Mit der Strategieformulierung werden zu untersuchende bzw. zu optimierende Zielparameter, Zielgrößen, Zielgruppen und Rahmenbedingungen konkretisiert. Im darauffolgenden Implementierungsschritt werden konkrete Maßnahmen zur Lösung des definierten Problems umgesetzt. Die praktische Umsetzung der Strategie endet mit einer Evaluation. Diese ist zentraler Bestandteil des Public Health Action Cycle und beschreibt die Ergebnisbewertung der Intervention anhand der in der Strategieformulierung festgelegten Zielparameter. Die Ergebnisse der abschließenden Evaluation können im Rahmen des zirkulären Modells den zuvor festgelegten Problemdefinitionen gegenübergestellt und abgeglichen werden. Sollten weiterhin Probleme bestehen oder andere Defizite infolge der Implementierung entstanden sein, kann eine erneute Anpassung der Strategie erfolgen (Kolip, 2006).

Die vorliegende Arbeit folgt diesem Modell und beschreibt im theoretischen Hintergrund die derzeitige Problemlage und die daraus abgeleiteten Forschungsbedarfe in der Sozialmedizin. Die Notwendigkeit zur Forschung wird primär durch die unklare Evidenzlage hinsichtlich des Sozialmedizinischen Begutachtungsprozesses und der fehlenden trägerübergreifenden Qualitätssicherung sozialmedizinischer Gutachten bestimmt. Der empirische Teil dieser Arbeit untergliedert sich in drei Teilstudien, die jeweils mit einer eigenen Methodik durchgeführt und beschrieben werden. Die hier dargelegten Studien zur Förderung der Qualitätssicherung in der Sozialmedizinischen Begutachtung bauen inhaltlich aufeinander auf und beschäftigen sich mit der:

- Systematischen Literaturrecherche zur Identifikation wissenschaftlich fundierter Literatur zur Evidenzbasierung in der Sozialmedizinischen Begutachtung (Teilstudie 1),
- Evaluation einer trägerübergreifenden Qualitätssicherung der DRV für die Sozialmedizinische Beurteilung bei Anträgen zur EM-Rente auf Basis eines Peer Review-Verfahrens (Teilstudie 2),
- Entwicklung und Evaluation einer Peer-Schulung für das trägerübergreifende Qualitätssicherungsverfahren (Teilstudie 3).

Die Strategieformulierung, die der Implementierungsphase vorausgeht, wird in den Darstellungen der Teilstudien separat beschrieben. Die Strategie wurde in der Planungsphase der Projekte „LitBegut“ und „EvaBegut“ gemeinsam mit der DRV Bund ausgearbeitet und konsentiert. Die Planung und Durchführung der systematischen Literaturrecherche zur Evidenzbasierung erfolgt in Anlehnung an das Projekt „SchuP“. Mit der Umsetzung der Teilstudien wurden die ausgearbeiteten Strategien implementiert und abschließend evaluiert. Aus den hier vorgestellten und diskutierten Ergebnissen werden abschließend konkrete Handlungsempfehlungen für die Praxis abgeleitet.

3 Teilstudie 1: Systematische Literaturrecherche

Die erste Teilstudie beschäftigt sich zunächst mit den Review-Objekten selbst, den sozialmedizinischen Gutachten. In einer systematischen Literaturrecherche wurden aktuelle evidenzbasierte Erkenntnisse in der Sozialmedizinischen Begutachtung untersucht. Die Recherche legte ihren Fokus primär auf den Ablauf und die Methodik der Begutachtung. Ziel war die Identifikation und Beurteilung von praktischen Instrumenten und Verfahren zur Erhöhung der Validität der Sozialmedizinischen Begutachtung.

Das systematische Review hat sich insbesondere seit den 1990er Jahren im medizinischen Bereich als Instrument der zielgerichteten Forschungssynthese etabliert. Im Unterschied zu einem nicht systematischen, narrativen Review wird ein systematischer und transparenter Forschungsüberblick, bezogen auf eine spezifische Fragestellung, dargelegt. Die systematische Recherche ist Ergebnis einer zielgerichteten Suche und definiert sich nicht als bloße Zusammenstellung von Zufallsfunden. In der Umsetzung beinhaltet dies die Vorabdefinition von Ein- und Ausschlusskriterien, die systematische Schlag- und Stichwortsuche in wissenschaftlichen Datenbanken sowie die präzise und transparente Dokumentation aller Einzelschritte der Literaturrecherche (Antes, 1998; Döring & Bortz, 2016, S. 158). Im Gegensatz zu narrativen Reviews ermöglichen systematische Übersichtsarbeiten eine objektivere Beurteilung der Datenlage und können auf diese Weise dazu beitragen, Evidenzen zu belegen (Tranfield, Denyer & Smart, 2003). In der Regel sind Artikel in Fachzeitschriften die akademischen Primärquellen, welche die Grundlage einer systematischen Recherche bilden. Eine systematische Recherche, ggf. ergänzt durch eine Meta-Analyse, gilt aktuell als Goldstandard zur Identifikation der besten verfügbaren Evidenz (Shekelle et al., 1999).

PICO-Schema. Zur Bearbeitung des gewählten Forschungsthemas im Rahmen einer systematischen Recherche ist es notwendig, relevante und bearbeitbare Fragestellungen zu formulieren und diese in für elektronischen Datenbanken geeignete Fragen zu übertragen (Kleibel & Mayer, 2011, S. 29). Bei der systematischen Darstellung des Forschungsstandes sollten folgende Fragestellungen berücksichtigt werden:

- Welche zentralen Befunde zum Thema haben sich ergeben?
- Welche Anwendungsfelder wurden in der Forschung zu diesem Themengebiet angesprochen?
- Inwiefern wurden aus den Forschungsergebnissen bereits Praxismaßnahmen abgeleitet?
- Wie ist der aktuelle Forschungsstand insgesamt zu beurteilen? (Döring & Bortz, 2016, S. 163).

Die Anzahl gefundener Literaturstellen, die Qualität sowie die Treffergenauigkeit der Auswahl der Publikationen hängen maßgeblich von der gewählten Suchstrategie ab. Zur Eingrenzung und Präzisierung des zu untersuchenden Forschungsgegenstandes hat sich das PICO-Schema als hilfreich erwiesen. Das Schema unterstützt den Anwender bei der Formulierung suchtauglicher Fragestellungen. Bei der Anwendung werden vor Recherchebeginn die zu untersuchende Population (P), die Intervention/en (I), Vergleichs- oder Kontrollgruppen (C) sowie das zu untersuchende Outcome (O) festgelegt. Zusätzlich können weitere Parameter wie Studientyp, ein spezifischer Suchzeitraum oder die Umgebung bzw. das Setting der zu untersuchenden Studien eingegrenzt werden (Blümle, Lagrèze & Motschall, 2018; Guba, 2008; Guyatt et al., 2011).

Bei der Leitlinienerstellung hat die systematische Übersichtsarbeit einen hohen Stellenwert. Die Qualität dieser Arbeiten hängt jedoch stark davon ab, welche Publikationen in die Untersuchung eingeschlossen und wie transparent und ausführlich die Ergebnisse berichtet werden. Folglich variiert die Qualität von systematischen Übersichtsarbeiten stark (Mulrow, 1987). Systematische Reviews nutzen ein standardisiertes Vorgehen, um bei den verschiedenen Suchschritten und -strategien eine hohe Transparenz und Vergleichbarkeit der Suchergebnisse zu garantieren.

PRISMA-Statement. Zur Gewährleistung der methodischen Qualitätsstandards von systematischen Literaturarbeiten nutzt die vorliegende Untersuchung das Statement zu Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-Statement). Das PRISMA-Statement wurde mit dem Ziel der vollständigen und transparenten Dokumentation von systematischen Reviews und Meta-Analysen entwickelt. Das Statement ist eine Weiterentwicklung des QUORUM-Statements, das ursprünglich ausschließlich für Meta-Analysen konzipiert wurde. Das PRISMA-Statement besteht aus einer 27-Punkte-Checkliste und einem Flussdiagramm für die transparente Darstellung der Literaturrecherche. Es legt seinen Schwerpunkt auf die Aggregation randomisierter Studien, kann aber auch Grundlage für das Berichten von systematischen Übersichten anderer Studien sein. Das PRISMA-Statement kann bei der kritischen Bewertung publizierter systematischer Übersichten hilfreich sein, versteht sich qua Definition jedoch nicht als Instrument zur Qualitätsbeurteilung. Obgleich die Checkliste nicht als Grundlage von Qualitätsskalen verwendet wird, führt die transparente Berichterstattung bei vorhandenen Mängeln schneller zum Entdecken systematischer Fehler (Moher, Liberati, Tetzlaff & Altman, 2011). Die Anwendung des PRISMA-Statements kann daher indirekt sowohl die Qualität der Berichterstattung als auch die methodische Qualität von Reviews erhöhen (Panic, Leoncini, de Belvis, Ricciardi & Boccia, 2013).

Evidenz in der Sozialmedizinischen Begutachtung. Bezugnehmend zur Evidenz der Sozialmedizinischen Begutachtung der RV haben Strahl und Vogel (2016) in einem systematischen Review versucht, den Erkenntnisstand zur Begutachtung bei EM-Renten und Anträgen zur Rehabilitation zu aggregieren (vgl. Strahl & Vogel, 2016). Der sehr hohe Anteil an narrativen Reviews (Evidenzlevel: IV) und Beobachtungsstudien (Evidenzlevel: III) ließ den Schluss zu, dass die Evidenzbasis der 52 in das Review eingeschlossenen Studien zur Sozialmedizinischen Begutachtung „limitiert“ sei. Vor allem Publikationen, die einen Fragebogen zur Begutachtung bestimmter Störungen evaluierten, wurden größtenteils mit einem Evidenzlevel von III bewertet. Vor dem Hintergrund, dass Fragebögen üblicherweise nicht durch RCT-Studien mit mehreren Messzeitpunkten evaluiert werden, entspricht dieses Evidenzlevel auch dem eigentlichen Studienziel. Eine ähnliche Argumentation wurde auch für die eingeschlossenen Publikationen zur Begutachtungspraxis angewendet, die überwiegend dem Evidenzgrad IV aufweisen. So wird die Beschreibung des Begutachtungsvorgangs vorrangig durch die Beiträge von Fachleuten oder Expertengruppen, Konsensuskonferenzen oder der klinischen Erfahrungen anerkannter Autoritäten dargelegt, was grundsätzlich keine höherwertige Evidenz darstellt (Strahl & Vogel, 2016, S. 63).

Die niedrige Evidenzbasis ist Ausdruck der komplexen biopsychosozialen Verankerung des sozialmedizinischen Begutachtungsgeschehens, sodass die Planung von Studien mit höheren Evidenzgraden erschwert ist. Gleichzeitig wurde die Recherche nicht nach dem standardisierten PICO-Schema durchgeführt, da weder eine bestimmte Patientengruppe noch eine spezielle Intervention oder nach einer Kontrollgruppe gesucht werden konnte. Darüber hinaus wird in der Diskussion darauf hingewiesen, dass es nicht auszuschließen ist, dass in einer weiteren indikationsspezifischen Recherche weitere unentdeckte Literaturstellen gefunden werden könnten (Strahl & Vogel, 2016, S. 64)

3.1 Fragestellung

Ziel der Untersuchung ist die Bestandsaufnahme der evidenzbasierten Literatur zur Sozialmedizinischen Begutachtung in Ergänzung zum Forschungsprojekt „LitBegut“. Mit der vorliegenden Untersuchung werden die oben genannten Limitationen des Forschungsprojektes aufgegriffen und bearbeitet. Mit einer differenzierten und geänderten Suchstrategie werden im Folgenden die häufigsten medizinischen Indikationen bei EM-Rente in einer individuellen Literaturrecherche untersucht.

Vor allem für die kontinuierliche Entwicklung und Revision von Begutachtungsleitlinien sind evidenzbasierte Informationen von zentraler Bedeutung. Alle im Projekt „LitBegut“ identifizierten Begutachtungsleitlinien weisen die Qualitätsklassifikation S1 auf. Hierbei handelt es sich um Handlungsempfehlungen von Expertengruppen, die in einem informellen Verfahren konsentiert worden sind (Muche-Borowski & Kopp, 2011). Im Gegensatz dazu werden jedoch nur systematisch entwickelte evidenzbasierte Leitlinien, die auf den Ergebnissen der besten klinischen Evidenz basieren, als Goldstandard definiert (Gechter et al., 2013, S. 7). Auch wenn „LitBegut“ die Studienlage zur Sozialmedizinischen Begutachtung bereits beschrieben hat, versucht diese erste Teilstudie der vorliegenden Arbeit weitere Studien mit Bezug zur Sozialmedizinischen Begutachtung bei EM-Rente zu identifizieren. Die in Tabelle 11 dargestellten Hauptfragestellungen werden im Rahmen der vorliegenden Studie untersucht.

Tabelle 11: Fragestellungen der systematischen Literaturrecherche zur Identifikation wissenschaftlich fundierter Literatur zur Evidenzbasierung in der Sozialmedizinischen Begutachtung.

1. Hauptfragestellung
Wie ist der Stand der Evidenzbasierung der Sozialmedizinischen Begutachtung bei der Begutachtung von Anträgen zur EM-Rente?
2. Nebenfragestellungen
Gibt es evidente Entscheidungsalgorithmen für sozialmedizinische Gutachter?
Gibt es Lücken oder Schwächen in der Studienlage zur Sozialmedizinischen Begutachtung?

Der Gesamtfokus der vorliegenden Arbeit liegt auf der Weiterentwicklung der Qualitätssicherung in der Sozialmedizinischen Begutachtung bei EM-Renten. Infolgedessen wird die Begutachtung zum Zugang zu Rehabilitationsmaßnahmen von der Recherche ausgeschlossen. Die Fragestellungen werden getrennt nach den spezifischen Hauptindikationen (Innere Medizin/Allgemeinmedizin, Neurologie/Psychiatrie, Orthopädie/Chirurgie) betrachtet. Es ist zu erwarten, dass das bestehende Review von Strahl und Vogel (2016) mit den hier identifizierten Publikationen erweitert werden kann.

3.2 Untersuchungsdesign

3.2.1 Elektronische Datenbankrecherche

Wissenschaftliche Fachzeitschriften sowie Fachbücher sind zum größten Teil mittels wissenschaftlicher Literaturdatenbanken elektronisch verfügbar. Die Auswahl der Datenbanken orientiert sich in der Regel an dem zu untersuchenden Forschungsthema sowie am lokalen Angebot der Hochschulbibliotheken. Datenbanken enthalten bibliographischen Angaben, Kurzzusammenfassungen (Abstracts) und Referenzen eines Manuskriptes. Bei der Nutzung einer elektronischen Datenbank wird auf eine a priori definierte Liste von Suchbegriffen zurückgegriffen, die systematisch abgearbeitet wird. Die elektronische Recherche wird zusätzlich durch Thesauren unterstützt, in denen Synonyme und verwandte Suchbegriffe angezeigt werden (Döring & Bortz, 2016, S. 159-160).

Um eine möglichst umfassende Literatursuche durchführen zu können, wurde die systematische Literaturrecherche zur sozialmedizinischen Begutachtung bei EM-Renten in den Datenbanken MEDLINE, Web of Science, PsycINFO, PSYINDEX, Cochrane Central Register of Controlled Trials (CENTRAL) und der Cochrane Database of Systematic Reviews (Cochrane Reviews) durchgeführt.

MEDLINE. MEDLINE ist eine der größten Datenbanken für biomedizinische Literatur. Die Datenbank ist schwerpunktmäßig auf die Bereiche Human-, Zahn- und Veterinärmedizin, Krankenpflege, Gesundheitswesen und vorklinische Studien spezialisiert. Es werden vorrangig englischsprachige Quellen mit einem hohen Anteil US-amerikanischer Literatur gelistet. MEDLINE wird von der U.S. National Library of Medicine (NLM) betrieben und ist kostenfrei zugänglich und nutzbar (Gechter et al., 2013, S. 14).

Web of Science Core Collection (WoS CC). WoS CC ist eine von Thomson Reuters betriebene multidisziplinäre Datenbank. Im Gegensatz zu reinen Fachdatenbanken umfasst WoS CC interdisziplinäre Literaturquellen aus den Bereichen der Natur-, Sozial- und Geisteswissenschaften. Es können sowohl Primärpublikationen als auch Kongressbeiträge gesucht werden (Stock & Stock, 2003).

PsycINFO. PsycINFO ist eine von der American Psychological Association betriebene Datenbank für Psychologie sowie zu psychologierelevanten Gebieten wie Medizin, Psychiatrie, Krankenpflege, Soziologie, Erziehungswissenschaften, Pharmakologie, Physiologie, Linguistik, Anthropologie, Wirtschafts- und Rechtswissenschaften. Neben Journalartikeln beinhaltet die Datenbank auch die Auflistung von Büchern und Dissertationen aus mehr als 50 Ländern in 29 Sprachen (APA, 2018).

PSYINDEX. PSYINDEX ist eine Datenbank der Psychologie und deren verwandte Gebiete, die vom Leibniz-Zentrum für Psychologische Information und Dokumentation in Trier (ZPID) betrieben wird. Die Datenbank umfasst ausschließlich Publikationen aus dem deutschsprachigen Raum (Deutschland, Österreich, Schweiz). PSYINDEX besteht aus den Teilbereichen: „Literatur“, „AV-Medien“, „Interventionsprogramme“ und „Tests“, welche psychologische Literatur, Beschreibungen von psychologischen Interventionsprogrammen sowie Testbeschreibungen beinhaltet (ZPID, 2018).

Cochrane Central Register of Controlled Trials (CENTRAL). CENTRAL versteht sich als Datenbank zur Identifizierung von RCTs. Die Datenbank besteht aus Zitationseinträgen anderer Datenbanken, vor allem MEDLINE und EMBASE. Gleichzeitig umfasst CENTRAL auch Einträge der ‚Specialized Register‘ der Cochrane Review Groups sowie von Handsuche gefundene Artikel. Der Vorteil von CENTRAL gegenüber den o.g. genannten Datenbanken ist, dass durch die manuelle Handsuche in gedruckten Fachzeitschriften auch nationalsprachig publizierte Literatur identifiziert werden kann, die nicht in MEDLINE indexiert wird. Die Datenbank ist Teil der Cochrane Library und wird monatlich aktualisiert (Gechter et al., 2013, S. 14).

Cochrane Database of Systematic Reviews (CDSR). Systematische Reviews sind die wesentlichsten Publikationen der Cochrane Collaboration, einer Organisation, die Forschungsergebnisse systematisch organisiert und damit evidenzbasierte Entscheidungen erleichtert. Die Reviews gelten als Qualitätsstandard in der evidenzbasierten Gesundheitsversorgung und untersuchen die Wirksamkeit von Interventionen oder die Genauigkeit von Tests. Die Reviews werden extern begutachtet und regelmäßig aktualisiert. Die Abstracts und eine zusätzliche laienverständliche Zusammenfassung werden frei zur Verfügung gestellt (Cochrane Deutschland, 2018).

3.2.2 Suchstrategie

Die im Rahmen der vorliegenden Forschungsarbeit durchgeführte systematische Recherche zur Evidenzbasierung in der Sozialmedizinischen Begutachtung umfasst den Zeitraum der letzten 20 Jahre von Januar 1998 bis Oktober 2018. Die Suche erfolgte nach Freitextbegriffen und Medical Subject Headings (MeSH) in deutscher oder englischer Sprache. Der MeSH ist ein polyhierarchisches, kontrolliertes Schlagwortvokabular zur Beschreibung und Dokumentation medizinischer Literatur. Dabei werden Literaturstellen mit mehreren MeSH-Terms kodiert, welche die inhaltlichen Aspekte der Publikation verschlagworten. Weiterhin besteht die Möglichkeit, die Suche mit Hilfe von weiterführenden „Subheadings“ gezielt einzugrenzen. Aus der Kombination von

Freitextbegriffen und MeSH-Termini resultieren in der Regel die besten Suchergebnisse, da oftmals aktuelle Artikel noch nicht verschlagwortet sind (Gechter et al., 2013, S. 31).

Bei der Entwicklung einer Suchstrategie muss entweder auf Genauigkeit (Präzision) oder Vollständigkeit (Sensitivität) Wert gelegt werden. Während Sensitivität den Anteil der gefunden relevanten Literatur von allen relevanten Einträgen beschreibt, ist die Präzision der Anteil der gefundenen relevanten Datenbankeinträge von allen gefundenen Einträgen (vgl. Abb. 9).

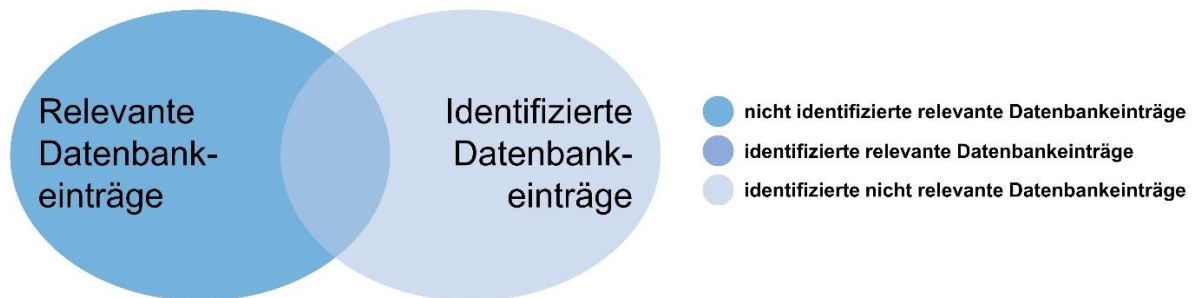


Abbildung 9: Darstellung der Sensitivität und Präzision bei der Literaturrecherche.

Durchführung der Recherche. Entsprechend der drei Hauptindikationsgruppen sozialmedizinischer Gutachten werden für die elektronische Literaturrecherche in den o.g. Datenbanken drei Suchbäume vorbereitet:

1. Sozialmedizinische Begutachtung bei Anträgen auf EM-Rente vor dem Hintergrund einer Erkrankung aus dem Bereich der Inneren Medizin/Allgemeinmedizin
2. Sozialmedizinische Begutachtung bei Anträgen auf EM-Rente vor dem Hintergrund einer Erkrankung aus dem Bereich der Neurologie/Psychiatrie
3. Sozialmedizinische Begutachtung bei Anträgen auf EM-Rente vor dem Hintergrund einer Erkrankung aus dem Bereich der Orthopädie/Chirurgie

Die zur Beantwortung der Fragestellung festgelegten Suchbegriffe wurden mit den entsprechenden Booleschen Operatoren (OR/AND) für die Datenbankrecherche zu einer Suchstrategie verknüpft. Die Freitextbegriffe wurden sowohl als Schlagwort sowie im Titel und Abstract der Publikationen gesucht. Diese Suchanfrage führte zu einer definierten Treffermenge, die in weiteren Schritten analysiert wurde (Gechter et al., 2013, S. 37). Abbildung 10 zeigt den schematischen Aufbau der Suchstrategie in Datenbanken.

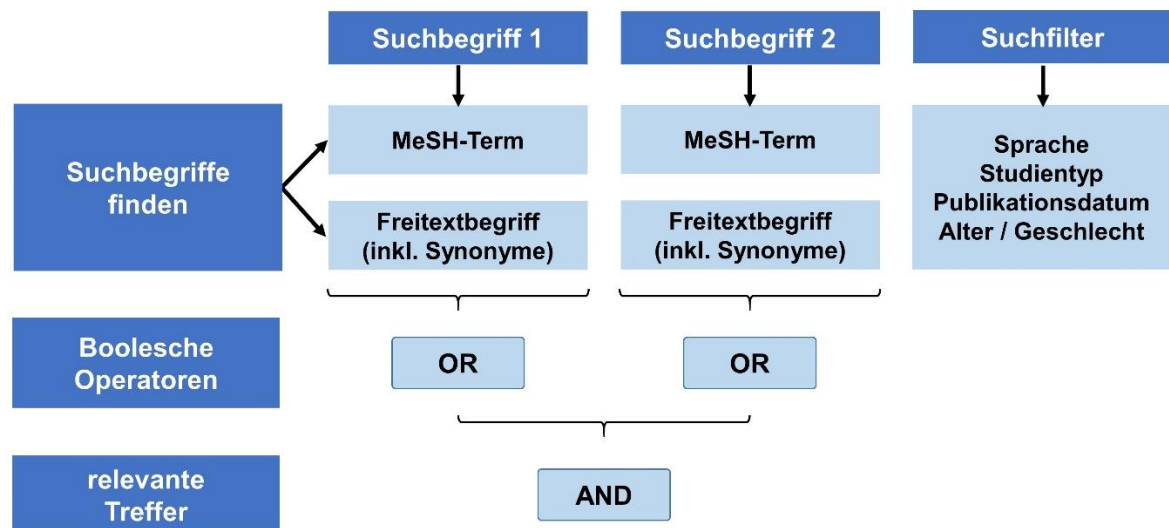


Abbildung 10: Schematische Darstellung der Suchstrategie nach Boeker und Motschall (modifiziert nach Gechter et al., 2013, S. 36).

Suchbegriffe. Die Schlagworte der Recherche wurden zu drei zentralen Modulen zugeordnet (vgl. Tab. 13). Modul A (Medizinische Hauptindikation), Modul B (Sozialmedizinische Begutachtung) und Modul C (Erwerbsminderungsrente/Funktionseinschränkung) wurden mit entsprechenden Booleschen Operatoren miteinander verknüpft. Innerhalb eines Modules wurden die Synonyme mit „OR“ und die Module miteinander mit „AND“ verknüpft (vgl. Abb. 10 und *Anhang 2*).

Die Auswahl der in Modul A aufgeführten Diagnosen orientierte sich an den häufigsten Diagnosen bei EM-Rente in der gesetzlichen Rentenversicherung des Berichtsjahres 2015 (GBE-Bund, 2018b). Für die Recherche wurde das etablierte PICO-Schema der evidenzbasierten Medizin für die zu untersuchende Fragestellung angepasst (vgl. Tab. 12).

Tabelle 12: PICO-Schema der systematischen Literaturrecherche zur Evidenzbasierung in der Sozialmedizinischen Begutachtung (Schema nach Schmucker, Motschall, Antes & Meerpohl, 2013, S.1392).

Patienten (P)	entspricht der medizinischen Hauptindikation (Modul A)
Intervention (I)	entspricht der Sozialmedizinischen Begutachtung der Funktionseinschränkungen bei Anträgen zur Erwerbsminderungsrente (Modul B/C)
Comparison (C)	entfällt, da kein Wirksamkeitsvergleich von Interventionen
Outcome (O)	entfällt, da kein spezieller klinischer Endpunkt untersucht wird

Tabelle 13: Schlagworte und Synonyme der elektronischen Datenbankrecherche.

Modul A Medizinische Hauptindikation	Modul B Sozialmedizinische Begutachtung	Modul C Erwerbsminderungsrente/ Funktionseinschränkung
Innere Medizin/ Allgemeinmedizin Neubildungen, neoplasm, Krebs cancer, tumor, tumour, carcinoma, Schlaganfall, stroke, apoplexy, COPD, CED, KHK, chronic ischemic heart disease, Diabetes mellitus, Arteriosklerose, arteriosclerosis, Kardiomyopathie, cardiomyopathy, Herzinsuffizienz, heart failure, heart insufficiency, cardiac insufficiency Neurologie/Psychiatrie psychische Störung, psychological disorder, psychiatric disorder, mental disorder, affektive Störung, affective disorder, mood disorder, Depression, Psychische und Verhaltensstörungen durch Alkohol, Psychological and behavioural disorders resulting from psychotropic substances, alcohol abuse, somatoforme Störung, somatoform disorder, Schizophrenie, schizophrenia, Angststörung, anxiety disorder, Multiple Sklerose, multiple sclerosis Orthopädie/Chirurgie Arthrose, arthrosis, joint disease, osteoarthritis, Rückenschmerzen, back pain, Spondylopathien, spondylopathies, Lumboischialgie, sciatica, radiculopathy, lumboischialgia	Assess* Begutachtung claim assessment disability assessment disability evaluation employee performance appraisal functional capacity assessment functional capacity evaluation functional evaluation functioning-based assessment impairment assessment impairment evaluation impairment of earning capacity assessment impairment of earning capacity evaluation incapacity assessment insurance medicine judgement Leistungsbeurteilung medical examination needs assessment physical examination practice guideline socio-medical assessment socio-medical evaluation symptom assessment symptom evaluation work capacity evaluation	compensation disability benefit disability pension disability retirement pay early retirement functional capacity pension social security benefit work disability

* Trunkierungszeichen: Abschneiden des Wortes zur Suche beliebig vieler Zeichen nach dem Wortstamm

Ein- und Ausschlusskriterien. Eingeschlossen wurden ausschließlich Publikationen, die sich thematisch mit der Begutachtungspraxis bei Anträgen auf EM-Rente auseinandersetzen. Ein Schwerpunkt liegt auf dem Ablauf des Begutachtungsprozesses per se und evidenzbasierten Entscheidungsalgorithmen. Üblicherweise werden zur besseren Vergleichbarkeit der Studien nur RCTs, kontrollierte Studien oder

Beobachtungsstudien in systematische Übersichtsarbeiten eingeschlossen. Unter der aus den Ergebnissen des „LitBegut“-Projekt begründeten Annahme, dass auf dem Gebiet der sozialmedizinischen Begutachtung wenig oder keine randomisierten Studien zur Verfügung stehen, wurde keine Eingrenzung der Literatur auf spezifische Studiendesigns vorgenommen. Dieses Vorgehen spiegelt auch den Anspruch der Arbeit zur Abbildung aller zur Verfügung stehenden Evidenzen wider. Eingeschlossen wurden nur Primärstudien in deutscher und englischer Sprache, die in wissenschaftlichen Zeitschriften publiziert wurden.

Publikationen wurden vom systematischen Review ausgeschlossen, wenn (A) kein Bezug zur Beurteilung der Leistungsfähigkeit und (B) kein Bezug zu EM-Rente erkennbar war. Diese Kriterien implizieren den Ausschluss folgender Studien:

- diagnostische Studien,
- Evaluationsstudien von Interventionen,
- Validierungsstudien von Assessmentinstrumenten,
- Prädiktor- und Korrelationsstudien für „Disability“,
- Prävalenzstudien,
- Studienprotokolle,
- Studien über Berufskrankheiten,
- Therapiewirksamkeitsstudien bei Patienten mit Erwerbsminderung.

Darüber hinaus wurden Publikationen ausgeschlossen, die bereits im Rahmen des Vorgängerprojektes „LitBegut“ identifiziert wurden. Beim Ausschluss von Studien bestand generell die Schwierigkeit bei der Differenzierung zwischen der allgemeinen Begutachtung von Behinderung, Beeinträchtigung und Invalidität, im englischen subsummiert mit dem Begriff „disability“ und der Begutachtung bei Anträgen auf EM-Rente, im englischen subsummiert mit dem Begriff „disability pension“.

Datenextraktion und -synthese. Im Anschluss an die Recherche wurden alle Datenbankergebnisse in das Literaturverwaltungsprogramm Mendeley (Mendeley Ltd.) übertragen. Nach der Entfernung von Duplikaten, wurden die verbliebene Beiträge in einer Vorauswahl anhand des Titels und des Abstracts auf Eignung geprüft. Die restlichen Literaturstellen wurden anhand ihrer Volltexte beurteilt.

Die Datenextraktion und -synthese entsprachen einer Kombination aus Evidence Mapping und Scoping Review. Evidence Maps bilden systematisch vorhandene und fehlende Evidenzen für medizinische Fragestellungen ab. Dazu wird eine Übersichtstabelle zum „aktuellen Forschungsstand im Hinblick auf die Studienanzahl, Studiencharakteristika (z. B. Studiendesign, geografische Angaben zur Studie), Patientencharakteristika und Studienergebnisse“ erstellt (Schmucker, Motschall, Antes & Meerpohl, 2013, S. 1391). Ergänzend wird im Scoping Review die Literatur deskriptiv narrativ zusammengefasst.

3.2.3 Beurteilung der Studienqualität

Durch den Einschluss aller Studiendesigns ergaben sich bei der studienübergreifenden Qualitätsbeurteilung Hindernisse, wie z. B. unterschiedliche, nicht vergleichbare Qualitätsbeurteilungstools (Schmucker et al., 2013). Es gibt jedoch zahlreiche Belege dafür, dass viele Studien methodisch schwach sind und vorhandene Mängel zu verzerrten Ergebnissen systematischer Übersichtsarbeiten führen können. Die Bewertung der methodischen Qualität sollte daher in systematischen Übersichtsarbeiten immer durchgeführt werden (Jüni, Altman & Egger, 2001).

In der vorliegenden Arbeit wurde das Mixed Methods Appraisal Tool (MMAT) für systematische Literaturrecherchen ohne Einschränkung des Studiendesigns verwendet. Das MMAT ist ein reliables Beurteilungsinstrument für mixed-methods Reviews. Die Interrater-Reliabilität des globalen Qualitätsergebnisses ist moderat (ICC 0,72) bis sehr gut (ICC 0,94). Es enthält fünf spezifische Beurteilungsdomänen für qualitative Studien, randomisiert-kontrollierte quantitative Studien, nicht randomisierte Vergleichsstudien, Beobachtungsstudien sowie mixed-methods Studien (Pace et al., 2012).

Narrative Reviews sind unsystematisch herangezogene, teilweise durch den Autor subjektiv geprägte Übersichtsarbeiten. Obgleich die Publikationen einen schnellen Überblick über ein Thema oder eine bestimmte Fragestellung ermöglichen, ist die Evidenz und die Qualität begrenzt. Durch die unsystematische Auswahl, Aufbereitung und Darstellung der eingeschlossenen Publikationen sind systematische Verzerrungen innerhalb der Ergebnisse nicht ausgeschlossen (Krause, Khan, Antoch, Kotzerke & Hosten, 2015; Scheidt et al., 2018). Da keine standardisierten Vorgaben für narrative Reviews existieren, ist auch eine einheitliche Qualitätsbeurteilungen dieser Studien nur begrenzt möglich.

3.2.4 Beurteilung der Evidenz

Zur Beurteilung der studienzugehörigen Evidenz wurde ein Schema in Anlehnung an Shekelle et al. (1999) verwendet. Das Schema teilt wissenschaftliche Beiträge anhand ihres Studiendesigns in vier hierarchische Evidenzklassen ein:

- Ia:** Evidenz durch Metaanalysen oder systematische Reviews von randomisiert-kontrollierten Studien
- Ib:** Evidenz durch zumindest eine randomisiert-kontrollierte Studie
- Ila:** Evidenz durch zumindest eine kontrollierte Studie (ohne Randomisierung)
- IIb:** Evidenz durch zumindest eine andere quasi-experimentelle Studie
- III:** Evidenz durch nicht-experimentelle, deskriptive Studien (z. B. Beobachtungsstudien, Korrelationsstudien)
- IV:** Evidenz durch Expertengruppen oder durch klinische Erfahrung erfahrener und anerkannter Personen.

Ausgehend von den Ergebnissen von Strahl und Vogel (2016) ist fraglich, ob die Übertragbarkeit der Methodik evidenzbasierter Medizin auf die besonderen Anforderungen der Sozialmedizinischen Begutachtung möglich sei. Eine ähnliche Problematik wird auch in der evidenzbasierten Prävention und Gesundheitsförderung (PGF) beschrieben.

Auch in der PGF sind Studien im RCT Design nicht immer sinnvoll umzusetzen, da geeignete Kontrollgruppen häufig fehlen bzw. nicht zu definieren sind. Die PGF fordert daher, die verstärkte Ausdifferenzierung und Aufwertung von Beobachtungsstudien. Entsprechend sollten für Fragestellungen, die durch ein randomisiert, kontrolliertes Studiendesign nicht adäquat beantwortet werden können, das RCT nicht automatisch als Ideal der Evidenz angesehen werden (Kliche, Koch, Lehmann & Töppich, 2006). Die Autoren (ebd.) empfehlen demgemäß eine modifizierte Anordnung der Evidenzklassen in Anlehnung an die Konventionen der Cochrane Collaboration (vgl. Tab. 14).

Tabelle 14: Adaptierte Evidenzklassen in PGF (modifiziert nach Kliche et al., 2006, S. 147).

Evidenz- klasse	Studienbeschreibung
I	• Meta-Analyse, systematisches Review möglichst hochwertiger Studien
II	• RCT
III	• clusterrandomisierte Studien (CRT) • nicht randomisierte experimentelle Studie • Beobachtungsstudien mit Kontrollgruppe • Beobachtungsstudie ohne Kontrollgruppe mit Mehrpunktmessung (mindestens vier Messzeitpunkte) und mehrdimensionaler Erhebung • Beobachtungsstudie ohne Kontrollgruppe mit multimethodischer Triangulation (qualitative Verlaufsforschung zur Verlaufs- und Wirkfaktoren)
IV	• Beobachtungsstudie ohne Kontrollgruppe (Querschnittsstudie, Prä-Post-Studie, Fallserien) • Strukturierter Expertenkonsens (Pluralität aller Perspektiven bei der Expertenauswahl, transparente Konsensregeln, Protokollierung und systematische Bearbeitung von Kontroversen und Mindermeinungen, Verknüpfung der Konsensergebnisse mit empirischem Kenntnisstand)
V	• Fallberichte, Expertenmeinungen, unstrukturierter Expertenkonsens

Aus den o.g. Gründen verwendete das hier vorliegende systematische Review zum Stand der Evidenzbasierung der Sozialmedizinischen Begutachtung im Folgenden zur Beurteilung der Evidenz das adaptierte Evidenzschema nach Kliche et al. (2006).

3.3 Ergebnisse

Das Review wurde im Einklang mit den PRISMA-Guidelines im Oktober 2018 durchgeführt (siehe *Anhang 3*). Durch die Datenbankrecherche wurden $N = 17.079$ Publikationen identifiziert. Nach dem Ausstreichen von Duplikaten ($n = 5.059$) wurden 12.020 Dokumente in die Vorauswahl aufgenommen. Nach Ausschluss von $n = 11.916$ Referenzen anhand von Titel und Abstract wurden $n = 104$ Referenzen zur Volltextsichtung ausgewählt, von denen $n = 88$ ausgeschlossen wurden, da sie die Selektionskriterien nicht erfüllten. Die Gründe für die Ausschlüsse sind in der Tabelle im *Anhang 4* zusammengefasst. Nach der Eignungsbeurteilung der Volltexte verblieben 16 Dokumente, welche für die Darstellung der Evidenz in der Sozialmedizinischen Begutachtung bei EM-Renten von Relevanz waren und in die Auswertung einbezogen wurden (Abb. 11).

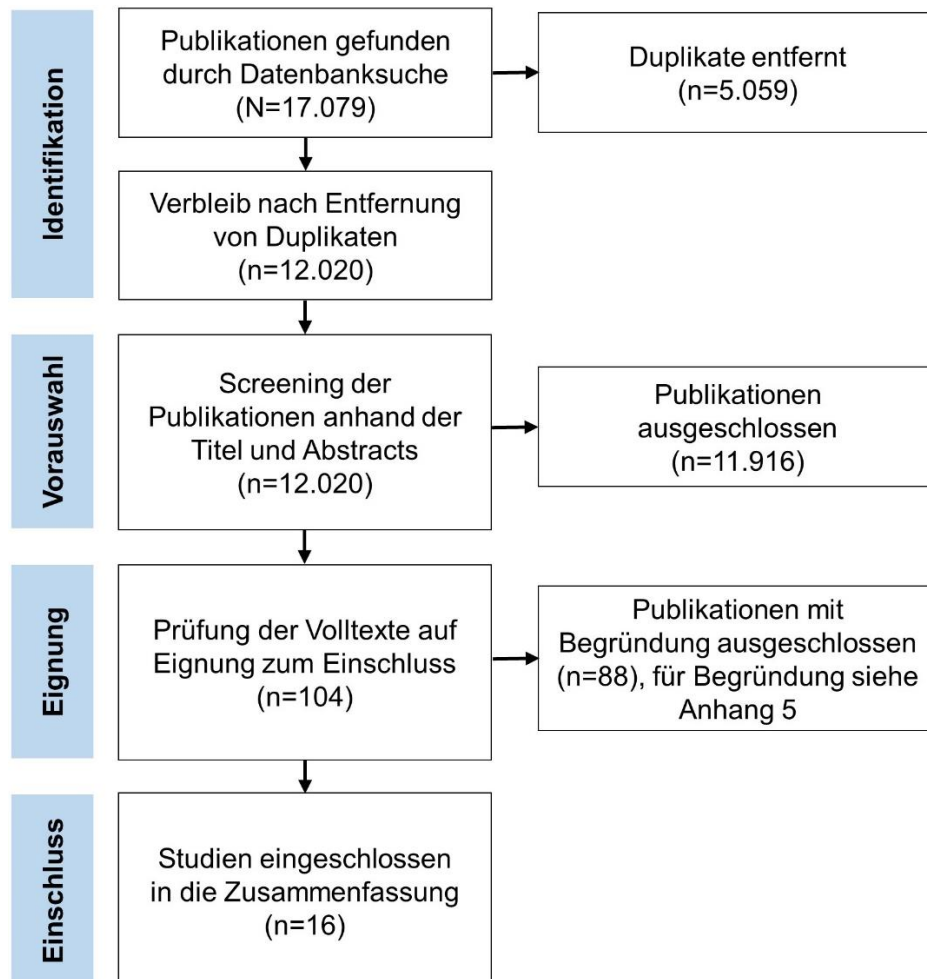


Abbildung 11: Flussdiagramm der verschiedenen Phasen der systematischen Recherche (modifiziertes PRISMA-Diagramm nach Moher, Liberati, Tetzlaff & Altman, 2011, S. e11).

Identifiziert wurden ein systematisches Literaturreview, fünf deskriptive Beobachtungsstudien und zehn narrative Reviews. Die Mehrzahl der eingeschlossenen Studien stammten aus Deutschland ($n = 10$) und Kanada ($n = 2$). Vier weitere Studien wurden in den USA, Niederlanden, Schweiz und Australien angefertigt.

3.3.1 Merkmale und inhaltliche Charakteristika der eingeschlossenen Studien

Bezüglich ihres Themenspektrums weisen die identifizierten Publikationen eine hohe Heterogenität auf. Sie lassen sich weitestgehend in drei Themenbereiche einteilen: (1) allgemeine und indikationsspezifische Hinweise zur Begutachtung, (2) ICF und ICF Core Sets sowie (3) Begutachtung aus Perspektive der gutachterlich tätigen Ärzte.

3.3.1.1 Hinweise zur Begutachtung

Die Publikationen der ersten inhaltlichen Themengruppe enthielten primär Publikationen mit allgemeinen oder indikationsspezifischen Hinweisen zur Durchführung einer Leistungsbegutachtung bei Beantragung einer EM-Rente. Unter den elf Publikationen dieser Kategorie befanden sich eine systematische Literaturrecherche (Canela et al., 2016), eine Beobachtungsstudie (Dell-Kuster et al., 2014) und neun narrative Reviews (Backman, 2004; Canela et al., 2015; Dodgshun & Malios, 2011; Hammer, 2002; Hoffmann, Vahlensieck & Zermann, 2016; Kellner, 2002; Kolenda, 1998; Rogers & Payne, 2006; Schiltenswolf, 2007).

In ihrer systematischen Literaturrecherche untersuchten Canela und Kollegen (2016) Testinstrumente zur Messung der Funktionsfähigkeit von Patienten mit psychischen Störungen in der Sozialmedizinischen Begutachtung. Insgesamt wurden 19 Erhebungsinstrumente zur Messung der relevanten Funktionen identifiziert. In ihrer abschließenden Beurteilung der Ergebnisse kommen die Autoren jedoch zu dem Schluss, dass keines dieser Instrumente die grundsätzlichen Testgütekriterien (Objektivität, Reliabilität, Validität) erfüllt. In der Konsequenz der Ergebnisse wird für die gutachterliche Praxis kein Instrument für die Leistungsbeurteilung psychisch Kranker empfohlen.

In der Publikation von Backman (2004) zur Begutachtung von Patienten mit RA wurden sowohl Risikofaktoren der Arbeitsunfähigkeit, Möglichkeiten zur Erhebung von arbeitsbezogenen Problemlagen sowie Interventionen zur Prävention von Erwerbsminderung beschrieben. Zur Erhebung arbeitsbezogener Problemlagen empfiehlt der Autor die Work Instability Scale (WIS) und den Work Limitations Questionnaire (WLQ). Der WIS ist ein 23 Item umfassendes Screening-Instrument zur Erhebung des Risikos des Arbeitsplatzverlustes infolge einer Erkrankung, dessen Testgütekriterien jedoch nicht hinreichend untersucht wurden. Der zweite empfohlene Fragebogen ist der 25 Items umfassende WLQ zur Messung des Einflusses chronischer Erkrankung auf die Arbeitsleistung und -produktivität. Die Testgütekriterien seien in früheren Evaluierungsstudien für diesen Fragebogen demonstriert worden. Für beide Assessmentinstrumente werden keine konkreten Angaben über die erreichten Testgütekriterien gemacht. Der Artikel endet mit der Empfehlung des Einsatzes der beiden Instrumente, um Erwerbsunfähigkeit gezielt identifizieren zu können.

In ihrem narrativen Review zur invalidenversicherungsrechtlichen Begutachtung in der Schweiz beschrieben Canela und Kollegen (2015) neben den Besonderheiten der schweizerischen Invaliditätsgesetzgebung anhand von rechtlichen Beispielen die Begutachtungsbesonderheiten von psychosozialen Faktoren und unklaren Beschwerden. Entgegen des biopsychosozialen Modells der WHO werden bei der Invaliditätsbemessung in der Schweiz soziale Komponenten ausgeklammert. Infolgedessen werden „Invaliditätsfremde“ Faktoren (Beeinträchtigungen durch Alter, Schulbildung, schlechte Sprachkenntnisse, schlechte Arbeitsmarktlage) bei der Feststellung von Invalidität nicht berücksichtigt. Zudem wird angenommen, dass psychosoziale Belastungen allein keine „Invalidität“ verursachen. Diese Annahme wird damit begründet, dass Leistungsminderungen bei neurotischen, somatoformen und Belastungsstörungen häufig durch äußere Umstände verursacht seien und bei einer zumutbaren Veränderung der Verhältnisse verschwinden würden. In der Begutachtung müssen Experten entsprechend zwischen psychosozialen und medizinischen Faktoren differenzieren und diese gewichten. Dabei ist die Genauigkeit der Erhebung psychosozialer Faktoren abhängig von der Zeit, die für das Gutachten zur Verfügung steht, dem Gutachter (Persönlichkeit/Fragetechnik) und dem Rentenantragsteller (Persönlichkeit) selbst. Bei anhaltenden somatoformen Schmerzstörungen sollen zusätzlich Komorbiditäten, chronische körperliche Begleiterkrankungen, sozialer Rückzug, primärer Krankheitsgewinn sowie unzureichende bisherige ambulante und stationäre Behandlungsergebnisse erhoben werden. Darüber hinaus sind die diagnostischen Kriterien einiger schmerzassoziierter Erkrankungen wie z. B. Fibromyalgie und chronisches Erschöpfungssyndrom oder Depression und chronisches Erschöpfungssyndrom überlappend. In der Begutachtung der Differentialdiagnostik wird ihnen eine besondere Bedeutung zugesprochen. Die Autoren schlussfolgern, dass langfristig Funktionsbeschreibungen, z. B. mit dem Mini-ICF zur Beschreibung von Funktionsstörungen bei psychischen Störungen, eine höhere Gewichtung in der Begutachtung haben sollte.

Dodgshun und Malois (2011) entwickelten eine Anleitung zur Beantragung einer finanziellen Arbeitsentschädigung für Allgemeinmediziner in Australien. Die Autoren führen potenzielle Arbeitseinschränkungen auf, die während der körperlichen Untersuchung in der Sozialmedizinischen Begutachtung geprüft werden sollen. Je nach betroffener Körperregion vermeiden die Patienten das körpernahe Heben von Gegenständen bis zu 10 kg, langes Stehen, kräftiges Schieben oder Ziehen, Bücken des Rückens über einen bestimmten Bereich hinaus, wiederholtes Bücken, Kniebeugen, Verwendung eines Arms über der Schulterhöhe, die wiederholte Verwendung eines Armes oder den Kontakt mit Ölen oder Lösungsmitteln. Typische Fallstricke in der Begutachtung seien zudem das Übersehen von psychosozialen Problemlagen und psychischen Störungen, die

unzureichende Beschreibung der Auswirkung von Schmerzen auf die Aktivitäten des täglichen Lebens und des Berufs als auch die nicht ausreichend durchgeführte Bewertung des Bedarfs an physikalischen oder psychologischen Therapien. Darüber hinaus wurde herausgestellt, dass die Kommunikation zwischen Arzt und Begutachteten bei der Erfassung der Erwerbsminderung besonders zentral ist, um die subjektiven Erwartungen des Antragstellers bezüglich der Rückkehr an den Arbeitsplatz detailliert zu erfassen.

Hammer (2002) beschrieb in einem narrativen Review die zur Einschätzung der Leistungsfähigkeit notwendigen Untersuchungen und Messinstrumente am Beispiel der Begutachtung von Patienten mit RA. Grundsätzlich verweist die Publikation auf die Hinweise zur Begutachtung für die gesetzliche Rentenversicherung (vgl. Kapitel 1.5.2). Es werden darüber hinaus spezifische Vorgehensweisen für die Begutachtung von Menschen mit RA genannt (anamnestische Erhebung von Rheumaknoten und Hautveränderungen, Beweglichkeits- und Schmerzprüfung aller betroffenen Gelenke, obligate Wirbelsäulenuntersuchung, Gehstreckenbestimmung, Funktionsprüfung von Komplexfunktionen, wie Faustschluss, Pinzettengriff und Griffstärke). Für die Beurteilung der systemischen Entzündungsreaktion sollten zudem im Labor Entzündungsparameter (C-reaktives Protein, Blutsenkungsgeschwindigkeit) und rheumatologisch-immunologische Parameter (Rheumafaktoren) im Blutserum bestimmt und ein allgemeines Blutbild sowie ein Differentialblutbild erstellt werden. Zur Einschätzung der eingetretenen Destruktivität der Gelenke und Wirbelkörper sind Röntgenuntersuchungen der auffälligen Gelenke obligat. Laut Hammer (2002) gibt es keine festgelegte Leistungsdiagnostik für Patienten mit rheumatischen Erkrankungen, da beispielsweise deutliche Abnutzungserscheinungen in Röntgenbildern nicht zwangsläufig mit individuellen Einschränkungen zusammenhängen. Entsprechend beruht die Beurteilung der Leistungsfähigkeit bei RA sowohl auf klinisch messbaren Anhaltspunkten als auch auf subjektiv wahrgenommenen Beeinträchtigungen seitens des Versicherten. Zur Beeinträchtigungvalidierung wurden sieben Messinstrumente (u.a. Funktionsfragebogen Hannover (FFbH-RA), Health Assessment Questionnaire (HAQ), Disease Activity Score 28 (DAS28)) benannt.

Kellner (2002) ergänzte die Literatur über die Begutachtung rheumatischer Erkrankungen mit seinem Beitrag über die Leistungsbeurteilung von Patienten mit seronegativen Spondylarthropathien (Morbus Bechterew). Für die Diagnostik des Rückenschmerzes des entzündlichen Typus müssen vier der fünf Diagnosekriterien (Krankheitsbeginn vor dem 40. Lebensjahr, schleichender Beschwerdebeginn, Dauer länger als drei Monate, Morgensteifigkeit, Besserung durch Bewegung) erfüllt sein. Anamnese und klinische Untersuchung wurden als die wichtigsten Säulen der klinischen Begutachtung beschrieben, wobei auch validierte Fragebögen und Assessments wie der HAQ, der Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), der Bath Ankylosing Spondylitis Functional

Index (BASFI), der Bath Ankylosing Spondylitis patient Global score (BAS-G), der Bath Ankylosing Spondylitis Metrology Index (BASMI) sowie der Bath Ankylosing Spondylitis Radiology Index (BASRI) eingesetzt werden sollten. Bei der klinischen Untersuchung sind spezielle Funktionsuntersuchungen (Schober'sches Zeichen, Ott'sches Zeichen, Fingerbodenabstand, Hinterhauptwandabstand, Differenz der maximalen Thoraxexpansion bei In- und Expiration) anzuwenden, sowie die Klopfschmerzhaftigkeit der Wirbelsäule und der Gelenkstatus aller peripheren Gelenke zu prüfen. Laboruntersuchung und bildgebende Verfahren ähneln der Untersuchung von Patienten mit RA. Die abschließende Beurteilung der Leistungsfähigkeit orientiert sich bei seronegativen Spondylarthropathien primär an den Funktionen und der strukturellen Veränderungen der Wirbelsäule und Gelenke. Der Artikel beschreibt Orientierungskriterien, die aufsteigend für einen bestimmten Grad der EM bzw. Grad der Behinderung (GdB) stehen:

- 10-20 % Einschränkung: Inaktives/wenig aktives Frühstadium mit nur geringer, reversibler, kompensierbarer Funktionseinschränkung und Beschwerden
- 30-50 % Einschränkung: Irreversible strukturelle Funktionseinschränkungen mäßigen Grades bei chronischer bzw. fortgeschrittener Erkrankung mit persistierender und/oder variabler Aktivität mit überwiegend kompensierter Statik
- 60-90 % Einschränkung: Hochgradig progrediente, die Statik beeinflussende und mit irreversiblen Funktionseinbußen einhergehende Wirbelsäulenversteifung in Fehlform mit oder ohne kompensierbare oder irreversible funktionsgeminderte Begleitarthritis bei fortgeschrittener Erkrankung mit persistierender oder variabler Aktivität
- 90-100 % Einschränkung: Endstadium der Erkrankung mit aufgehobener Funktion der Wirbelsäule und/oder destruierender Arthritis peripherer Gelenke mit Beteiligung viszeraler Organe (Kellner, 2002, S. 650).

Der Ablauf der Sozialmedizinische Begutachtung bei Patienten mit Prostatakarzinom wird von Hoffmann, Vahlensieck und Zermann (2016) beschrieben. Neben der Darstellung von zu beachtenden Begleitsymptomen und Komorbiditäten (wie z. B. Harninkontinenz, Blasenfunktionsstörung, sexuelle Funktionsstörungen, Hitzewallungen, Fatigue-Syndrom, Depression, Anämie, kognitive Funktionsstörungen bei hormonablativen Therapien, Gynäkomastie und Brustschmerz bei antiandrogenen Monotherapien, Osteoporose bei Hormontherapien, Stoffwechselstörungen sowie Schmerz und Dysästhesie bei Chemotherapie) werden konkrete Anhaltspunkte für die Abschätzung des positiven und negativen Leistungsbildes bei Prostatakarzinom genannt. Nach dem Modell wird der zeitliche Umfang des Restleistungsvermögens (6 Std. und mehr, 3-6 Std. und mehr, 3 Std. und weniger) determiniert durch die Art der Behandlung (kurativ oder palliativ), dem Tumorstadium (lokal, fortgeschritten lokalregionär oder metastasiert), der noch zu

verrichtenden Arbeitsschwere und Arbeitshaltung, der Arbeitsorganisation (Schichtdienst), psychomentalen Funktionen, Sinnesfunktionen, bewegungsbezogenen Funktionen, kardiopulmonale Funktionen sowie relevante Gefährdungs- und Belastungsfaktoren (Kontextfaktoren).

Kolenda (1998) gab in seinem narrativen Review einen Überblick über die Sozialmedizinische Beurteilung und Beratung von Patienten mit koronarer Herzkrankheit. Wichtige Bestandteil der Begutachtung dieses Versichertenkollektivs sind die Anamnese, körperliche Untersuchung, Ruhe-EKG, Echokardiographie, Belastungs-EKG, Langzeit-EKG, Lungenfunktionsuntersuchung sowie Röntgen-Thorax-Befund. Zudem ist die Erhebung von relevanten Begleiterkrankungen, wie z. B. Diabetes mellitus, periphere arterielle Verschlusskrankheit, zerebrale Ischämien, COPD und Schlafapnoe, bedeutsam. Ferner sind bei der Beurteilung die körperliche und psychische Belastbarkeit sowie die geistige Anpassungsfähigkeit zu berücksichtigen. Der Autor (ebd.) nennt sieben konkrete Faktoren, die eine berufliche Wiedereingliederung beeinflussen und entsprechend beachtet werden müssen: medizinische Befunde, Alter, Berufsgruppenzugehörigkeit, psychosoziale Faktoren, Einstellung der Arbeitgeber, Zeitdauer der Abwesenheit vom Beruf, Situation am Markt. Abschließend entwickelte Kolenda (1998) eine synoptische Darstellung des Zusammenhangs zwischen den erhobenen medizinischen Befunden (aus Ergometrie, Echokardiographie, Langzeit-EKG), der Leistungsbeeinträchtigung (keine, bei leichten Belastungen, bei mittelschweren Belastungen, in Ruhe), der zumutbaren beruflichen Tätigkeit (keine Tätigkeit möglich, leichte Arbeiten, mittelschwere Arbeiten), der möglichen Teilnahme an einer ambulanten Herzgruppe und dem voraussichtlich daraus resultierenden GdB:

- 0-10 GdB: keine Einschränkung in der Ergometrie, keine bis leichte Funktionsstörungen im Echokardiogramm, keine relevanten Rhythmusstörungen im Langzeit-EKG, keine Leistungsbeeinträchtigungen im Alltag, leichte und mittelschwere Arbeiten noch zumutbar,
- 20-40 GdB: bis 100 Watt Belastbarkeit in der Ergometrie, leichte bis mäßige linksventrikuläre Funktionsstörungen in der Echokardiographie, keine malignen Rhythmusstörungen im Langzeit-EKG, Leistungsbeeinträchtigung bei mittelschwerer Belastung, leichte Arbeiten noch zumutbar,
- 50-80 GdB: bis 50-75 Watt Belastbarkeit in der Ergometrie, Dilatation des linken Ventrikels, mäßige bis ausgeprägte systolische linksventrikuläre Funktionsstörungen im Echokardiogramm, höhergradige ventrikuläre Rhythmusstörungen, ggf. Vorhofflimmern im Langzeit-EKG, Leistungsbeeinträchtigung bei leichter Belastung, häufig keine berufliche Tätigkeit mehr zumutbar,

- 90-100 GdB: Ergometrie nicht durchführbar, ausgeprägte Dilatation des linken Ventrikels, ausgeprägte systolische linksventrikuläre Funktionsstörung im Echokardiogramm, maligne ventrikuläre Rhythmusstörung möglich, ggf. Vorhofflimmern im Langzeit-EKG, Leistungsbeeinträchtigung bereits in Ruhe, keine berufliche Tätigkeit mehr möglich.

Die Begutachtung von muskuloskelettalen Schmerzen wurde im narrativen Review von Schiltenswolf (2007) näher beschrieben. Der Artikel basiert auf den Inhalten einer interdisziplinären AWMF S2-Leitlinie. Bei Begutachtung von Schmerzen sollte das biopsychosoziale Schmerzverständnis zugrunde gelegt werden. Grundsätzlich kann Schmerz aus gutachterlicher Sicht eine Begleitsymptomatik einer Schädigung des Nervensystems oder anderer Gewebsarten, eine Gewebeschädigung mit psychischer Komorbidität oder ein Leitsymptom einer psychischen Erkrankung sein. Neben einer ausführlichen Anamnese (insb. Schmerzanamnese) sollte die klinische und apparative Diagnostik bei Verdacht auf psychischer Schmerzgenese durch ein strukturiertes klinisches Interview (wie z. B. dem Strukturiertes Klinisches Interview für DSM-IV [SKID], dem Composite International Diagnostic Interview [CIDI] oder dem Diagnostischem Interview bei psychischen Störungen [Mini-DIPS] ergänzt werden. Nach Aussage des Autors hängt die abschließende Beurteilung von Schmerzen von der zugrunde liegenden Gesundheitsstörung ab. Die Begutachtung physischer Schmerzsyndrome hat die Art der Grunderkrankung, den Schweregrad der Erkrankung, die Störungen der Aktivität und Partizipation sowie die Zumutbarkeit von weiteren Interventionen zu beachten. Die Beurteilung des Schweregrades von Schmerzen aufgrund von psychischen Störungen bedarf dagegen der Erhebung der gesicherten Diagnose der Grunderkrankung, der weiteren Komorbiditäten, der psychosozialen Auswirkungen, dem Grad der Chronifizierung, die Art der Vorbehandlung sowie die Angemessenheit einer psychotherapeutischen Behandlung. Zudem ist bei der Begutachtung von psychischen Störungen auf symptomverstärkende Darstellungsformen zu achten. Dies beinhaltet die Prüfung auf Simulation (d.h. das bewusste Vortäuschen einer Störung), Aggravation (d.h. die bewusste verschlimmernder Darstellung einer Störung) und Verdeutlichungstendenzen (d.h. die Tendenz den Gutachter vom Vorhandensein einer Störung überzeugen zu wollen).

Ergänzend hierzu, beschrieben Rogers und Payne (2006) in einem narrativen Review empirisch validierte Nachweisstrategien zur Detektion von vorgetäuschten Erkrankungen bei Antragsteller zur finanziellen Kompensation. Im Gegensatz zu allgemeineren Übersichtsarbeiten, beschreibt die Publikation vier Erhebungsinstrumente zum Einsatz bei Verdacht auf symptomverzerrenden Darstellungen. Die Autoren (ebd.) beschreiben das Minnesota Multiphasic Personality Inventory-2 (MMPI-2), das Personality Assessment Inventory (PAI), das Structured Interview of Reported Symptoms (SIRS) und das Trauma

Symptom Inventory (TSI). Im Falle einer beantragten finanziellen Kompensation bei Erwerbsminderung empfehlen Rogers und Payne (2006) den kombinierten Einsatz der F-Psychiatrie Skala des MMPI-2 und eines strukturierten Interviews, wie des SIRS. Die Identifikation beruht auf der Strategie seltene psychische Symptome zu identifizieren. Diese werden häufig von Patienten verwendet, die sich der Seltenheit dieser jedoch nicht bewusst sind und daher ein unerwartetes hohes Ausmaß an Symptomen berichten. Diese Erkennungsstrategie kann durch unwahrscheinliche Muster der Psychopathologie auf dem PAI ergänzt werden.

Die Publikation von Dell-Kuster und Kollegen (2014) untersuchte die Übereinstimmung der sozialmedizinischen Urteile zwischen Experten, behandelnden Ärzten und Antragstellern und zieht daraus indirekt Implikationen für die Begutachtungspraxis. In einer indikationsübergreifenden monozentrischen, quantitativen Querschnittsstudie zur kategorialen Interrater-Reliabilität wurde die Übereinstimmung der Leistungsfähigkeitsbeurteilungen zwischen Antragstellern, behandelnden Ärzten und einem medizinischen Expertenteam analysiert. Es wurden die konsekutiven Daten von 3.462 Antragstellern von EM-Renten im Zeitraum von Januar 2005 bis Dezember 2008 in die Studie aufgenommen. Die Leistungsfähigkeit wurde für die letzte Tätigkeit und auf dem allgemeinen Arbeitsmarkt in Prozentwerte von 0 % (nicht arbeitsfähig) bis 100 % (voll arbeitsfähig) bewertet und im Anschluss zu Kategorien (≤ 30 % = starke Einschränkung, 31-50 % = moderate bis starke Einschränkung, 51-79 % = milde bis moderate Einschränkung, 80-99 % = milde Einschränkung, 100 % = keine Einschränkung) zugeordnet. Die Autoren fanden erhebliche Unterschiede zwischen dem sozialmedizinisch erfahrenem, multidisziplinären Expertenteam und den Antragstellern sowie deren behandelnde Ärzte. Die Funktionsfähigkeit wurde seitens des Expertenteams im Mittel stets höher eingeschätzt als vom Antragsteller selbst oder dessen behandelnden Arzt (prozentuale Übereinstimmung entspricht für die Funktionsfähigkeit im letzten Beruf 51 %). Die Übereinstimmung bezüglich der Funktionsfähigkeit auf dem allgemeinen Arbeitsmarkt war mit 15 % zwischen Experten und Antragstellern und 20% zwischen Experten und behandelnden Ärzten bedeutend geringer. Für die Begutachtungspraxis bedeuten diese Ergebnisse, dass ein Antragsteller, der zuvor von seinem behandelnden Arzt bei der Beantragung der EM-Rente voll oder teilweise unterstützt wurde, bei negativen Gutachterbescheid voraussichtlich weniger bereit ist, sich mit Engagement in das Erwerbsleben zu integrieren. In der Konsequenz forderten die Autoren einerseits eine sorgfältige und standardisierte Sozialmedizinische Begutachtung, um die Akzeptanz des Begutachtungsergebnisses zu erhöhen. Andererseits könnten Schulungen und Weiterbildungen zu den Kriterien der Begutachtung für niedergelassene (Haus-)Ärzte das

Verständnis der Versicherungsmedizin erhöhen und Divergenz bei der Leistungsbeurteilung von Patienten verringern.

Die zusammenfassenden Ergebnisse der Publikationen werden in Abhängigkeit vom Studien-Outcome in Tabelle 15 berichtet. Die Tabelle beinhaltet Erstautor, Jahr der Publikation, Entstehungsland, Studientyp, Fragestellung (ggf. Methodik) und eine kurze Zusammenfassung der Ergebnisse.

Tabelle 15: Inhaltliche Charakteristika von Studien mit praktischen Hinweise zur Sozialmedizinischen Begutachtung.

Autor/Jahr/Land	Fragestellung/Hypothese	Methode						relevante Ergebnisse zur Begutachtungspraxis
		Meta-Analyse/ Sys. Review	RCT	Quasi-experimentelle Studie	Beobachtungsstudie	Strukturierter Expertenkonsens	Fallberichte, Expertenmeinungen	
Backman (2004) Kanada	Beschreibung von Risikofaktoren für Erwerbsminderung und Assessment-instrumenten zur Beurteilung von Arbeitseinschränkungen	--	--	--	--	--	●	Risikofaktoren sind körperlich anstrengende Arbeiten, mangelnde Autonomie bei der Arbeit, höhere Schmerzzustände, niedrigerer Funktionsstatus und niedrigerer Bildungsstand. Zur Beurteilung arbeitsbedingter Probleme wird der WIS und WLQ empfohlen.
Canela et al. (2015) Deutschland	Beschreibung der für den medizinischen Experten relevanten juristischen Aspekte bei der Begutachtung in der Schweiz	--	--	--	--	--	●	Bei anhaltenden somatoformen Schmerzstörungen sollen Komorbiditäten, chronische körperliche Begleiterkrankungen, sozialer Rückzug, primärer Krankheitsgewinn sowie unzureichende bisherige ambulante und stationäre Behandlungsergebnisse erhoben werden. Bei der Beurteilung wird das Mini-ICF zur Beschreibung von Funktionsstörungen bei psychischen Störungen empfohlen.
Canela et al. (2016) Deutschland	Identifikation von Assessment-instrumenten zur gutachterlichen Funktions- und Leistungsfähigkeitsbeurteilung Systematische Pubmed-Recherche von 13.771 Datenbanktreffern	●	--	--	--	--	--	Identifikation von 19 Selbstrating- und Fremdeinschätzungsinstrumenten (MELBA, O-AFP, WAI, AOF, WRFQ, DASI, IMA, MC, FAI, HSRS, GAS, SOFAS, GAF, OFS, WHODAS, GSDS, WHODAS 2.0, Mini-ICF App, SSA-BH), die jedoch nicht speziell für psychiatrische Fragestellungen konzipiert und validiert wurden.
Dodgshun et al. (2011) Australien	Beschreibung einer Anleitung für Hausärzte zur Beantragung eines finanziellen Arbeitsausgleiches für Patienten mit chronischen Erkrankungen	--	--	--	--	--	●	Beschreibung von relevanten körperlichen Untersuchungen in der Begutachtung, typischen Fallstricken, wie das Übersehen von psychosozialen Problemlagen und psychischen Störungen, die unzureichende Nutzung von Funktionsdiagnosen sowie die Betonung der Kommunikation zwischen Arzt und Gutachter als zentrales Element der Begutachtung.

3.3 Systematische Literaturrecherche: Ergebnisse

Autor/Jahr/Land	Fragestellung/Hypothese	Methode						relevante Ergebnisse zur Begutachtungspraxis
		Meta-Analyse/ Sys. Review	RCT	Quasi-experimentelle Studie	Beobachtungsstudie	Strukturierter Expertenkonsens	Fallberichte, Expertenmeinungen	
Dell-Kuster et al. (2014) Schweiz	Untersuchung der Übereinstimmung sozialmedizinischer Urteile zwischen Experten, behandelnden Ärzten und Antragstellern Analyse der Interrater-Reliabilität (prozentuale Übereinstimmung) anhand von n = 3.462 konsekutiven EM-Rentenfällen im Zeitraum von Januar 2005 bis Dezember 2008	--	--	--	●	--	--	In der Analyse wurde die Funktionsfähigkeit von Sozialmedizinischen Experten immer höher eingeschätzt als vom Antragsteller oder dessen behandelnden Arzt. Vor allem die Funktionsbeurteilung für den allgemeinen Arbeitsmarkt zeigte besonders große Differenzen. Die mittlere Differenz zwischen Experten und Antragsteller betrug 57%, zwischen Experte und behandelnder Arzt 46%.
Hammer (2002) Deutschland	Beschreibung des Ablaufs der Sozialmedizinischen Begutachtung von Patienten mit RA	--	--	--	--	--	●	Beschreibung von obligaten klinischen Untersuchungen (Gelenkbeweglichkeit, charakteristische Deformierungen, Schmerzen), hämatologischen und klinischen Laborparametern sowie rheumatologisch-immunologischen Parametern, technischen Untersuchungen (Röntgen) und Empfehlung von Messinstrumenten (FFbH, HAQ, DAS28, VAS, MSS, BDI, State-Trait-Angstinventar) zur Beurteilung des Leistungsvermögens.
Hoffmann e al. (2016) Deutschland	Beschreibung von relevanten Funktionsstörungen und Abschätzung des positiven und negativen Leistungsbildes von Patienten mit Prostatakarzinom	--	--	--	--	--	●	Beschreibung von Funktionsdefiziten, die während der Begutachtung berücksichtigt werden sollten (z. B. u.a. Erschöpfung, Harninkontinenz, sexuelle Funktionsstörungen Depression). Bereitstellung eines Modells zur Abschätzung der positiven und negativen Leistungsbildes in Bezug auf die Art der Behandlung, Arbeitsschwere, Arbeitshaltung, Arbeitsorganisation, psychomentele Funktionen, Sinnesfunktionen, bewegungsbezogenen Funktionen, kardiopulmonalen Funktionen sowie relevanten Gefährdungs- und Belastungsfaktoren.

Autor/Jahr/Land	Fragestellung/Hypothese	Methode						relevante Ergebnisse zur Begutachtungspraxis
		Meta-Analyse/ Sys. Review	RCT	Quasi-experimentelle Studie	Beobachtungsstudie	Strukturierter Expertenkonsens	Fallberichte, Expertenmeinungen	
Kellner (2002) Deutschland	Beschreibung des Ablaufs der Sozialmedizinischen Begutachtung von Patienten mit seronegativen Spondylarthropathien	--	--	--	--	--	●	Beschreibung von Diagnosekriterien, Besonderheiten der Anamnese (Krankheitsbeginn und -verlauf, Wirbel- und Gelenkschmerzen, extra-skeletale Manifestationen), diagnosespezifische klinischer Untersuchungen, Laboruntersuchung (Entzündungsparameter, Hämatologie, Rheumaserologische Parameter, Serologien arthritogener Parameter), bildgebender Verfahren (Röntgen, Arthrosonographie, Kernspintomographie) sowie Empfehlung von Messinstrumenten (HAQ, BASDAI, BASFI, BAS-G, BASMI, BASRI). Ferner werden abgestufte Orientierungskriterien für die Minderung der Erwerbsfähigkeit beschrieben.
Kolenda (1998) Deutschland	Beschreibung des Ablaufs der Sozialmedizinischen Begutachtung von Patienten mit koronarer Herzkrankheit	--	--	--	--	--	●	Nennung essentieller Bestandteile der Begutachtung, wie Anamnese, körperlicher Untersuchung, Ruhe-EKG, Echokardiographie, Belastungs-EKG, Langzeit-EKG, Lungenfunktionsuntersuchung sowie Röntgen-Thorax-Befund. Ferner wird in einem Schema der Zusammenhang zwischen den erhobenen medizinischen Befunden (aus Ergometrie, Echokardiographie, Langzeit-EKG), der Leistungsbeeinträchtigung, der zumutbaren beruflichen Tätigkeit, der möglichen Teilnahme an einer ambulanten Herzgruppe und dem voraussichtlich daraus resultierenden GdB beschrieben.
Rogers et al. (2006) USA	Untersuchung von kontext-abhängigen Faktoren, welche die Beschwerdedarstellung gegenüber dem Arzt beeinflussen und Identifikation von Assessmentverfahren zum Aufdecken von vorgetäuschten psychischen Beschwerden	--	--	--	--	--	●	Zur Detektion von vorgetäuschten Beschwerden bei Antragstellungen für finanzielle Kompensation, werden vier Erhebungsinstrumente zum Einsatz bei Verdacht auf symptomverzerrenden Darstellungen beschrieben (MMPI-2 PAI, SIRS, TSI). Es wird der Einsatz einer Kombination des MMPI-2 und eines strukturierten Interviews empfohlen.

Autor/Jahr/Land	Fragestellung/ Hypothese	Methode						relevante Ergebnisse zur Begutachtungspraxis
		Meta-Analyse/ Sys. Review	RCT	Quasi-experimentelle Studie	Beobachtungsstudie	Strukturierter Expertenkonsens	Fallberichte, Expertenmeinungen	
Schiltenswolf (2007) Deutschland	Beschreibung des Ablaufs der Begutachtung von Patienten mit muskuloskelettalen Schmerzen anhand einer S2-Leitlinie die Begutachtung von Schmerzen	--	--	--	--	--	●	Nennung essentieller Bestandteil der Begutachtung, wie (Schmerz-) Anamnese, klinische und apparative Diagnostik und bei Verdacht auf psychischer Schmerzgenese ein strukturiertes klinisches Interview (SKID, CIDI, Mini-DIPS). Für die abschließende Leistungsbeurteilen sollte die Grunderkrankung, Schweregrad, Beeinträchtigung der Aktivität und Partizipation, Zumutbarkeit von weiteren Interventionen, Komorbiditäten, psychosozialen Auswirkung, Grad der Chronifizierung, Vorbehandlung sowie Angemessenheit einer psychotherapeutischen Behandlung berücksichtigt werden.

3.3.1.2 Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit

Die Publikationen zum Themenkomplex ICF befassen sich mit dem Einsatz von ICF Core Sets zur Erstellung Sozialmedizinischer Gutachten (Kirschneck et al., 2008, 2015) sowie dem Assessment von Funktionseinschränkungen unter der Perspektive der ICF (Linden, 2017). Unter den drei identifizierten Publikationen in dieser Kategorie befanden sich zwei retrospektiv-qualitative Beobachtungsstudien (Kirschneck et al., 2008, 2015) und ein narratives Review (Linden, 2017).

Beide retrospektiv-qualitativen Studien von Kirschneck und Kollegen (2008, 2015) beschrieben Daten desselben Projektes und wurden im Kapitel 1.4.2 dieser Arbeit bereits kurz zusammengefasst. Die Arbeiten beruhen auf der Linking-Methode, bei der Sozialmedizinische Gutachten durch unabhängige Gesundheitsfachpersonen in die Sprache der ICF übersetzt wurden. In ihrer ersten Studie aus dem Jahr 2008 untersuchten Kirschneck und Kollegen, ob die ICF Core Sets für lumbale Rückenschmerzen und das generalisierte Schmerzsyndrom als Grundlage zur Erstellung von Sozialmedizinischen Gutachten dienen können. Hierzu sollten in Zusammenarbeit mit der DRV Bund und vier weiteren Regionalträgern der RV 300 anonyme Sozialmedizinische Gutachten mit den folgenden Erstdiagnosen analysiert werden: M42 „Osteochondrose der Wirbelsäule“, M45 „Spondylitis ankylosans“, M46 „Sonstige entzündliche Spondylopathien“, M47 „Spondylose/Arthrose der Wirbelsäule“, M48 „Sonstige Spondylopathien“, M51 „Sonstige Bandscheibenschäden“, M53 „Instabilitäten der Wirbelsäule“, M54 „Rückenschmerzen“, M99 „Biomechanische Funktionsstörungen, andernorts nicht klassifiziert“ und F45.4 „anhaltende somatoforme Schmerzstörung“. In einem mehrschrittigen Verfahren wurden die Gutachten zunächst gelesen und im Anschluss in inhaltliche Konzepte gekennzeichnet, die in die Sprache der ICF übersetzt werden sollen. Zwei Gesundheitsexperten ordnen nachfolgend die präzise ICF Kategorie dem inhaltlichen Konzept zu. Die Übereinstimmung zwischen den Gesundheitsexperten galt als Indikator dafür, welche ICF Kategorie gelinkt wird. Die gelinkten ICF Kategorien wurden im Anschluss mit den ICF Core Sets verglichen. Erste Ergebnisse zeigten, dass die Struktur der Sozialmedizinischen Gutachten stark heterogen war, infolgedessen eine direkte und indirekte Vergleichbarkeit der ICF Codes nur schwer möglich sei. Kirschneck und Kollegen (2008) schlussfolgerten jedoch, dass die ICF als Strukturierungshilfe bei der Erstellung von Gutachten deren Objektivität maßgeblich erhöhen könnte.

In ihrer Folgepublikation untersuchten Kirschneck und Kollegen (2015) anhand der beschriebenen Methodik, „ob die krankheitsspezifischen ICF Core Sets und andere internationale Ansätze, wie die kurzen ICF Core Sets für die berufliche Rehabilitation

(ICF-bR) und das ICF Core Set zur Evaluierung von Behinderung im Dienste der sozialen Sicherheit (ICF-EUMASS), die Inhalte der sozialmedizinischen Gutachten abbilden“ (Kirschneck et al., 2015, S. 92). Aus insgesamt 600 zur Verfügung gestellten anonymisierten Gutachten wurden 156 Gutachten zur Beurteilung chronisch lumbaler Rückenschmerzen und 138 Gutachten zur Beurteilung des generalisierten Schmerzsyndroms analysiert. Die Ergebnisse zeigen, dass der Gutachtenabschnitt „Sozialmedizinische Leistungsbeurteilung“ in allen Gutachten mittels der ICF nicht kodiert werden konnte. Ebenso wurden Interventionen, Informationen über Familienmitglieder, GdB und suizidale Tendenzen nicht in die Sprache der ICF überführt. Insgesamt beinhalten die Gutachten etwa 40 % aller 1.424 zur Verfügung stehenden ICF Codes. Die Varianz wurde minimiert, indem nur die Codes weiterverarbeitet wurden, die in über 50 % der gelinkten Gutachten kodiert worden waren (chronischen Rückenschmerzen: 77 ICF-Kategorien/generalisiertes Schmerzsyndrom: 80 ICF-Kategorien). Die Autoren schlussfolgerten, dass dieser Anteil der Codes weitgehend alle Aspekte der funktionalen Gesundheit der krankheitsspezifischen ICF Core Sets, ICF-bR und des ICF-EUMASS abbilden. Gleichzeitig wurde jedoch betont, dass eine Mehrheit der gelinkten ICF Kategorien, die wiederum die Individualität der Begutachtung widerspiegeln, nicht in den Core Sets enthalten sind. Entsprechend ist zur Feststellung des qualitativen und quantitativen Leistungsvermögens eine fallspezifische individuelle Kodierung notwendig, die über die Inhalte der untersuchten ICF Core Sets hinaus geht. Um die inhaltliche Varianz zu verringern und den Aufbau der Sozialmedizinischen Gutachten zu standardisieren wurde zudem empfohlen, den Einsatz von an ICF angelehnten Checklisten voranzutreiben.

In seinem narrativen Review gab Linden (2017) einen Überblick über die Definition und möglichen Assessmentinstrumente von (Funktions-)Beeinträchtigungen (engl.: „disability“) vor dem Hintergrund des ICF-Modells bei Patienten mit psychischen Störungen. Behinderung und Beeinträchtigungen sind kontextbezogen und werden definiert als Begrenzung der benötigten Kapazitäten, um am täglichen Leben teilzunehmen. Im Sinne der ICF ist Behinderung ein mehrdimensionales Konzept, das Beeinträchtigung von Körperstrukturen oder -funktionen, Kapazitätsbeschränkungen, Umwelt sowie Teilhabe und Partizipation umfasst. Die Publikation beschreibt unterschiedliche Testverfahren, um dieses Kapazitäts-Kontext-Interaktionsmodell in der Begutachtung adäquat zu erfassen. Zur Erfassung des Ausmaßes von Beeinträchtigung und Aktivitäten benannte Linden (2017) elf Testinstrumente und deren Zieldimensionen:

- (Instrumental) Activities of Daily Living (IADL): generelle Pflegebedürftigkeit,
- Barthel Index: Bewertung alltäglichen Fähigkeiten,
- Global Assessment Scale (GAS): globales Assessment der Funktionsfähigkeit,
- Global Assessment of Functioning (GAF): globales Assessment der Funktionsfähigkeit,
- Personal and Social Performance Scale (PSP): psychosoziales Funktionsniveau,
- Social and Occupational Functioning Assessment Scale (SOFAS): generelle psychologische Probleme,
- Groningen Social Disability Schedule (GSDS-II): halb-strukturiertes Interview zur Erfassung der Teilhabe und Partizipation,
- Index zur Messung von Einschränkungen der Teilhabe (IMET): ICF basierte Beurteilung der Einschränkungen in Alltag, familiären und häuslichen Verpflichtungen, Aktivitäten, Freizeit, sozialen Aktivitäten, persönlichen Beziehungen, Sexualleben und Umgang mit Stressoren,
- ICF-AT-50-Psych: Aktivitäten- und Teilhabebeschränkungen
- World Health Organization Disability Assessment Schedule II (WHODAS 2.0): Aktivitäten- und Teilhabebeschränkungen
- Mini-ICF-APP: Aktivitäten- und Teilhabebeschränkungen

Die zusammenfassenden Ergebnisse der Studien mit dem Schwerpunkt ICF werden in Abhängigkeit vom Studien-Outcome in Tabelle 16 berichtet. Die Tabelle beinhaltet Erstautor, Jahr der Publikation, Entstehungsland, Studientyp, Fragestellung (ggf. Methodik) und eine kurze Zusammenfassung der Ergebnisse.

Tabelle 16: Inhaltliche Charakteristika von Studien mit Bezug zur Anwendung der ICF in der Sozialmedizinischen Begutachtung.

Autor/Jahr/Land	Fragestellung/Hypothese	Methode						relevante Ergebnisse zur Begutachtungspraxis
		Meta-Analyse/ Sys. Review	RCT	Quasi-experimentelle Studie	Beobachtungsstudie	Strukturierter Expertenkonsens	Fallberichte, Expertenmeinungen	
Kirschneck et al. (2008) Deutschland	Übersetzung der Inhalte sozialmedizinischer Gutachten in die Sprache der ICF zur Prüfung, ob krankheitsspezifische ICF Core Sets, ICF-bR und ICF-EUMASS die Inhalte sozialmedizinischer Gutachten abbilden	--	--	--	●	--	--	Bericht des Studienprotokolls und vorläufiger Analysen beim Stand von n = 10 bearbeiteten Gutachten. Erste Ergebnisse zeigten eine hohe Variabilität der Struktur der Sozialmedizinischen Gutachten. Der Einsatz der ICF als Strukturierungshilfe bei der Erstellung von Gutachten wird empfohlen.
Kirschneck et al. (2015) Deutschland	Qualitative Analyse n = 294 sozialmedizinische Gutachten aus n = 5 Rentenversicherungsträgern; Linkage-Methode	--	--	--	●	--	--	Analyse von 156 Gutachten bei chronisch lumbalen Rückenschmerzen und 138 Gutachten bei generalisierten Schmerzsyndroms. Nicht alle Gutachtenbestandteile konnten in ICF Codes umgewandelt und kodiert werden (z. B. Aspekte der Sozialmedizinischen Leistungsbeurteilung). 77 ICF-Kategorien wurden in mehr als 50 % der Gutachten bei chronisch lumbalen Rückenschmerzen, 80 ICF-Kategorien beim generalisierten Schmerzsyndrom gefunden. Weitgehend alle Aspekte der funktionalen Gesundheit der krankheitsspezifischen ICF Core Sets, ICF-bR und des ICF-EUMASS werden damit abgebildet. Die Mehrheit der gelinkten ICF Kategorien aus den Sozialmedizinischen Gutachten sind jedoch nicht in den Core Sets enthalten. Die Beurteilung des Leistungsvermögens geht über die Inhalte der untersuchten ICF Core Sets hinaus.
Linden (2017) Deutschland	Darstellung von Definition und Assessmentinstrumenten zur Messung von (Funktions-) Beeinträchtigungen vor dem Hintergrund der ICF bei Patienten mit psychischen Störungen	--	--	--	--	--	●	Behinderung und Funktionsbeeinträchtigung wird als mehrdimensionales Konzept definiert, das Beeinträchtigung von Körperstrukturen oder -funktionen, Kapazitätsbeschränkungen, Umwelt und Teilhabe und Partizipation umfasst. Es werden Empfehlungen für Messinstrumente gegeben (ADL, Barthel Index, GAS, GAF, PSP, SOFAS, GSDS-II, IMET, ICF-AT-50-Psych, WHODAS 2.0, Mini-ICF-APP).

3.3.1.3 Begutachtung aus Arztperspektive

Zwei Studien konnten zum Themenkomplex Arztperspektive identifiziert werden. Diese untersuchten anhand qualitativen Fokusgruppen- und Einzelinterviews die Sicht der Sozialmedizinischen Experten zum Ablauf der Begutachtung. Während sich eine Studie mit den Erfahrungen und Schwierigkeiten bei der Begutachtung von Menschen mit psychischen Störungen befasste (Dewa, Hees, Trojanowski & Schene, 2015), beschrieb die andere Publikation die Begutachtungsschwerpunkte sowie Erfahrungen mit Leitlinien bei der Beurteilung der Erwerbsminderung von Patienten mit Tumorerkrankungen (van Muijen, Duijts, Kornet-van der Aa, van der Beek & Anema, 2015).

Dewa und Kollegen (2015) untersuchten die Erfahrungen von Klinikern im Zusammenhang mit der Durchführung medizinischer Beurteilungen im Rahmen der Begutachtung von Menschen mit psychischen Störungen. An der Studie nahmen insgesamt $n = 13$ interdisziplinäre Gesundheitsexperten teil. Die Teilnehmer kamen aus den Fachdisziplinen Sozialarbeit ($n = 1$), Ergotherapie ($n = 1$), Psychologie ($n = 2$), Allgemeinmedizin ($n = 2$) und Psychiatrie ($n = 7$). Es wurden zwei Fokusgruppen (bestehend aus $n = 2$ bzw. $n = 3$ Teilnehmern) und acht Einzelinterviews durchgeführt. Aus den Interviews ging hervor, dass die Validierung der Diagnose und der Beschwerden durch die Abhängigkeit von der Selbstausskunft des Versicherten und dem Mangel an objektiven Daten erschwert sei. Häufig könne eine genaue Diagnose nicht bestimmt werden. Darüber hinaus wurde angemerkt, dass psychiatrische Diagnosen einer zeitlichen Veränderung unterliegen und die als ursprüngliche Ursache der Funktionseinschränkungen identifizierte Diagnose möglicherweise nicht die Diagnose ist, die die anhaltende Erwerbsminderung erklärt. Die Gewährung der Erwerbsminderung sollte daher nicht allein auf Grundlage einer Primärdiagnose erfolgen. Die Interviewpartner berichteten, dass die Diagnose eher Anhaltspunkt für die daraus folgenden Beeinträchtigungen sei. Bei Begutachtung der Funktionsfähigkeit können Menschen mit gleichen Symptomen in ihrer Funktionsfähigkeit unterschiedlich beeinträchtigt sein, dies müsse in der Begutachtung herausgearbeitet werden. Darüber hinaus seien psychosoziale und Umweltfaktoren sowie die Charakteristik des bisherigen Arbeitsplatzes zu berücksichtigen. Von Bedeutung sind zudem kognitive Funktionen wie Verarbeitungsgeschwindigkeit, Exekutivfunktion und Arbeitsgedächtnis.

In der Studie von van Muijen und Kollegen (2015) wurden Aspekte der Sozialmedizinischen Begutachtung aus Sicht von Versicherungsmedizinern bei der Begutachtung von Patienten mit Tumorerkrankungen untersucht. An der Studie nahmen insgesamt $n = 29$ Versicherungsmediziner teil. Vier Teilnehmer befanden sich noch in der Ausbildung. Die Range der Beschäftigungsdauer in der Sozialversicherung reichte von 2 bis 39 Jahren. Es wurden drei konsekutive Themenlisten-geleitete Fokusgruppen durchgeführt. Die

vorstrukturierte Themenliste umfasste die generellen Erfahrungen von Versicherungsmedizinern, Krebs assoziiertes Fatigue-Syndrom, die Functional Abilities List (FAL), leitlinientreue Begutachtung sowie Prognoseregeln. Bei der Beurteilung von Tumorkranken fühlen sich die begutachtenden Ärzte besonders um eine korrekte Beurteilung der Arbeitsunfähigkeit verpflichtet. Entsprechend wurde in den Interviews Faktoren benannt, die bei der Begutachtung der Return-to-Work Fähigkeit von Tumorkranken erhoben werden sollten. Dazu zählen u.a. personbezogene Faktoren, wie Soziodemographie (Alter, sozioökonomischer Status, Einkommen, Familienstand), physikalische Domänen (physischer Allgemeinzustand, Funktionseinschränkungen, Einschränkungen im alltäglichen und sozialen Leben, allgemeines Wohlbefinden) sowie psychologische Domänen (Erwartungen, Kognition, Motivation, Coping, Progredienzangst, Stress, Bedeutung der Arbeit). Ferner sollten krankheitsbezogene Faktoren (Tumorstadium, Prognose, Nebenwirkungen, Komorbidität, Art der Behandlung), arbeitsbezogene Faktoren (Erwerbsstatus, Arbeitszufriedenheit, Arbeitsschwere und -belastung, Schichtarbeit, Arbeitsstress) und umweltbezogene Faktoren (Unterstützung durch Partner, Kollegen, Arbeitgeber, finanzielle Unterstützungen, Arbeitsmarktsituation) untersucht werden. Ähnliche Faktoren werden bei Untersuchung des Fatigue-Syndroms herangezogen. Zur Messung der Funktionsfähigkeit wurde die FAL als angemessenes Assessmentinstrument betrachtet, das jedoch in Bezug auf Aspekte der kognitiven Funktionen Defizite aufweist. Seitens der Nutzer wurde die Einhaltung von Leit- oder Richtlinien sowohl negativ als auch positiv beurteilt. Einerseits werden Leitlinien als hilfreiche Informationsquelle betrachtet, die Ausgangspunkt für die Beantwortung gutachtenspezifischer Fragestellungen sein können und die sozialmedizinische Entscheidung im Wesentlichen unterstützen. Andererseits sei die Verwendung von Leitlinien zu zeitaufwendig, diese seien häufig veraltet und zu unspezifisch. Zudem würden Leitlinien auch bei deren Anwendung häufig nicht die eigentliche Beurteilung der funktionellen Fähigkeiten inhaltlich ausreichend untermauern.

Die zusammenfassenden Ergebnisse der zwei Studien werden in Abhängigkeit vom Studien-Outcome in Tabelle 17 berichtet. Die Tabelle beinhaltet Erstautor, Jahr der Publikation, Entstehungsland, Studientyp, Fragestellung (ggf. Methodik) und eine kurze Zusammenfassung der Ergebnisse.

Tabelle 17: Inhaltliche Charakteristika von Studien zur Sichtweise der Versicherungsmedizin auf die Sozialmedizinische Begutachtung.

Autor/Jahr/Land	Fragestellung/ Hypothese	Methode						relevante Ergebnisse zur Begutachtungspraxis
		Meta-Analyse/ Sys. Review	RCT	Quasi-experimentelle Studie	Beobachtungsstudie	Strukturierter Expertenkonsens	Fallberichte, Expertenmeinungen	
Dewa et al. (2015) Kanada	Untersuchung der Erfahrungen von Ärzten und anderen Klinikern bei der Beurteilung von Funktionseinschränkungen im Zusammenhang mit psychischen Störungen Qualitative Fokusgruppen- und Einzelinterviews; n = 13 Teilnehmer unterschiedlicher Fachdisziplinen	--	--	--	●	--	--	Die Beurteilung der Erwerbsunfähigkeit bei psychischen Störungen wird durch das Fehlen objektiver Daten erschwert wahrgenommen. Weitere genannte Erfahrungsberichte belegen, dass sich im zeitlichen Verlauf der Erkrankung die Diagnosen und Funktionseinschränkungen verändern können, was eine valide Prognose behindern kann. Darüber hinaus sollten psychosoziale und Umweltfaktoren, die Charakteristik des bisherigen Arbeitsplatzes und kognitive Funktionen erhoben werden.
van Muijen et al. (2015) Niederlande	Beschreibung von Aspekten, die Ärzte bei der Beurteilung von Erwerbsunfähigkeit bei Tumorpatienten berücksichtigen und ihre Erfahrungen mit Leitlinien Qualitative Fokusgruppeninterviews; n = 29 Ärzte	--	--	--	●	--	--	Bei der Beurteilung der Erwerbsfähigkeit sollten medizinische und nicht-medizinische Faktoren berücksichtigt werden. Krankheitsbezogene Faktoren sind Tumorstadium, Prognose, Nebenwirkungen, Komorbidität und Art der Behandlung. Nicht-medizinische Faktoren in Bezug auf die Person, Arbeit und soziales Umfeld, sollten in Fällen, in denen medizinische Probleme nicht primär im Vordergrund stehen, berücksichtigt werden. Zur Messung der Funktionsfähigkeit wird die FAL empfohlen. Die Nutzung von Leitlinien wird sowohl positiv als auch negativ beurteilt.

3.3.2 Qualität der eingeschlossenen Studien

Die Qualitätsbewertung wurde mit dem MMAT beurteilt und ist überblicksartig in Tabelle 19 zusammengestellt (vgl. Hong et al., 2018). Das Beurteilungsschema zeigt, dass 15 der 16 Publikationen zur Sozialmedizinischen Begutachtung einen „qualitativen Forschungsansatz“ aufweisen. Als qualitativ werden im MMAT alle Studiendesigns subsummiert, die keine quantitativen Daten berichten. Dies beinhaltet sowohl narrative Reviews, systematische Literaturrecherchen sowie Beobachtungsstudien mit qualitativen Analysemethoden. Elf der 16 untersuchten Studien erfüllten weniger als 50 % der vordefinierten Qualitätskriterien. Lediglich $\frac{1}{4}$ der eingeschlossenen Publikationen erfüllte alle Qualitätskriterien des MMAT. Aggregiert über alle Studien beträgt die Qualität, definiert als Grad der MMAT-Kriterienerfüllung, 56,25 %, was als mangelhaft (<60 %) beurteilt werden kann.

Vor allem narrative Reviews erfüllten durchschnittlich weniger Qualitätskriterien als Studiendesigns höherer Evidenzklassen. So konnte in keinem der Reviews valide dargestellt werden, ob die unsystematisch erhobene Literatur geeignet ist, um die Forschungsfrage adäquat zu beantworten (Kriterium 1.2), ob die Interpretation der Ergebnisse durch die unsystematisch erhobenen Daten ausreichend belegt ist (Kriterium 1.4) und ob eine schlüssige Kohärenz zwischen den unsystematisch erhobenen Datenquellen, der Erhebung, der Analyse und der Interpretation vorliegt (Kriterium 1.5).

3.3.3 Evidenz der eingeschlossenen Studien

Die Evidenz der Publikationen erreicht unter Verwendung des Evidenzklassenmodells von Kliche und Kollegen (2006) hauptsächlich niedrigere Evidenzen (Tab. 18). Die Einteilung in die fünf Evidenzklassen orientiert sich primär an den zugrunde gelegten Studiendesigns der Publikationen. Durch den Einschluss von primär narrativen Reviews und Beobachtungsstudien ohne Kontrollgruppe mit nur einem Messzeitpunkt wurden insgesamt zehn Studien die Evidenzklasse V und fünf Studien die Evidenzklasse IV zugeordnet. Lediglich das identifizierte systematische Review von Canela und Kollegen (2016) erreichte die Evidenzklasse I.

Tabelle 18: Evidenzklassen der eingeschlossenen Studien.

Studie von	Methodik	Evidenz- klasse
Backman (2004)	Narratives Review	V
Canela et al. (2016)	Systematische Literaturrecherche	I
Canela et al. (2015)	Narratives Review	V
Dell-Kuster et al. (2015)	Analyse Interrater-Reliabilität	IV
Dewa et al. (2015)	Qualitative Fokusgruppen	IV
Dodsgun et al. (2011)	Narratives Review	V
Hammer (2002)	Narratives Review	V
Hoffmann et al. (2016)	Narratives Review	V
Kellner (2002)	Narratives Review	V
Kirschneck et al. (2008)	Linkage-Methode, qualitative Analyse	IV
Kirschneck et al. (2015)	Linkage-Methode, qualitative Analyse	IV
Kolenda (1998)	Narratives Review	V
Linden (2017)	Narratives Review	V
Rogers et al. (2006)	Narratives Review	V
Schiltewolf (2007)	Narratives Review	V
van Muijen et al. (2015)	Qualitative Fokusgruppen	IV

Tabelle 19: Qualitätsbeurteilung der eingeschlossenen Studien anhand des Mixed Methods Appraisal Tools nach Hong et al. (2018).

Methodological quality criteria	Studienbeurteilung															
	1	2*	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Screening questions (for all types)																
S1. Are there clear research questions?	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
S2. Do the collected data allow to address the research questions?	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Qualitative (incl. narrative review research)																
1.1. Is the qualitative approach appropriate to answer the research question?	●	●	●	--	●	●	●	●	●	●	●	●	●	●	●	●
1.2. Are the qualitative data collection methods adequate to address the research question?	○	●	○	--	●	○	○	○	○	●	●	○	○	○	○	●
1.3. Are the findings adequately derived from the data?	●	●	●	--	●	○	●	●	●	○	●	●	●	●	●	●
1.4. Is the interpretation of results sufficiently substantiated by data?	○	●	○	--	●	○	○	○	○	○	●	○	○	○	○	●
1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?	○	●	○	--	●	○	○	○	○	○	●	○	○	○	○	●
Quantitative randomized controlled trials																
2.1. Is randomization appropriately performed?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2.2. Are the groups comparable at baseline?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2.3. Are there complete outcome data?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2.4. Are outcome assessors blinded to the intervention provided?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2.5. Did the participants adhere to the assigned intervention?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

* Das MMAT sieht die Beurteilung von systematischen Literaturarbeiten laut Manual per Definition nicht vor. In der vorliegenden Arbeit wurde die Qualität dennoch anhand der definierten Qualitätskriterien beurteilt.

Methodological quality criteria	Studienbeurteilung															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>Quantitative non-randomized</i>																
3.1. Are the participants representative of the target population?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
3.3. Are there complete outcome data?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
3.4. Are the confounders accounted for in the design and analysis?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<i>Quantitative descriptive</i>																
4.1. Is the sampling strategy relevant to address the research question?	--	--	--	●	--	--	--	--	--	--	--	--	--	--	--	--
4.2. Is the sample representative of the target population?	--	--	--	●	--	--	--	--	--	--	--	--	--	--	--	--
4.3. Are the measurements appropriate?	--	--	--	○	--	--	--	--	--	--	--	--	--	--	--	--
4.4. Is the risk of nonresponse bias low?	--	--	--	●	--	--	--	--	--	--	--	--	--	--	--	--
4.5. Is the statistical analysis appropriate to answer the research question?	--	--	--	●	--	--	--	--	--	--	--	--	--	--	--	--

Methodological quality criteria	Studienbeurteilung															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mixed methods																
5.1. Is there an adequate rationale for using a mixed methods design to address the research question?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
5.2. Are the different components of the study effectively integrated to answer the research question?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Quality Ranking																

(1) Backman, 2004; (2) Canela et al., 2016; (3) Canela et al., 2015; (4) Dell-Kuster et al., 2014; (5) Dewa et al., 2015; (6) Dodsgun et al., 2011; (7) Hammer, 2002; (8) Hoffmann et al., 2016; (9) Kellner, 2002; (10) Kirschneck et al., 2008; (11) Kirschneck et al., 2015; (12) Kolenda, 1998; (13) Linden, 2017; (14) Rogers et al., 2006; (15) Schiltenswolf, 2007; (16) van Muijen et al., 2015

● = Kriterium erfüllt, ○ = Kriterium nicht erfüllt, -- = Kriterium nicht zutreffend

Qualitätsbeurteilung:

- 20% der MMAT-Kriterien erfüllt
- 40% der MMAT-Kriterien erfüllt
- 60% der MMAT-Kriterien erfüllt
- 80% der MMAT-Kriterien erfüllt
- 100% der MMAT-Kriterien erfüllt

3.4 Diskussion

3.4.1 Allgemeine Diskussion

Die Teilstudie stellt den zweiten Versuch dar, eine systematische Literaturübersicht zur Evidenzbasierung der Sozialmedizinischen Begutachtung bei Anträgen auf EM-Renten zu erstellen. Das vorliegende systematische Review hatte zum Ziel, evidenzbasierte Publikationen im Kontext der Sozialmedizinischen Begutachtung zu identifizieren. Die Recherche macht mit etwa 12.000 identifizierten Publikationen den Umfang potenziell relevanter Publikationen zur Thematik und damit die Bedeutung der Fragestellung deutlich. Jedoch wurde die Suchstrategie auch sehr sensitiv konzipiert, um ein möglichst großes Spektrum an Publikationen einschließen zu können. Die Recherche wurde auf die häufigsten medizinischen Indikationen einer EM-Rente (vgl. GBE-Bund, 2018b) beschränkt, nachdem in der Vorläuferstudie keine Einschränkungen bei der Recherche gemacht worden sind.

Obgleich in systematischen Übersichtsarbeiten üblicherweise zur besseren Vergleichbarkeit der Studienergebnisse ausschließlich RCTs einbezogen werden, wurde in der vorliegenden Arbeit keine Beschränkung des Artikeltyps vorgenommen. Untersucht wurden alle Artikel seit dem Jahr 1998, die zur Thematik „Begutachtung und Erwerbsminderung“ publiziert wurden. Nach Analyse der Ergebnisse erfüllten 16 Publikationen die Ein- und Ausschlusskriterien.

Zentrale Befunde und Anwendungsfelder. Innerhalb der Recherche wurden mehrere zentrale Forschungsthemen identifiziert, welche mit der Sozialmedizinischen Begutachtung verknüpft sind. Durch die Ausarbeitung der ICF, der Verabschiedung des SGB IX in Deutschland im Jahr 2001 und auch durch das Inkrafttreten der UN-Behindertenrechtskonvention im Jahr 2009 hat sich das Anforderungsprofil für Gutachten geändert. Vor allem die mehrdimensionale ICF wird zunehmend in der Begutachtung als relevantes Strukturierungskonzept angesehen. Diesem Umstand entsprechend finden sich im Review mehrere Publikationen wieder, die sich damit auseinandersetzen, wie die ICF in die Sozialmedizinische Begutachtung eingebunden werden kann (Kirschneck et al., 2008, 2015; Linden, 2017). Die Entwicklung eines ICF Core-Sets für die Sozialmedizinische Beurteilung (Brage et al., 2008) und ICF basierte Assessmentinstrumente zur Bestimmung relevanter Teilhabe- und Leistungseinschränkungen bieten einen ersten Ansatzpunkt, das Konzept in die Begutachtung einfließen zu lassen. Gleichzeitig wurde deutlich, dass vor allem die Anwendung der ICF im Bereich der individuellen Begutachtung zukünftig verstärkt in den Fokus der Wissenschaft rücken sollte. Neben der indizierten Anwendung als

„Forschungs- und Diagnostikinstrument“ zur Messung von Aktivitäten und Teilhabe sowie personbezogenen und Umweltfaktoren (vgl. DIMDI/WHO, 2005, S. 11), war es bisher nicht möglich, die ICF als „Instrument der Versorgung“ zur Beurteilung der Leistungsfähigkeit zu nutzen. So stellten Kirschneck und Kollegen (2008, 2015) fest, dass die eigentliche Leistungsbeurteilung mittels ICF nicht kodiert werden kann. Andere Gutachtenabschnitte profitieren hingegen von der ICF und den aus ihr abgeleiteten ICF-Core Sets, als möglichen Leitfaden für die Begutachtung. Die Core Sets können helfen, relevante Aspekte einer Erkrankung strukturiert mittels Anamnese und/oder körperlicher Untersuchung zu erheben. Eine Ableitung der Leistungsfähigkeit ergibt sich daraus jedoch nicht.

An dieser Stelle der Recherche lässt sich auch ein Hinweis auf die Komplexität und Individualität des sozialmedizinischen Begutachtungsgeschehens finden. Kirschneck und Kollegen (2015) zeigten, dass bei der Kodierung von Gutachten mehr als 40 % aller verfügbaren ICF-Codes vergeben werden konnten. Diese Variabilität unterstreicht, dass kein Gutachtenfall gleich ist und je nach Fallkonstellation und medizinischer Grunderkrankung verschiedene Aspekte bei der Beurteilung der Leistungsfähigkeit herangezogen werden müssen. Entsprechend ist anzunehmen, dass der Prozess der Gutachtenerhebung als auch das schriftliche Gutachten selbst zwischen verschiedenen Gutachtern sehr heterogen ausfällt, was neben interindividuellen Unterschieden wiederum Ausdruck der Komplexität der Begutachtung sein kann. Ergänzend berichten Canela und Kollegen (2015), dass die Begutachtung abhängig von der Länge der Untersuchung, vom Gutachter (Persönlichkeit/Fragetechnik) sowie vom Rentenantragsteller (Persönlichkeit) ist. Diese Faktoren führen wahrscheinlich dazu, dass die Interrater-Reliabilität von Gutachtern bei der Beurteilung desselben Versicherten eher niedrig ausfallen wird.

Weitere Publikationen befassen sich mit allgemeinen Hinweisen bzw. indikationsspezifischen Empfehlungen für die Begutachtungspraxis. So wird primär in narrativen Reviews dargestellt, wie eine vollständige und transparente Begutachtung durchzuführen und zu protokollieren ist. Diese Publikationen orientieren sich indikationsspezifisch an das Schema zum Aufbau sozialmedizinischer Gutachten der RV (VDR, 2001) und beschreiben relevante zu erhebende Bestandteil der Anamnese und der Befunde. Auch in dieser systematischen Literaturrecherche erscheint die Datenlage von Artikeln zur Begutachtung, die in wissenschaftlichen Zeitschriften publiziert wurden, limitiert. Die identifizierten narrativen Reviews erreichen lediglich die Evidenzklasse V und enthalten allenfalls Hinweise zur Begutachtung für einige ausgewählte medizinische Indikationen. Diese sind zudem aufgrund des Evidenzniveaus nicht verallgemeinerbar. Zusätzlich basieren die narrativen Reviews auf einer unsystematischen Datensammlung, was sich methodisch negativ auf die gesamte Studienqualität, gemessen mit dem MMAT, auswirkt.

Darüber hinaus wird in der sozialmedizinischen Forschung diskutiert, welche Assessmentinstrumente objektive und valide Ergebnisse zur Vorhersage der Arbeits- und Leistungsfähigkeit eines Patienten haben. Zu diesem Thema fand sich eine einzige systematische Literaturrecherche. Canela und Kollegen (2016) untersuchen die psychometrischen Eigenschaften von Assessments für die Beurteilung der Funktionsfähigkeit bei Patienten mit psychiatrischen Störungen. Die Autoren können aufgrund der fehlenden Validierung für keines der untersuchten Assessments eine Empfehlung zum Einsatz in der Sozialmedizinischen Begutachtung aussprechen.

Limitationen der Studie. Die Ergebnisse des systematischen Reviews können durch mehrere Fehlerquellen (Bias) beeinflusst werden. Das Review kann dem sog. „search bias“ unterliegen, bei dem keine adäquaten Suchtermini oder Suchstrategien benutzt wurden, um alle relevanten Studien eines Themas zu finden. Die in der vorliegenden Studie gewählte sensitive Suchstrategie führte zu einer hohen Zahl an falsch-positiven Ergebnissen. Zusätzlich konnte die Recherche nicht nach dem PICO-Schema umgesetzt werden, da aufgrund des zu untersuchenden Themas weder nach einem Wirksamkeitsvergleich von Interventionen noch nach speziellen klinischen Endpunkten gesucht wurde. Der verwendete Suchalgorithmus sollte Literatur mit Bezug zur Begutachtung bei EM-Rente identifizieren. Es ist nicht auszuschließen, dass bestimmte indikationsbezogene Literaturstellen unentdeckt blieben bzw. der Suchalgorithmus zur Bearbeitung der Fragestellung inadäquat gewählt war. Beispielsweise gibt es voraussichtlich mehr Literatur zur Validierung von Assessments, bei denen die Funktionsfähigkeit gemessen werden. Sofern dort in den Schlagworten allerdings kein ausdrücklicher Bezug zur Begutachtung (im Sinne der beschriebenen Schlagworte) genommen wurde, konnte diese Literatur nicht berücksichtigt werden. Trotz einer im Prinzip sensitiven Suchstrategie war eine Beschränkung der Suche notwendig, um die anfallenden Suchergebnisse angemessen verwerten zu können.

Als weitere mögliche Fehlerquelle muss genannt werden, dass ausschließlich deutsch- und englischsprachige Publikationen in die Recherche eingeschlossen wurden und es zum Ausschluss relevanten Studien in anderen Sprachen gekommen sein könnte.

Darüber hinaus wurden durch die sensitive Suchstrategie eine große Anzahl an Publikationen identifiziert, die vom Verfasser dieser Arbeit anhand der Ein- und Ausschlusskriterien überprüft wurden. Da die Ein- und Ausschlusskriterien für sehr heterogene Publikationsarten im mixed-methods Reviews angewendet wurden, ist ein Selektionsbias durch den durchführenden Wissenschaftler nicht auszuschließen. Außerdem nahm an der vorliegenden Teilstudie kein zweiter Reviewer teil, der validierend

die Liste der identifizierten Publikationen anhand der Ein- und Ausschlusskriterien gescreent hat. Ein Selektionsbias hätte dadurch minimiert werden können.

Da die Beurteilung der Studienqualität mit einem Studiendesign-übergreifenden Tool (MMAT) durchgeführt wurde, kann nicht ausgeschlossen werden, dass die Qualität einiger eingeschlossener Studien über- bzw. unterschätzt wurde. Der parallele Einsatz mehrerer, auf das entsprechende Studiendesign zugeschnittener, Beurteilungsbögen hätte ein differenzierteres Bild der Qualität der eingeschlossenen Studien geben können. Gleichzeitig sei an dieser Stelle jedoch angemerkt, dass der MMAT ausgewählt wurde, um die Qualität möglichst aller eingeschlossener Studien miteinander vergleichen zu können.

Ferner kann es zu Fehlern bei der Erhebung von Daten aus den Studien gekommen sein. Die Datenextraktion gestaltete sich aufgrund der hohen Heterogenität der Studien als schwierig, da keine klar zu extrahierenden Variablen vorgegeben werden konnten.

3.4.2 Beantwortung der Fragestellung

Die Literaturrecherche zur Evidenzbasierung in der Sozialmedizinischen Begutachtung untersuchte die Fragestellungen, die in Tabelle 11 zusammengetragen wurden:

1. Wie ist der Stand der Evidenzbasierung der Sozialmedizinischen Begutachtung bei der Beurteilung von Anträgen zur EM-Rente?

Nach Beurteilung der Evidenz durch das adaptierte Bewertungsschema von Kliche und Kollegen (2006), das den Evidenzgrad anhand des Studiendesigns (d.h. mit der Möglichkeit der Verallgemeinerung der Ergebnisse beurteilt) wurden für die Mehrheit der eingeschlossenen Studien keine höheren Evidenzgrade festgestellt. Auffällig ist der hohe Anteil an narrativen Reviews (Evidenzklasse: V) und Beobachtungsstudien (Evidenzklasse: IV).

Die Ergebnisse bestätigen die Resultate des „LitBegut“ Forschungsprojektes der DRV Bund, nach dem in der Sozialmedizinischen Begutachtungsliteratur die Evidenzbasierung häufig auf Expertenkonsens „limitiert“ sei. Von den in der vorliegenden Studie insgesamt 16 Artikeln wurden zehn mit dem Evidenzgrad V beurteilt. Diese Beiträge beschreiben ausschließlich Positionen von Fachleuten oder Expertenkreisen, Konsensuskonferenzen und/oder klinische Erfahrungen anerkannter Autoritäten und beinhalten keine aktiv erhobenen Forschungsdaten. Qualitativ hochwertige Studiendesigns (RCT, quasi-experimentelle Studien etc.) sind in diesem Themenfeld rar. Dies könnte darin begründet sein, dass sie einer weitreichenden Planung und Lösung eigenständiger Untersuchungsprobleme bedürfen. Dementsprechend ist die vergleichsweise hohe Anzahl an narrativen Reviews bei

der sozialmedizinischen Sachaufklärung Ausdruck der Komplexität des sozialmedizinischen Begutachtungsgeschehens, welches sich nur unter größeren Anstrengungen mit höheren Evidenzklassen „beforschen“ lässt. Diese Komplexität beruht nicht zuletzt auf der Individualität des Begutachtungsgeschehens auf Basis der biopsychosozialen Verankerung in der Sozialmedizin. In diesem Zusammenhang zeigt sich zudem eine Divergenz zwischen der individuellen Begutachtung und der allgemeinen Forderung nach evidenten und verallgemeinerbaren Forschungsergebnissen, die originär nur mit RCTs und deren Aggregation durch Meta-Analysen erreicht werden können.

Dieser Sachverhalt spiegelt sich auch in den Leitlinien zur Sozialmedizinischen Begutachtung wider (siehe Kap. 1.6.2). Diese Leitlinien wurde ausnahmslos als S1-Leitlinien konzipiert und wurden daher ausschließlich durch repräsentativ zusammengesetzte Expertengruppen im informellen Konsens erstellt (vgl. AWMF, 2018). Infolgedessen, dass es keine hinreichend geklärten evidenten Ergebnisse zur Prognose einer Leistungseinschränkung gibt, ist eine andere Evidenzstufe für die Entwicklung einer Leitlinie kaum realisierbar. Im Allgemeinen besteht in der Sozialmedizin die Problematik, dass das komplexe sozialmedizinische Begutachtungs- und Interaktionsgeschehen die Möglichkeiten der evidenzbasierten Forschung determiniert.

2. Gibt es evidente Entscheidungsalgorithmen für sozialmedizinische Gutachter?

Aus dem Portfolio der elf identifizierten Studien zur Begutachtungspraxis konnten drei Publikationen identifiziert werden, die einen konkreten Algorithmus zur Einschätzung der Leistungsfähigkeit beschreiben (Hoffmann et al., 2016; Kellner, 2002; Kolenda, 1998). Davon stellt jedoch lediglich eine Publikation die konkrete Ableitung des zeitlichen Restleistungsvermögens im Rahmen der Sozialmedizinischen Begutachtung bei Prostatakarzinomen dar. Die beiden anderen Studien beinhalten Bewertungskriterien zur Beurteilung des GdB in Prozent bei seronegativer Spondylarthropathien und KHK. Die Bewertungen werden gestützt durch medizinische Befunde, Leistungsbeeinträchtigungen, der zumutbaren beruflichen Tätigkeiten (Arbeitsschwere, Arbeitshaltung, Arbeitsorganisation), relevanten Gefährdungs- und Belastungsfaktoren sowie dem Stadium der Erkrankung. Die Evidenz der Entscheidungsalgorithmen wird in allen drei Studien mit der Evidenzklasse V bewertet, da sie in narrativen Reviews entwickelt wurden.

Im Lehrbuch zur Sozialmedizinischen Begutachtung für die GKV finden sich im diagnosespezifischen Teil ebenfalls Hinweise zur Sozialmedizinische Begutachtung. Infolge der Komplexität und Individualität des Sozialmedizinischen Begutachtungsgeschehens werden an dieser Stelle ebenfalls keine allgemeingültigen Entscheidungsalgorithmen publiziert. Stattdessen werden indikationsspezifische

Begutachungskriterien aus den ICF Bereichen „Körperstrukturen und -funktionen“ sowie „Aktivitäten und Teilhabe“ formuliert, die während der Begutachtung erhoben und berücksichtigt werden. Eine Verknüpfung zur Leistungsfähigkeit wird nicht hergestellt.

3. Gibt es Lücken oder Schwächen in der Studienlage zur Begutachtung?

Obgleich in der Recherche nach den wichtigsten Diagnosen für Beantragung einer EM-Rente gesucht wurde, konnten innerhalb der Ergebnisse nur Studien zu KHK (Kolenda, 1998), muskuloskelettalen Schmerzen (Schiltenswolf, 2007), (Prostata-)Karzinomen (Hoffmann et al., 2016; van Muijen et al., 2015), psychische Störungen (Canela et al., 2016; Dewa et al., 2015; Linden, 2017), rheumatoider Arthritis (Backman, 2004; Hammer, 2002) sowie seronegativen Spondylarthropathien (Kellner, 2002) gefunden werden. Weitere Themenbereiche waren die Anwendung der ICF sowie die Untersuchung der Erfahrungen von Ärzten und Klinikern bei der Beurteilung von Funktionseinschränkungen. Es wurden keine Publikationen zu Krankheiten des Atmungssystems, Gastroenterologische und Stoffwechselkrankheiten, Adipositas, Mamma-Karzinom, Neurologische Krankheiten sowie Bandscheiben- und bandscheibenassoziierten Erkrankungen identifiziert.

Weitere Schwächen sind in der eigentlichen Evidenz und in der Qualität der Studien zu suchen. Aufgrund des Studiendesigns (viele narrative Reviews) scheint auch die Studienqualität limitiert zu sein, da bestimmte Kriterien des MMAT nicht beantwortet werden können. Wie bereits ausgeführt, beruhen die Limitationen der Studienlage auf der multidimensionalen biopsychosozialen Verankerung der Sozialmedizin, die keine höherwertigen Studiendesigns zulässt.

Die identifizierten Studien eignen sich aus diesem Grund auch nur begrenzt zur Weiterentwicklung sozialmedizinischer Leitlinien. Alle durch die DRV veröffentlichten Leitlinien sind S1 Leitlinien. Sowohl die vorliegende Arbeit als auch die Ergebnisse des Forschungsprojektes „LitBegut“ konnten keine relevanten Studien identifizieren, welche die Evidenzbasierung bei der Erstellung und Überarbeitung dieser Leitlinien erhöht. Gleichwohl sei angemerkt, dass derzeit von den durch die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) aufgelisteten 963 Leitlinien (Stand: 25.11.2018) insgesamt 363 auch der Evidenzklasse S1 angehören. Dabei ist die Gültigkeit einer Leitlinie nicht ausschließlich durch deren methodischer Qualität determiniert, sondern auch vom Bedarf, der Thematik, der Dissemination, der Handhabbarkeit sowie deren Auswirkungen auf Ergebnisse der Versorgungs- bzw. Begutachtungsleistung (Muche-Borowski & Kopp, 2011). Entsprechend sind Studienlage und Evidenz der Leitlinien zwar limitiert, die Gewichtung für die Praxis der gutachterlich tätigen Ärzte jedoch bedeutend.

3.5 Handlungsempfehlungen

Die Qualität der Sozialmedizinischen Begutachtung hängt von der adäquaten Durchführung verschiedener Teilbereiche der Gutachtenerstellung ab. Dazu gehören:

- die Einschätzungen der Auswirkungen bestimmter Störungen von Körperfunktionen und -strukturen auf die Einschränkung der Funktionsfähigkeit (im Sinne der ICF). Diese Einschätzungen werden maßgeblich beeinflusst von Anzahl und Wechselwirkung der vorhandenen Störungen (Multimorbidität),
- die Einschätzung der Auswirkungen der diagnostizierten eingeschränkten Funktionsfähigkeit auf die Teilhabe und speziell auf die Leistungseinschränkung im Erwerbsleben. Hier werden die Kontextfaktoren im Sinne der ICF bedeutsam. Einerseits können Umweltaspekte eine Leistungs- und Teilhabeeinschränkung beeinflussen bzw. vermindern, andererseits die Einstellungen, Motivation und Lebensweise von Patienten diese Einschränkung modulieren,
- die Einschätzung über die Wahrscheinlichkeiten, durch gesundheitliche Behandlungen eine Verbesserung der Einschränkung zu erzielen.

Eine evidenzbasierte Weiterentwicklung der Sozialmedizinischen Begutachtung muss sich auf wissenschaftliche Wissensressourcen in allen oben genannten Bereichen beziehen können. Die in dieser Studie vorgenommene Recherche nach wissenschaftlicher Evidenz hat sich allgemein auf Studien zur Sozialmedizinischen Begutachtung gerichtet und keine höhere Evidenzklassen identifizieren können.

Literaturrecherche anhand der gutachterlichen Argumentationskette. Im Rahmen einer erneuten Literaturrecherche zur Evidenz in der Sozialmedizinischen Begutachtung sollten die o.g. Teilbereiche einzeln und unabhängig voneinander untersucht werden, um die dort vorhandenen Evidenzen zu erheben. Diese stammen möglicherweise aus anderen Wissenschaftsfeldern und müssen in einem weiteren Bearbeitungsschritt auf die Sozialmedizinische Begutachtung übertragen werden. Die vorliegende Untersuchung zeigt auch, dass eine indikationsspezifische Recherche zu allgemein gefasst zu sein scheint, um relevante Evidenzen zu identifizieren. Das Fokussieren auf einzelne Bestandteile des Begutachtungsprozesses könnte ein Ansatz für weitere Literaturrecherchen auf diesem Themengebiet sein. In dieser Recherche sollte daher auch die gutachterliche Argumentationskette bestehend aus „Fragestellung“, „Informationssammlung“, „Informationsbewertung“ und „Beantwortung der Fragestellung“ (vgl. Abb. 6) im Einzelnen berücksichtigt werden. Dazu müssten gänzlich andere Fragestellungen und Suchtermini erarbeitet werden, die nicht mehr per Definition mit dem Begriff der Sozialmedizinischen Begutachtung verknüpft sein müssen.

Bezogen auf den Argumentationskettenbaustein „Informationssammlung“ (bestehend aus Anamnese und körperlicher Untersuchung) sollte mittels einer Literaturrecherche systematisch und indikationsspezifisch untersucht werden, welche Assessmentverfahren zur Messung der Funktionsfähigkeit für jede einzelne Erkrankung vorliegen und eingesetzt werden können, welche anamnestischen Parameter bei der Befundung essenziell sind und welche Parameter einen signifikanten Einfluss auf Erwerbsfähigkeit des Patienten ausüben (siehe Kap. 3.5, Abschnitt Prädiktoranalysen). Bezogen auf die beiden miteinander verknüpften Bausteine „Informationsbewertung“ (bestehend aus Funktionsdiagnosen und Epikrise) und „Beantwortung der Fragestellung“ (bestehend aus der Sozialmedizinischen Beurteilung der Leistungsfähigkeit und des Reha-Bedarfs) sollte mittels einer weiteren Literaturrecherche untersucht werden, welche Diagnosen oder Fallkonstellationen typischerweise zur Einschränkung der Leistungsfähigkeit führen.

Bisher wurde ausschließlich versucht mittels „globaler“ Literaturrecherchen die Evidenz in der Begutachtung zu erfassen. Die bisherigen Anstrengungen belegen jedoch, dass über diese top-down Vorgehensweise sehr heterogene Ergebnisse identifiziert wurden und ein Rückschluss auf die konkrete Arbeitsweise eines Gutachters nicht gezogen werden konnte. Um evidente Entscheidungsalgorithmen für Gutachter auszuarbeiten, wird daher nicht empfohlen, explizit nach solchen Algorithmen zu suchen, sondern die einzelnen Begutachtungsschritte mit ihren jeweiligen eigenen Fragestellungen zu untersuchen. Im Sinne eines induktiven Bottom-Up-Ansatzes könnten somit einzelne evidente Ergebnisse zu einem übergeordnetem Entscheidungsalgorithmus verknüpft werden.

Literaturrecherche zu Prädiktoranalysen. In der vorliegenden Recherche wurden 25 Prädiktoranalysen identifiziert, die nicht in das Review aufgenommen wurden, da kein direkter Zusammenhang mit der Sozialmedizinischen Begutachtung bestand. Die Analyse solcher Studiendesigns gibt jedoch zusätzliche Hinweise über relevante zu begutachtende Risikofaktoren oder Folgeerkrankungen. Die ausgeschlossenen Prädiktoranalysen untersuchten hauptsächlich Variablen zur Vorhersage der Arbeits(un)fähigkeit (work disability/work ability/sick leave) und Invaliditätsrente (disability pension). Häufig benannte Prädiktoren, wie Geschlecht, Alter, AU-Tage, wahrgenommene gesundheitliche Beschwerden, Einschränkungen der körperlichen Aktivität, Komorbiditäten, schwere körperliche Arbeit usw. geben einerseits Auskunft über die statistische Wahrscheinlichkeit, arbeitsunfähig oder erwerbsgemindert zu werden, können andererseits jedoch nicht die genaue Einschränkung des Restleistungsvermögens vorhersagen. Da die Leistungs- oder Funktionsfähigkeit im Sinne der ICF durch die Wechselwirkungen zwischen den Anforderungen der Arbeitsumwelt und der aktuellen Leistungsfähigkeit abgeleitet wird,

können Prädiktoren allenfalls als Risikofaktoren dienen, welche in die abschließende sozialmedizinische Leistungsbeurteilung einfließen können.

Prospektiv sollten diese Prädiktoren dennoch in einer separaten Literaturrecherche erfasst und indikationsspezifisch ausgewertet werden, um die bestehenden Sozialmedizinischen Begutachtungsleitlinien dahingehend zu ergänzen. Die Aufgabe des Gutachters besteht dann darin, die erhobenen Information zu bewerten und korrekt zu verknüpfen. Es ist davon auszugehen, dass die identifizierten Studien dieser Literaturrecherche höhere Evidenzniveaus erreichen werden, da Prädiktoranalysen häufig im Rahmen von Fall-Kontroll-Studien oder prospektiven Kohortenstudien (Evidenzklasse III nach Kliche et al. (2006)) untersucht werden.

3.6 Kernbotschaft

Das vorliegende Review lässt darauf schließen, dass die Evidenzlage zur Sozialmedizinischen Begutachtung bei Anträgen auf EM-Rente begrenzt ist. Einige konkrete Hinweise für die Begutachtung der Erwerbsfähigkeit ausgewählter Erkrankungen konnten identifiziert werden. Offen bleibt jedoch die ubiquitäre Übertragbarkeit der Ergebnisse auf die gutachterliche Praxis, da die Ergebnisse zumeist auf narrativen Reviews beruhen. Die gefundenen Studien stammen selten aus einer umfassenden Forschungsstrategie und -programm, sondern sind Ergebnis singulärer Forschungsgruppen und Antragsinitiativen. Eine hohe Evidenzbasierung kann die Leitlinienentwicklung, das professionelle Handeln der gutachterlich tätigen Ärzte sowie die Stellung der Sozialmedizinischen Begutachtung im Gesundheitswesen nachhaltig positiv beeinflussen. Angesichts der großen Bedeutung der Qualität der Sozialmedizinischen Begutachtung, sowohl für den einzelnen betroffenen Menschen als auch für die DRV, wäre zukünftig eine konsentrierte Forschungsförderung zur Validität und Reliabilität des Begutachtungsgeschehens indiziert.

4 Teilstudie 2: Interrater-Reliabilität des Peer Review Verfahrens

Die zweite Teilstudie beschäftigt sich mit der Qualitätssicherung sozialmedizinischer Gutachten zur EM-Rente. Mittels eines für den sozialmedizinischen Forschungskontext konzipierten und durchgeführten Peer Review-Verfahrens wird in Kooperation mit zwölf RV-Trägern die Umsetzbarkeit und Interrater-Reliabilität des durch die DRV Bund und der Universität Würzburg entwickelten trägerübergreifenden Qualitätssicherungssystems untersucht. Ziel der Untersuchung ist die Validierung und ggf. Weiterentwicklung des Qualitätssicherungsprogramms der DRV Bund. In der vorliegenden Teilstudie werden neben dem Einsatz quantitativer Methoden auch mittels qualitativer Verfahren die Gründe für fehlende Übereinstimmungen der teilnehmenden Sozialmediziner untersucht.

Dem Qualitätssicherungskonzept für die Sozialmedizinische Begutachtung, welches ausgehend von den Vorschlägen SOMEKO auf Initiative der DRV durch die Projektgruppe „Qualitätssicherung der Sozialmedizinischen Begutachtung“ erarbeitet wurde, liegt als Methode ein Peer-Review-Verfahren für die trägerübergreifende Qualitätsprüfung zu Grunde. Dabei sollen Gutachten anhand der Qualitätskriterien, die die verschiedenen qualitativen Aspekte eines Gutachtens widerspiegeln, durch erfahrene Sozialmediziner aus den Häusern der Rentenversicherungsträger bewertet werden (vgl. Kapitel 1.7). Die systematische Rückmeldung der Ergebnisse soll die Transparenz der Gutachten für Nutzer (z. B. Ärzte, Auftraggeber) erhöhen und eine trägerübergreifende Harmonisierung und nachhaltige Verbesserung des Begutachtungsgeschehens ermöglichen.

Die Daten zur Interrater-Reliabilität wurden in Teilen bereits in der Zeitschrift „Das Gesundheitswesen“ publiziert (Strahl et al., 2016). Im Gegensatz zu den veröffentlichten Studiendaten beruhen die in dieser Arbeit dargestellten Hauptanalysen zur Interrater-Reliabilität jedoch auf einer anderen Teilmenge. Berichten Strahl und Kollegen (ebd.) ihre Ergebnisse auf Basis von 771 Reviews, die jeweils über Peer-Paare ausgewertet wurden, werden in der vorliegenden Teilstudie 325 Reviews für die Interrater-Reliabilitätsanalyse ausgewertet, die simultan von allen Peers beurteilt wurden. Dieses Analyseverfahren bietet die Möglichkeit, die Überstimmung gepoolt für alle Peers abzubilden und die Tendenzen zu extremen Mängelbeurteilungen besser zu erklären.

Ergänzend werden qualitative Analysen der freitextlichen Mängelbegründungen der Peers berichtet, die bisher nicht publiziert wurden.

4.1 Fragestellung

Ziel der Untersuchung ist die Analyse der Interrater-Reliabilität eines Peer Review-Verfahrens als Qualitätssicherungsmethode im Bereich der Sozialmedizinischen Begutachtung unter der Nutzung des „Prüffragenkatalogs zur Qualitätssicherung sozialmedizinischer Gutachten“ (vgl. DRV, 2009c; Löffler et al., 2010). Eine hohe Übereinstimmung bei der Prüfung sozialmedizinischer Gutachten ist Voraussetzung für die Akzeptanz des Verfahrens als Qualitätssicherungsmethode in der Sozialmedizin. Mit der vorliegenden Untersuchung soll das trägerübergreifende Review-Verfahren der DRV Bund bewertet und die Ergebnisse von Strahl und Kollegen (2016) validiert werden.

Die Überprüfung der Interrater-Reliabilität ist vor dem Hintergrund der Evaluationsergebnisse der Peer-Review-Verfahrens in der medizinischen Rehabilitation in Deutschland dringend geboten. Farin et al. (2003) berichten, dass die Übereinstimmung bezüglich des Peer-Review-Urteils eines einzelnen Peers im Verfahren der gesetzlichen Rentenversicherung nicht immer zufriedenstellend ausgefallen sei. Ferner stellte Neuderth (2004, S. 230) bei der Erprobung eines Peer Reviews für die gesetzliche Unfallversicherung fest, dass die Güte des Verfahrens nicht ausreicht, wenn davon auszugehen ist, dass in der Routine jeder zu beurteilende Fall nur von einem Peer bewertet würde.

Neben der Höhe der Interrater-Reliabilität ist es ebenso von Bedeutung, aus welchen Gründen Unterschiede bei der Beurteilung zustande kommen. Hieraus könnte man mögliche Schulungsbedarfe für im Verfahren tätige Peers ableiten. Ziel der Untersuchung ist die Identifikation von Gründen für intraindividuelle Peer-Bewertungsmaßstäbe. Das Review-Verfahren selbst bietet die Möglichkeit der qualitativen Analyse aller Mängelbegründungen, da festgestellte Defizite im Gutachten durch die beurteilenden Peers schriftlich dokumentiert werden müssen. Entsprechend werden die in Tabelle 20 dargestellten Hauptfragestellungen im Rahmen der vorliegenden Studie untersucht.

Tabelle 20: Fragestellungen zur Analyse der Interrater-Reliabilität des Peer Review Verfahrens.

1. Hauptfragestellung
Wie hoch ist die Interrater-Reliabilität des Peer Review-Verfahrens zur Qualitätssicherung in der Sozialmedizinischen Begutachtung?
2. Nebenfragestellungen
Welche freitextlichen Mängelbegründungen werden genutzt, um detektierte Gutachtendefizite zu begründen?
Können typische Qualitätsproblemstellungen in den freitextlichen Mängelbegründungen identifiziert und systematisiert werden?

4.2 Untersuchungsdesign

Zur Prüfung der Interrater-Reliabilität wurde ein indikations- und trägerübergreifendes Peer Review-Verfahren im Rahmen einer Pilotstudie durchgeführt. Dabei beurteilten Sozialmediziner die Gutachten anderer RV-Träger anhand eines Prüffragenkatalogs zur Qualitätssicherung sozialmedizinischer Gutachten. Der Prüffragenkatalog ist im Manual zum Peer Review hinterlegt. Die Studie wurde in enger Kooperation mit der DRV Bund, Bereich Sozialmedizin und den teilnehmenden Regionalträgern durchgeführt. Das Projekt untergliederte sich in drei Arbeitsphasen:

- 1. Rekrutierung und Schulung:** Nach Rekrutierung der an der Studie teilnehmenden Sozialmediziner wurden diese in einem zweitägigen Peer-Schulungsworkshop im Umgang mit dem Manual und Prüffragenkatalog zum Peer Review-Verfahren geschult. Im Anschluss erfolgte eine erste Erprobung des Manuals anhand von drei anonymisierten Beispielgutachten, die von der DRV zur Verfügung gestellt wurden.
- 2. Peer Review:** In dieser Phase wurden die sozialmedizinischen Gutachten zu den entsprechenden Peers zugeordnet und das Peer Review durchgeführt.
- 3. Kalibrierung:** Nach Abschluss des Reviews wurden alle Peers in einen zweitägigen Workshop zur Kalibrierung der eigenen Beurteilungsmaßstäbe eingeladen. Die Teilnehmer erhielten Feedback über das Beurteilerverhalten und bekamen die Möglichkeit, kontrovers beurteilte Gutachten zu diskutieren und erneut zu beurteilen.

Technische Umsetzung des Reviews. Die Bearbeitung und der Versand der Reviews erfolgten ausschließlich in digitaler Form. Die Rentenversicherung Bund lizenzierte die Evaluations- und Umfragesoftware EvaSys®, um das Peer Review zu realisieren. Dazu wurde jedes Gutachten durch die DRV mit einer individuellen TAN-Nummer versehen. Durch Eingabe der TAN-Nummer konnte jeder Peer den digitalisierten Prüffragenkatalog einsehen und bearbeiten. Nach Abschluss des Reviews wurden die Ergebnisse an die DRV Bund, Bereich Sozialmedizin gesandt und von dort in pseudonymisierter Form den bearbeitenden Wissenschaftlern der Universität Würzburg zur Verfügung gestellt.

Datenschutz und Ethik. Das Datenschutzkonzept wurde vom Datenschutzbeauftragten der DRV Bund geprüft und bewilligt. Weiterhin wurden sämtliche Regionalträger der RV durch die DRV Bund angeschrieben und über das Projekt informiert sowie deren Einverständnis zur Teilnahme eingeholt. Für die Auswertung der Forschungsdaten wurde kein Ethikvotum beantragt, da die bereitgestellten Daten durch die DRV Bund vollständig pseudonymisiert wurden. Entsprechend dieses Vorgehens ließen sich keine direkten Rückschlüsse auf den einzelnen Peer oder RV-Träger ziehen.

4.2.1 Rekrutierung

Die Rekrutierung der am Peer Review teilnehmenden Peers erfolgte in Zusammenarbeit mit der DRV Bund, Bereich Sozialmedizin. Bei einer Auftaktveranstaltung in Berlin wurde das Vorhaben den Leitenden Ärzten aller 16 RV-Träger vorgestellt und um Teilnahme am Projekt gebeten. Die Teilnahme der Träger und Peers war freiwillig. Insgesamt erklärten sich 21 Sozialmediziner von zwölf RV-Trägern bereit, an der Untersuchung teilzunehmen.

4.2.2 Schulung

In einem zweitägigen Peer-Schulungsworkshop wurden die als Peers gemeldeten Sozialmediziner im Umgang mit dem Manual zum Peer Review-Verfahren geschult. Nachdem eine allgemeine Einführung in das Manual und den Prüffragenkatalog gegeben wurde, sollten die Peers anhand von zwei Beispielgutachten konkret mit der Handhabung des Instruments vertraut gemacht werden.

Die Schulung beinhaltete einen Vortrag zur Entwicklung der Qualitätssicherung innerhalb der DRV sowie die Vorstellung aller Prüffragen und deren hinterlegte Mängelkategorien. Ausgehend vom SOMEKO-Abschlussbericht (VDR, 2004) wurden die vier Projektinitiativen „Sozialmedizinisches Glossar“, „Selbstauskunftsbogen zum Leistungsantrag“, „Leitlinien für die Sozialmedizinische Begutachtung“ und „Qualitätssicherung sozialmedizinischer Begutachtung“ inhaltlich vorgestellt. Als Grundlage für das spätere Arbeiten mit Manual und Prüffragenkatalog wird das Manual an beispielhaft ausgewählten Prüffragen in der Großgruppe vorgestellt. Da der Fokus in der sich anschließenden Kleingruppenphase auf dem Umgang mit dem Manual und dem Prüfbogen lag, wurden grundsätzliche Debatten über die Qualitätskriterien und Prüffragen selbst vermieden.

Die anschließende Gruppenarbeit wurde in drei Kleingruppen mit jeweils sieben Personen realisiert. Dabei wurde darauf geachtet, dass Sozialmediziner, die schon am bisherigen Entwicklungsprozess des Manuals beteiligt waren, systematisch auf die drei Kleingruppen verteilt wurden. Ziel war es, jeder Arbeitsgruppe mindestens einen Peer zuzuordnen, der im weitesten Sinne mit dem Manual vertraut war. Jede Gruppe wurde durch einen methodisch versierten Moderator der Universität Würzburg angeleitet. In den Kleingruppen wurden die Prüffragen anhand des ersten Gutachtens sukzessiv mit den Peers bearbeitet. Nach dem Verlesen einer Prüffrage, beurteilte jeder Peer diese mündlich. Divergentes Antwortverhalten wurde im Diskurs durch die Peers angeglichen. Ziel des ersten Tages war das Kennenlernen und Einschätzen jeder Prüffrage in Form einer diskursiven konsensorientierten Kleingruppenarbeit. Am zweiten Schulungstag wurde neben der Beurteilung eines weiteren Gutachtens die Dokumentation festgestellter Mängel in der

Kleingruppe geübt. Die Peers sollten in die Lage versetzt werden, sinn- und gehaltvolle Begründungen für ihre Beurteilung zu formulieren.

Probegutachten. Nach der Schulung bearbeiteten alle Peers drei gleiche anonymisierte sozialmedizinische Gutachten, um den autonomen Umgang mit dem Manual zu üben und zu verfestigen. Mit einer teilstandardisierten Telefonbefragung aller Peers wurde die Schulungsphase abgeschlossen. Das leitfadengestützte Telefoninterview erhob Daten zum Bearbeitungsstand der Probegutachten, zu Schwierigkeiten und offene Fragen im Umgang mit dem Manual, zu technischen Schwierigkeiten mit EvaSys® sowie allgemeine Anmerkungen.

Ein Peer verließ das Projekt ohne Angabe von Gründen nach der Schulungsphase. Infolgedessen wurde die Methodik des Peer Reviews für 20 Peers konzipiert.

4.2.3 Peer Review

Nach der Schulung der Peers und Erprobung des Manuals folgte eine fünfmonatige Reviewphase. In diesem Zeitraum beurteilten die Peers sozialmedizinische Erstgutachten zur EM-Rente im Rahmen des Peer Reviews. Der Zeitraum wurde a priori nicht festgelegt, sondern richtete sich nach dem Rücklauf der Reviews. Da die Gutachten i.d.R. berufsbegleitend beurteilt wurden, sollten alle Peers einen ausreichenden zeitlichen Rahmen bekommen, der benötigt wurde, um möglichst alle Reviews abzuschließen.

Reviewobjekte (Gutachten). Im Rahmen der Gesamtstudie wurden insgesamt 260 anonymisierte Gutachten beurteilt. Bei den Gutachten handelte es sich um eine Zufallsauswahl der teilnehmenden RV-Träger, wobei jeder der zwölf teilnehmenden Träger mindestens 20 Gutachten zur Verfügung stellt. Diese insgesamt 240 Gutachten wurden von je zwei Peers beurteilt. Ergänzend wurden 20 zufällige Gutachten ausgewählt, die von allen 20 Peers anhand des Prüffragenkatalogs beurteilt werden sollen. Die Regionalträger nahmen eine Anonymisierung bezüglich des Trägers, des Antragstellers und des Gutachters vor, sodass keine Rückschlüsse auf die beteiligten Personen und Institutionen möglich sind. Die Gutachten entstammen aus den drei übergeordneten medizinischen Indikationsbereichen Innere Medizin/Allgemeinmedizin, Neurologie/Psychiatrie sowie Orthopädie/Chirurgie.

Zum Einschluss in die Studie und Teilnahme am Projekt mussten die sozialmedizinischen Gutachten folgende Einschlusskriterien erfüllen:

- Erstgutachten auf EM-Rente,
- Gutachten mit körperlicher Untersuchung,
- letzter Rehabilitationsantrag des Antragstellers muss länger als ein Jahr zurück liegen.

Die Zuordnung der Peers zu den Gutachten erfolgte randomisiert. Um die Interrater-Reliabilität bestimmen zu können, müssen die Gutachten von mindestens zwei Sozialmediziner*innen beurteilt werden. Für die Aufteilung der zu beurteilenden Gutachten wurde in Anlehnung an Neuderth (2004, S. 133-134) ein Verteilungsschema entwickelt, welches folgende Anforderungen erfüllt:

- jedes Gutachten wird von mindestens zwei Peers beurteilt,
- die Peers beurteilen keine Fälle aus ihrem eigenen RV-Träger,
- die Fälle eines Trägers sollten möglichst auf viele verschiedene Peers verteilt werden, um intraindividuelle Beurteilerunterschiede auszugleichen
- aus jedem RV-Träger sollen nicht mehr 23 Gutachten beurteilt werden.

Tabelle 21 zeigt einen Ausschnitt des konstruierten Verteilungsschemas der Zuordnungssystematik.

Tabelle 21: Ausschnitt aus Schema für die Gutachtenverteilung auf die Peers.

Träger	Gutachten	Peer 1	Peer 2	Peer 3	Peer 4	Peer 5	Peer 6	Peer 7	Peer 8	Peer 9	Peer 10	Peer 11	Peer 12	Peer 13	Peer 14	Peer 15	Peer 16	Peer 17	Peer 18	Peer 19	Peer 20
Träger	#1					X	X														
	#2						X	X													
	#3							X	X												
	#4								X	X											
	#5									X	X										
	#6										X	X									
	#7											X	X								
	#8												X	X							
	#9	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
	#10														X	X	X	X	X	X	X
	#11															X	X	X	X	X	X
	#12																X	X	X	X	X
	#13																	X	X	X	X
	#14																		X	X	X
	#15	X																		X	X
	#16	X	X																		X
	#17		X	X																	
	#18			X	X																
	#19				X	X															
	#20					X	X														

Am linken Rand des Schemas sind die Rentenversicherungsträger abgetragen (Spalte „Träger“). Die weiteren Spalten des Verteilungsschemas repräsentieren die Gutachten sowie die einzelnen Peers. Die zweite Spalte (Gutachten) kennzeichnet die Gutachten, die für das Peer Review-Verfahren ausgewählt wurden. Die Peers, an die das jeweilige Gutachten zum Beurteilen gesendet wurde, sind in den Tabellenzeilen mit einem „X“ markiert. Dem Ausschnitt ist zu entnehmen, dass immer zwei Peers ein Gutachten beurteilen sollen. Da kein Peer Fälle aus seinem eigenen Träger erhalten darf, sind die entsprechenden Felder im Verteilungsschema „geblockt“ (Peer₁₃). Nach dem Verteilungsschema waren insgesamt 20 „Peer-Doubletten“ vorgesehen. Dafür wurden 20 feste „Peer-Paare“ gebildet, die jeweils eine bestimmte Anzahl an Gutachten (n = 12) parallel beurteilten. Gutachten #9 wurde zusätzlich von allen Peers bearbeitet.

Beurteilungskriterien (Prüffragenkatalog). Das Manual zum Peer Review-Verfahren mit seinem Prüffragenkatalog (vgl. Kap. 1.7.1) bildete die Bewertungsgrundlage für die Qualitätssicherung. Jedes Gutachten wurde anhand von sechs Einzelkriterien und des übergeordneten Kriteriums beurteilt. Während das übergeordnete Kriterium „Nachvollziehbarkeit des Gutachtens“ im Kap. 1.7.1 genauer beschrieben wurde, wird im Folgenden näher auf die Fragen und Inhalte der Einzelkriterien eingegangen.

Das Einzelkriterium „**Formale Gestaltung**“ wird durch zwei Prüffragen abgebildet, die die formalen Anforderungen an das Gutachten gemäß den Gliederungsvorgaben (vgl. VDR, 2001) und die Nutzung des einheitlichen Formularsatz der DRV, bestehend aus Deck- und Schlussblatt Teil 1 und 2, prüfen. Die Qualitätsbeurteilung dieser Prüffragen bezieht sich auf formale Aspekte des Gutachtens. Die Untersuchung der Vollständigkeit oder Nachvollziehbarkeit der Darstellung sind nicht Gegenstand der Bewertung dieses Einzelkriteriums (Löffler et al., 2010, Anhang 12, S. 6).

Das Einzelkriterium „**Verständlichkeit**“ wird mittels vier Prüffragen geprüft. Dabei beurteilen die Peers die sprachlich unmissverständliche Ausdrucksweise des Gutachtens, den Umgang mit sozialrechtlichen Begriffen und Aussagen, die allgemeinverständliche Erklärung aller im Gutachten genutzten Fachbegriffe und Abkürzungen in den Diagnosen und der Epikrise sowie die richtige Anwendung der sozialmedizinischen Terminologie (Löffler et al., 2010, Anhang 12, S. 9).

Das Einzelkriterium „**Transparenz**“ wird mittels drei Prüffragen geprüft. Es wird überprüft, ob die relevanten verwendeten Verfahren, Methoden und Instrumente zur Befundung des Antragstellers im Gutachten beschrieben wurden. Zusätzlich sollen alle Messwerte und Referenzbereiche der selbst erhobenen apparativ-technischen und psychometrischen Untersuchungen angegeben werden. Ferner wird die transparente Darstellung der Herkunft von medizinischen Informationen untersucht (Löffler et al., 2010, Anhang 12, S. 14).

Das Einzelkriterium „**Vollständigkeit**“ wird mittels neun Prüffragen evaluiert. Das Kriterium prüft ausschließlich die Vollständigkeit der Gutachteninhalte. Da keine Aussagen über die Richtigkeit der Angaben gemacht werden können, ist dieser Aspekt seitens der Peers nicht zu beurteilen. Geprüft wird die vollständig dokumentierte Erhebung der Anamnese, Befunde, Diagnosen und Epikrise sowie der Beantwortung der sozialmedizinischen Fragestellung. Hierbei sind auch Stellungnahmen zu bisherigen Therapien, zu weiteren therapeutischen Optionen sowie zu Abweichungen in der Beurteilung gegenüber Vorgutachten zu beurteilen (Löffler et al., 2010, Anhang 12, S. 18). Dieses Einzelkriterium definiert indirekt den inhaltlich definierten Standard, der seitens der DRV an Gutachten herangetragen wird.

Das Einzelkriterium „**Medizinisch-wissenschaftliche Grundlagen**“ wird wieder durch zwei Prüffragen abgebildet, mit denen der beurteilende Peer prüft, inwieweit der anerkannte Stand medizinischer Erkenntnisse eingehalten und die Begutachtungsliteratur der DRV berücksichtigt wurde (Löffler et al., 2010, Anhang 12, S. 30).

Das Einzelkriterium „**Wirtschaftlichkeit**“ wird ebenfalls durch zwei Prüffragen abgebildet. Im Mittelpunkt der Prüfung steht die Wirtschaftlichkeit der eingesetzten Diagnostik. Die Peers beurteilen, ob die eingesetzte Diagnostik ausreichend bzw. notwendig war, um in der Epikrise Schlüsse auf das Restleistungsvermögen ziehen zu können (Löffler et al., 2010, Anhang 12, S. 33).

In Tabelle 22 ist der Wortlaut der einzelnen Prüffragen der Einzelkriterien überblicksartig dargestellt. Der vollständige Prüffragenkatalog befindet sich im *Anhang 5*.

Tabelle 22: Überblick Prüffragenkatalog des Manuals zum Peer Review-Verfahren (Löffler et al., 2010, Anhang 12, S.4).

Einzelkriterium A: Formale Gestaltung	
Prüffrage A.1	Inwieweit werden die Gliederungsvorgaben gemäß den Anforderungen eingehalten?
Prüffrage A.2	Inwieweit wird der einheitliche Formularsatz bestehend aus Deck- und Schlussblatt verwendet?
Einzelkriterium B: Verständlichkeit	
Prüffrage B.1	Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich?
Prüffrage B.2	Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt?
Prüffrage B.3	Inwieweit wird die sozialmedizinische Terminologie richtig angewendet?
Prüffrage B.4	Inwieweit werden sozialrechtliche Ausführungen/Folgerungen unterlassen?
Einzelkriterium C: Transparenz	
Prüffrage C.1	Inwieweit ist die Herkunft von medizinischen Informationen erkennbar?
Prüffrage C.2	Inwieweit wird dargestellt, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben werden?
Prüffrage C.3	Inwieweit wird dargestellt, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind?
Einzelkriterium D: Vollständigkeit	
Prüffrage D.1	Inwiefern wird die Anamnese vollständig dargestellt?
Prüffrage D.2	Inwieweit sind die bei der Beantwortung der sozialmedizinischen Fragestellung zugrunde gelegten Befunde dokumentiert?
Prüffrage D.3	Inwieweit werden die Diagnosen dargestellt und mit Funktionseinschränkungen (sog. Funktionsdiagnosen) formuliert?
Prüffrage D.4	Inwieweit werden die vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen in der Epikrise erfasst?

Fortsetzung Tabelle 22 Überblick Prüffragenkatalog des Manuals zum Peer Review-Verfahren (Löffler et al., 2010, Anhang 12, S.4).

Einzelkriterium D: Vollständigkeit	
Prüffrage D.5	Inwieweit werden die im Gutachten aufgeführten Befunde in der Epikrise erfasst?
Prüffrage D.6	Inwieweit werden die festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit im Erwerbsleben bewertet?
Prüffrage D.7	Inwiefern erfolgt eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen?
Prüffrage D.8	Inwieweit werden wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten dargelegt?
Prüffrage D.9	Inwieweit werden die sozialmedizinischen Fragestellungen vollständig beantwortet?
Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen	
Prüffrage E.1	Inwieweit wird der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet?
Prüffrage E.2	Inwieweit wird die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet?
Einzelkriterium F: Wirtschaftlichkeit	
Prüffrage F.1	Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig?
Prüffrage F.2	Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend?

Datengrundlage. Während der Reviewphase reduzierte sich nochmals die Zahl der am Projekt teilnehmenden Peers. Ein weiterer Peer zog ohne Angabe von Gründen seine Einverständniserklärung zur Teilnahme am Pilotprojekt zurück. Insgesamt gingen die Daten von 19 Peers in die Auswertung ein.

Von den 20 Gutachten, die von allen Peers beurteilt werden sollten¹, waren 349 Reviews zu erwarten, von denen 325 tatsächlich erstellt wurden (Quote: 93 %). Die weiteren 240 Gutachten wurden mit Hilfe des oben beschriebenen Verteilungsschemas (Tab. 21) auf die Peers verteilt. Durch den Ausfall des einen Peers konnten für die weiteren Analysen zwei Peer-Paarungen nicht berücksichtigt werden. Bei effektiv 18 in die Auswertung eingehenden „Peer-Doubletten“ wurde planmäßig von 456 zu erwartenden Reviews ausgegangen. Von diesen wurden bis zum Abschluss der Reviewphase 446 erstellt (Quote: 98 %). Insgesamt ergeben sich daraus 771 erstellte Reviews (Quote: 96 %, bezogen auf 805 erwartete Reviews).

¹ ausgenommen waren hier Gutachten aus dem eigenen RV-Träger

4.2.4 Kalibrierung

Nach Beendigung der Reviewphase fand ein weiterer zweitägiger Workshop mit allen am Verfahren beteiligten Sozialmedizinerinnen statt. Ziel des Workshops war die Darstellung und Diskussion der unterschiedlichen Beurteilungsmaßstäbe der Peers. Sowohl die Peer-Schulung als auch der dem Peer Review angeschlossene Workshop sollten dazu beitragen, die eigenen Bewertungsmaßstäbe bei der Beurteilung von Gutachten zu kalibrieren. Die Kalibrierung dient dazu, den individuellen Peer-Bias zu minimieren, um über alle Beurteiler möglichst homogene Antworttendenzen zu erzeugen. Die Lageabweichungen innerhalb der Bewertungen sollten aufgelöst sowie ggf. Vorschläge zur Optimierung der Formulierungen der Mängelkategorien und Prüffragen entwickelt werden. Der Kalibrierungsworkshop beinhaltete eine allgemeine Diskussion der Reviewphase sowie Kleingruppenarbeit mit dem Ziel, das Schulungsverfahren für die Peers und/oder Elemente bzw. Formulierungen des Manuals hinsichtlich einer höheren Beurteiler-Übereinstimmung zu optimieren. Ergänzend wurde mittels eines Fragebogens die Wichtigkeit der einzelnen Prüffragen und des übergeordneten Kriteriums für das Qualitätssicherungsverfahren erhoben.

4.2.5 Reliabilitätsanalyse

Die Übereinstimmung von zwei oder mehreren Beurteilern wird als Interrater-Reliabilität bezeichnet. Definiert als Zuverlässigkeit oder Genauigkeit der Messung, ist sie im Peer Review das wesentlich zu prüfende Gütekriterium, da im Verfahren die Qualität verschiedener RV-Träger zuverlässig miteinander verglichen werden soll (Neuderth, 2004, S. 128; Wirtz & Caspar, 2002, S. 15).

Laut Manual besteht nicht zwangsläufig ein linearer Zusammenhang zwischen den Einzelkriterien und dem übergeordneten Kriterium. Während Mängel in den Einzelkriterien die Verwertbarkeit des Gutachtens nicht zwangsläufig in Frage stellen, hat das übergeordnete Kriterium die Funktion eines K.O.-Kriteriums. Ausgehend von dieser Modellannahme des hierarchischen Aufbaus des Prüffragenkatalogs, stellt das **übergeordnete Kriterium** der Nachvollziehbarkeit des Gutachtens die **Hauptzielgröße** der Analyse dar.

Die Beurteilung der Interrater-Reliabilität folgt dabei der Argumentation von Wirtz und Caspar, dass „... es nicht eine beste Methode gibt, Reliabilität oder Übereinstimmung zu messen. Einzelne Koeffizienten bilden jeweils eine bestimmte Eigenschaft der erhobenen Datenstruktur ab und spiegeln somit jeweils unterschiedliche Informationen wieder. [...] Entsprechend ist es in der Regel günstig, mehrere Koeffizienten parallel zu berichten, damit ein möglichst vollständiges Bild der Datenstruktur erkennbar wird“ (Wirtz & Caspar, 2002,

S. 23). Der prozentuale Anteil der Reviews, in denen zwei oder mehrere Peers das gleiche Urteil abgeben, ist intuitiv der einfachste Indikator für die Güte eines Verfahrens und wird als prozentuale Übereinstimmung bezeichnet. Diese Form der Übereinstimmung beschreibt zwar einen allgemeinverständlichen Indikator für die Güte des Peer Reviews, sie ist jedoch nur bedingt aussagekräftig. So besteht die Gefahr, dass die Güte des Verfahrens überschätzt wird, da die Indikatoren auch aufgrund einer rein zufälligen Übereinstimmung der Peers in den Bewertungen hohe Werte erreichen können.

In der Forschungsliteratur werden deshalb eine Reihe von komplexen Koeffizienten beschrieben, um Faktoren wie z. B. die zufällige Übereinstimmung zu berücksichtigen, was die Höhe der Güteindikatoren beeinflussen kann. Auf diese Weise besteht die Möglichkeit, eine Überschätzung der Güte zu vermeiden. Die Koeffizienten untergliedern sich weiterhin in *Maße der Beurteiler-Übereinstimmung* und *Maße der Beurteiler-Reliabilität*. Maße der Beurteiler-Übereinstimmung differenzieren nicht zwischen den unterschiedlichen Möglichkeiten der Nicht-Übereinstimmung. Bewertet beispielsweise Peer 1 das übergeordnete Kriterium in der Ampelbewertung mit „grün“ und Peer 2 mit „gelb“, so ist diese Nicht-Übereinstimmung äquivalent der Bewertung von Peer 1 mit „grün“ und Peer 2 mit „rot“. Im Gegensatz dazu berücksichtigen die Maße der Beurteiler-Reliabilität die unterschiedlichen Formen der Nicht-Übereinstimmung. Beurteilt beispielsweise Peer 1 das übergeordnete Kriterium mit „grün“ und Peer 2 mit „rot“, so bestimmt dies im stärkeren Maße die Reliabilität, als die Bewertung von Peer 1 mit „grün“ und Peer 2 mit „gelb“² (vgl. Abb. 12).

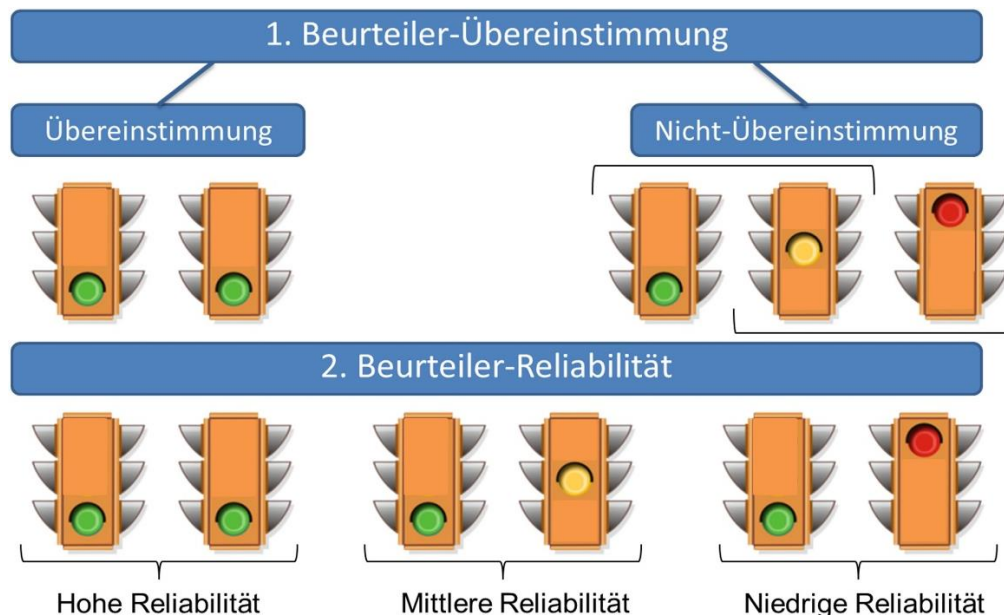


Abbildung 12: Schema zur Beurteiler-Übereinstimmung und Beurteiler-Reliabilität.

² die gewählten Beispiele sind nicht erschöpfend und gelten für alle weiteren Beurteilungskombinationen

Neben der prozentualen Übereinstimmung wurden weitere spezifische Koeffizienten berechnet. Alle Analysen werden mit Excel und SPSS für Windows (Version 24.0, IBM) durchgeführt.

4.2.5.1 Maße der Beurteiler-Übereinstimmung

Cohens Kappa (k) berücksichtigt die prozentuale Übereinstimmung der Peerbewertungen und ist auch gegenüber den zufällig zu erwartenden Übereinstimmungen korrigiert. Es ist deswegen das am häufigsten verwendete Maß der Beurteilerübereinstimmung. Ein Wert um 0 zeigt eine nicht übereinstimmende Bewertung durch die Peers an, ein Wert von 1 die perfekte Übereinstimmung, ein Wert von -1 die völlig gegenläufige Einschätzung.

4.2.5.2 Maße der Beurteiler-Reliabilität auf dem Ordinalskalenniveau

Kendalls Tau-b (r_{tau}) und **Spearman's Rho (r_{rho})** beschreiben die Ausprägung, in dem die Urteile von je zwei Peers kovariieren. Als Korrelationsmaße berücksichtigen sie die unterschiedlichen Formen der Nicht-Übereinstimmung. Ein Wert von 0 zeigt an, dass die Urteile von je zwei Peers nicht kovariieren. Je höher die Urteile kovariieren, desto mehr nähert sich der Koeffizient dem Wert 1 an. Je mehr sich der Koeffizient dem Wert -1 annähert, desto gegenläufiger sind die Peerbewertungen.

Kendalls Konkordanzkoeffizient W (r_w) ist ähnlich Spearman's Rho, kann jedoch für mehr als zwei Peers berechnet werden. Der Koeffizient gibt den Anteil der Varianz zwischen den Ratingrängen der verschiedenen Objekte an der Gesamtvarianz der Ränge an. Ein Wert nahe 1 zeigt an, dass die Peers vergleichbare Ratingränge vergeben, ein Wert von 0, dass dies nicht der Fall ist (vgl. Bortz & Lienert, 2008; Bortz et al., 2008; Gisev, Bell & Chen, 2013).

4.2.5.3 Besonderheiten der Interpretation

Bei der Analyse der Interrater-Reliabilität hat die Signifikanz der berechneten Koeffizienten lediglich eine untergeordnete Bedeutung. Regelhaft wird bei nicht-signifikanten Maßen davon ausgegangen, dass sie das Ergebnis rein zufälliger Beurteilungen seien. Bei der Bestimmung der Interrater-Reliabilität ist die Frage, ob sich das Beurteilerverhalten vom Zufall abhebt, jedoch nicht entscheidend. Von zentraler Bedeutung ist stattdessen die absolute Ausprägung der berechneten Koeffizienten (Wirtz & Caspar, 2002, S. 25).

Eine weitere Besonderheit besteht darin, dass allgemeingültige Kriterien für die Güte von Koeffizienten der Beurteilerübereinstimmung meist nicht angegeben werden können. Häufig existieren lediglich grobe Interpretationsrichtlinien (Wirtz & Caspar, 2002, S. 25), die in der Literatur, je nach Autor und Koeffizient, abweichende Bewertungsempfehlungen beschreiben.

Altman (1991) beschreibt folgende Richtwerte für die Übereinstimmung:

- $<0,20$ schwach,
- $0,21-0,40$ leicht,
- $0,41-0,60$ mittelmäßig,
- $0,61-0,80$ gut,
- $0,81-1,00$ sehr gut.

Dagegen beschreiben Landis und Koch (1977, S. 165) folgende Normwerte:

- $<0,00$ sehr gering,
- $0,00-0,20$ schwach,
- $0,21-0,40$ ausreichend,
- $0,41-0,60$ mittelmäßig
- $0,61-0,80$ gut
- $0,81-1,00$ sehr gut.

Weitere Interpretationsnormen von Döring und Bortz (2016, S. 669) sowie von Cohen (1988, S. 82) gehen davon aus, dass der Zusammenhang im Sinne eines bivariaten Korrelationsmaßes als standardisiertes Effektgrößenmaß wie folgt zu interpretieren ist:

- $0,11-0,30$ klein,
- $0,31-0,50$ mittel,
- $0,51-1,00$ stark.

Vor dem Hintergrund der Literatur, wird Reliabilität zumeist erst ab einer **Höhe von 0,7** als **befriedigend** interpretiert.

Im Gegensatz zu anderen Reliabilitätsmaßen, ist die Interrater-Reliabilität durch eine Vielzahl von Faktoren störanfällig. Beeinflussende Faktoren sind z. B. die Definition des zu messenden Merkmals (ein breit definiertes Merkmal führt in der Regel zu niedrigeren Koeffizienten als klar umschriebene Merkmale), die Anzahl der Skalenkategorien (wenige Kategorien verringern in der Regel die Höhe der Reliabilitäts-koeffizienten), aber auch spezielle Effekte wie z. B. bevorzugte Kategorienverwendung durch einzelne Peers oder Bewertungstendenzen.

Daher soll die Interpretation der Ergebnisse im Kontext inhaltlicher Überlegungen und in Bezug auf die Ergebnisse vergleichbarer Studien erfolgen (Wirtz & Caspar, 2002, S. 25). Goldman (1994) kategorisierte in seiner Metaanalyse die Studienergebnisse an einem Schwellenwert von 0,4. Empirische Vergleichswerte für das Peer Review-Verfahren auf nationaler Ebene liegen vor allem im Rahmen der Qualitätssicherung der Reha-Entlassungsberichte vor. In der Studie von Farin et al. (2003) lag der Reliabilitätskoeffizient Kendalls W für den Bereich „Sozialmedizinischen Stellungnahme“, gemittelt über alle Indikationsbereiche, bei 0,39, der Median bei 0,36. Entsprechend dieser Ergebnisse wird für das Peer Review-Verfahren im Rahmen der Qualitätssicherung der Sozialmedizinischen Begutachtung als kritische Orientierungsmarke eine Reliabilität von 0,4 angenommen.

4.2.6 Qualitative Inhaltsanalyse

Die Untersuchung der Beurteilerkonkordanz wurde durch eine Analyse der freitextlichen Mängeldokumentationen ergänzt. Dazu wurden die freitextlich formulierten Mängelbegründungen inhaltlich analysiert. Die Inhaltsanalyse hatte zum Ziel, die dokumentierten Mängel systematisch zu beschreiben und die Vielzahl der Einzelbegründungen auf charakteristische Kernaussagen zu reduzieren, um typische Qualitätsproblemstellungen zu identifizieren und darzustellen. In der qualitativen Inhaltsanalyse bildet die Verwendung von Kategorien die wesentliche Methode. Dabei können entweder definierte Kategorien an die zu analysierende Texte herangetragen (deduktive Vorgehensweise) oder Kategorien aus den Ausgangsmaterial heraus (induktive Vorgehensweise) entwickelt werden (vgl. Mayring, 2003). Durch die Anwendung beider Methoden handelt es sich bei der vorliegenden Studie um eine Kombination aus strukturierter und zusammenfassender Inhaltsanalyse (Mayring & Fenzl, 2014, S. 547-548).

Die Ableitung prüffragenspezifischer Codes der Einzelkriterien wurde zunächst durch einen am Projekt beteiligten wissenschaftlichen Mitarbeiter vorgenommen. Die Identifikation der Textpassagen, die thematisch und inhaltlich über die definierten Mängelcharakterisierungen hinausreichten, sowie die Ableitung prüffragenübergreifender Kategorien und Generierung der Textcodes erfolgte im Konsensverfahren von zwei Kodierern.

Methodische Vorgehensweise: Einzelkriterien. Dieser beiden Vorgehensvarianten folgend, wurden zunächst die im Manual zum Peer Review hinterlegten vordefinierten Mängeldefinitionen der *Einzelkriterien* als eigenständige Analysekategorien ausgewertet. Auf diese Weise ergaben sich Kategorien, die spezifisch für jede Prüffrage sind. Anschließend wurde an den 325 Reviews der 20 Gutachten, die von allen Peers beurteilt wurden, die Textpassagen identifiziert, die thematisch und inhaltlich über die definierten Mängelcharakterisierungen hinausreichten. Diese Textpassagen wurden um die Themen und Inhalte reduziert, die als bedeutungsgleich mit den definierten Mängelbeschreibungen angesehen wurden. Die verbleibenden Textpassagen mit thematischem und inhaltlichem Mehrwert wurden geordnet, bedeutungsverwandte Textpassagen gebündelt und mit einer zusammenfassenden Kurzcharakterisierung beschrieben. Die Kurzcharakterisierungen wurden prüffragenübergreifend miteinander in Verbindung gesetzt, die Typisierungen weiter verdichtet und auf einem höheren Abstraktionsniveau im Sinne eines *Codes* verwendet. Das Vorgehen ist in Abbildung 13 vereinfacht schematisch dargestellt.

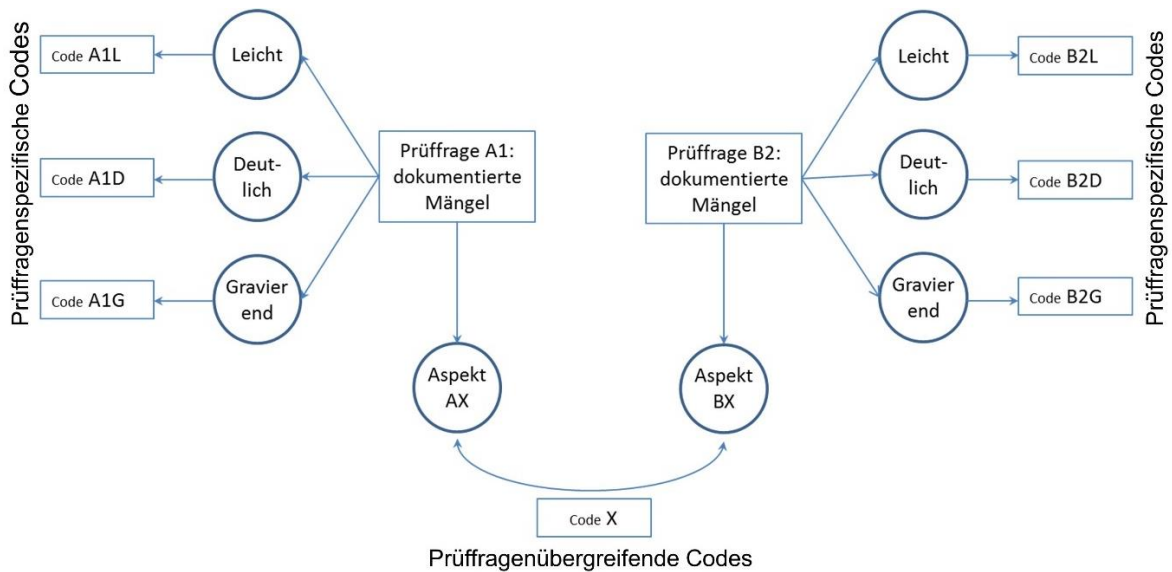


Abbildung 13: Schema des Vorgehens bei der qualitativen Inhaltsanalyse.

Prüffragenspezifische Codes wurden mit dem Kennzeichnungsbuchstaben der sechs Einzelkriterien des Manuals (A bis E), der Nummer der Prüffrage (1 bis 9) und der Stufe des Mängelbewertungsschemas (L für leichte Mängel, D für deutliche Mängel und G für gravierende Mängel) verschlüsselt. Wurden innerhalb einer Mängelstufe mehrere thematische Aspekte identifiziert, wurde diese in der 4. Stelle fortlaufend nummeriert. Prüffragenübergreifende Codes hingegen wurden mit einem X und einer fortlaufenden Nummer bezeichnet. Alle verwendeten Codes wurden in einer Codeliste beschrieben.

Die Codes wurden anschließend auf die freitextlich begründeten Qualitätsdefizite in den Reviews angewendet. Dabei wurde fortlaufend auf die Passung der vorliegenden Codes zu den zu analysierenden Textpassagen geachtet und diese bei Bedarf modifiziert bzw. weiter ausdifferenziert. Auf diese Weise war es möglich, alle thematischen und inhaltlichen Aspekte der freitextlichen Mängelbegründungen der Einzelkriterien auf einem höheren Abstraktionsniveau zuzuordnen.

Methodische Vorgehensweise: übergeordnetes Kriterium. Für die Analyse des *übergeordneten Kriteriums* wurden ausschließlich die Reviews der Gutachten herangezogen, die von allen Peers beurteilt wurden. Die Ampelbewertung des Kriteriums ist nicht durch streng vordefinierte Mangelkategorien gekennzeichnet. Gemäß der induktiven Vorgehensweise der qualitativen Inhaltsanalyse wurden die Codes daher aus den Freitexten der Peers heraus entwickelt.

Auswertungsstrategie. Im Anschluss der Codierung erfolgte für jede Prüffrage eine Auszählung der in den Begründungen für Mängelbewertungen verwendeten Codes. Mit dieser Methode wurde untersucht, welche und wie häufig Begründungen, die sich aus den hinterlegten Mängelkriterien ergeben, eingesetzt wurden. Zudem wurden Aspekte, die über die im Manual hinterlegten Kriterien hinausgehen, neben der inhaltlichen Einbindung auch quantitativ erfasst.

In einem weiteren Schritt wurden am Beispiel des übergeordneten Kriteriums die Codes mit den quantitativen Mängelbewertungen der Peers verknüpft. Anhand dieses sog. K.O.-Kriteriums als Hauptzielkriterium des Verfahrens wurde versucht mit dieser Gegenüberstellung Divergenzen und Fehler bei der Anwendung des Prüffragenkataloges aufzudecken.

Datengrundlage. Insgesamt wurden 4.970 freitextliche Mängelbegründungen bei Beurteilung der Einzelkriterien und 426 Mängelbegründungen beim übergeordneten Kriterium dokumentiert. Tabelle 23 zeigt die Anzahl freitextlicher Mängelbegründungen je Prüffrage.

Tabelle 23: Anzahl der freitextlichen Mängelbegründungen der Einzelkriterien.

Einzelkriterium	Prüffrage	Gutachten alle Peers	Häufigkeit Paare*	Gesamt Σ
Formale Gestaltung	A1	73	120	193
	A2	85	74	159
Verständlichkeit	B1	63	100	163
	B2	114	141	255
	B3	64	72	136
	B4	39	58	97
Transparenz	C1	110	158	268
	C2	78	74	152
	C3	121	117	238
Vollständigkeit	D1	136	181	317
	D2	82	123	205
	D3	190	259	449
	D4	127	148	275
	D5	114	137	251
	D6	104	129	233
	D7	184	237	421
	D8	99	113	212
	D9	162	199	361
Medizinisch-wissenschaftliche Grundlagen	E1	19	20	39
	E2	79	103	182
Wirtschaftlichkeit	F1	62	70	132
	F2	102	130	232
Übergeordnetes Kriterium	Ampel	191	235	426

* Anzahl der freitextlichen Mängelbegründungen, die beim Review durch die Peer-Paare dokumentiert wurden

4.3 Ergebnisse

4.3.1 Übereinstimmung der Probegutachten

Im Anschluss an die Schulung wurden von den am Projekt teilnehmenden Peers drei anonymisierte Probegutachten auf Grundlage des Manuals beurteilt und die jeweiligen Reviews mit Hilfe des Erfassungssystems EvaSys® dokumentiert. Insgesamt wurden 57 Reviews erstellt. Ein Peer konnte aus gesundheitlichen Gründen keine Probegutachten bearbeiten. Der digitale Zugriff war lediglich auf 56 der insgesamt 57 Reviews möglich. Ein nachweislich durchgeführtes Review ist bei der digitalen Übermittlung durch EvaSys® verloren gegangen. Für dieses Review lagen infolgedessen keine Daten vor.

Obgleich die Probegutachten ausschließlich zur Überprüfung der Schulung eingesetzt wurden, wurde explorativ die prozentuale Übereinstimmung der Peers global für alle Einzelkriterien und das übergeordnete Kriterium berechnet (Tab. 24).

Tabelle 24: Mittelwerte der Beurteiler-Konkordanz (325 Reviews).

Prüffrage	Prozentuale Übereinstimmung (%)		
	Probe- gutachten 1	Probe- gutachten 2	Probe- gutachten 3
Einzelkriterien	47,6 %	68,4 %	48,9 %
Übergeordnetes Kriterium	35,5 %	52,9 %	33,3 %

Die prozentuale Übereinstimmung der Einzelkriterien lag durchschnittlich bei 54,9 % mit einer Range von 47,6 % bis 68,4 %. Die Übereinstimmung der Peers bezüglich des übergeordneten Kriteriums lag zwischen 33,3 % und 52,9 %, im Mittel bei 40,6 %.

Nach Bearbeitung der Probegutachten wurde jeder Peer telefonisch kontaktiert und auf Basis eines Interviewleitfadens zu den Erfahrungen bei der Bearbeitung befragt. Die Probegutachten wurden überwiegend ohne Verständnisprobleme bearbeitet. Probleme wurden lediglich hinsichtlich der technischen Umsetzung berichtet, die durch die IT-Abteilungen des jeweiligen RV-Träger zeitnah gelöst werden konnten. Für das Erhebungssystem EvaSys® wurde die Möglichkeit vorgeschlagen, fertig gereviewte Prüfbögen auch vom Peer selbst speichern und ausdrucken lassen zu können. Eine andere Anmerkung bezog sich auf die Schriftgröße der Prüffragen, die nicht variabel konfigurierbar war. Schließlich wurde angemerkt, dass die TAN-Nummern zur leichteren Orientierung zusätzlich im Kopf des Prüfbogens im EvaSys® ersichtlich sein sollten. In Bezug auf die Anonymisierung der Gutachten wurde zudem beanstandet, dass das Alter bzw. zumindest das Geburtsjahr des Versicherten erkennbar bleiben sollte.

4.3.2 Interrater-Reliabilität der Einzelkriterien

Die mit SPSS und Excel berechneten Interrater-Übereinstimmungen und Reliabilitäten der Einzelkriterien, basierend auf 325 Reviews, waren je nach Koeffizienten sehr heterogen (Tab. 25).

Tabelle 25: Mittelwerte der Beurteiler-Konkordanz der Einzelkriterien (325 Reviews).

Prüffrage	Prozentuale Übereinstimmung	Cohens Kappa	Kendalls Tau-b	Spearman's Rho	Kendalls W
Einzelkriterium A: Formale Gestaltung					
Prüffrage A.1	68,6 %	0,12	0,12	0,13	0,18
Prüffrage A.2	83,1 %	0,76	0,81	0,83	0,87
Einzelkriterium B: Verständlichkeit					
Prüffrage B.1	70,7 %	0,25	0,2	0,21	0,26
Prüffrage B.2	50,6 %	0,11	0,28	0,3	0,34
Prüffrage B.3	68,6 %	-0,03	0,09	0,1	0,15
Prüffrage B.4	80,5 %	0,28	0,23	0,23	0,29
Einzelkriterium C: Transparenz					
Prüffrage C.1	54,4 %	0,13	0,20	0,21	0,26
Prüffrage C.2	66,8 %	0,20	0,30	0,31	0,36
Prüffrage C.3	58,3 %	0,14	0,23	0,24	0,29
Einzelkriterium D: Vollständigkeit					
Prüffrage D.1	49,3 %	0,16	0,19	0,21	0,25
Prüffrage D.2	65,6 %	0,22	0,23	0,25	0,29
Prüffrage D.3	49,3 %	0,25	0,36	0,39	0,42
Prüffrage D.4	46,0 %	0,07	0,08	0,09	0,14
Prüffrage D.5	52,9 %	0,14	0,15	0,16	0,21
Prüffrage D.6	53,1 %	0,06	0,03	0,03	0,09
Prüffrage D.7	40,0 %	0,15	0,27	0,30	0,33
Prüffrage D.8	80,2 %	0,27	0,19	0,19	0,25
Prüffrage D.9	43,8 %	0,09	0,10	0,11	0,16
Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen					
Prüffrage E.1	89,0 %	0,64	0,2	0,21	0,26
Prüffrage E.2	66,3 %	0,29	0,18	0,18	0,24
Einzelkriterium F Wirtschaftlichkeit					
Prüffrage F.1	72,9 %	0,04	0,14	0,14	0,2
Prüffrage F.2	57,5 %	0,16	0,15	0,16	0,21

Die Range der Reliabilität lag für Cohens Kappa zwischen -0,03 und 0,76, für Kendalls Tau-b zwischen 0,09 und 0,81, für Spearmans Rho zwischen 0,1 und 0,83 und für Kendalls Konkordanzkoeffizient W zwischen 0,16 und 0,89. Der Mittelwert von Kendalls Konkordanzkoeffizient W als validester Reliabilitätskoeffizient für mehrere gleichzeitig beurteilende Rater betrug 0,28 (Median: 0,26) für die Interrater-Reliabilität der Einzelkriterien. Prüffrage A2 erreicht in allen Reliabilitätskoeffizienten einen Wert von >0.7 . Demgegenüber stehen die Prüffragen B3, D4, D6 und D9 die auffallend niedrige Reliabilitätswerte erreichten.

Im Gegensatz zur niedrigen Reliabilität steht eine prozentuale Übereinstimmung in Höhe von im Mittel 62,2 % (Median: 62,0 %), mit einer Range von 40 % bis 89 %. Vor allem bei Prüffragen mit niedriger Interrater-Reliabilität war eine Inkongruenz zwischen der Reliabilität und der prozentualen Übereinstimmung zu verzeichnen. Während beispielsweise bei Prüffrage B3 der Kendalls Konkordanzkoeffizient W lediglich eine Höhe von $r_w = 0,15$ erreichte, betrug die prozentuale Übereinstimmung dieser Prüffrage 68,6 %.

4.3.3 Interrater-Reliabilität des übergeordneten Kriteriums

Die Interrater-Reliabilität des übergeordneten Kriteriums, basierend auf 325 Reviews, schwankte je nach herangezogenen Reliabilitätskoeffizienten zwischen 0,2 und 0,39. Die prozentuale Übereinstimmung lag bei 47,9 % (Median: 49,3 %). In Tabelle 26 sind alle berechneten Koeffizienten abgebildet.

Tabelle 26: Mittelwerte der Beurteiler-Konkordanz des übergeordneten Kriteriums (325 Reviews).

Prüffrage	Prozentuale Übereinstimmung	Cohens Kappa	Kendalls Tau-b	Spearmans Rho	Kendalls W
Übergeordnetes Kriterium					
Ampelbewertung	47,9 %	0,2	0,31	0,33	0,39

Die relativ niedrigen Ausprägungen der Koeffizienten der Ampelbewertung konnten durch die zusätzliche Einzelanalyse der beurteilten Gutachten veranschaulicht werden. In Abbildung 14 sind die prozentualen Anteile der drei Bewertungskategorien für 20 Gutachten dargestellt, die von allen Peers bewertet worden sind. Die Qualität des Gutachtens in der Abbildung links oben wurde von 88 % der Peers „grün“, von 12 % „gelb“ und von keinem Peer „rot“ beurteilt. Dagegen wurde die Nachvollziehbarkeit des Gutachtens rechts unten von 77 % „rot“, von keinem „gelb“ und von 23 % „grün“ beurteilt. Bei vielen dazwischen liegenden Gutachten fielen die Bewertungen heterogener aus.

4.3 Interrater-Reliabilität des Peer Review Verfahrens: Ergebnisse

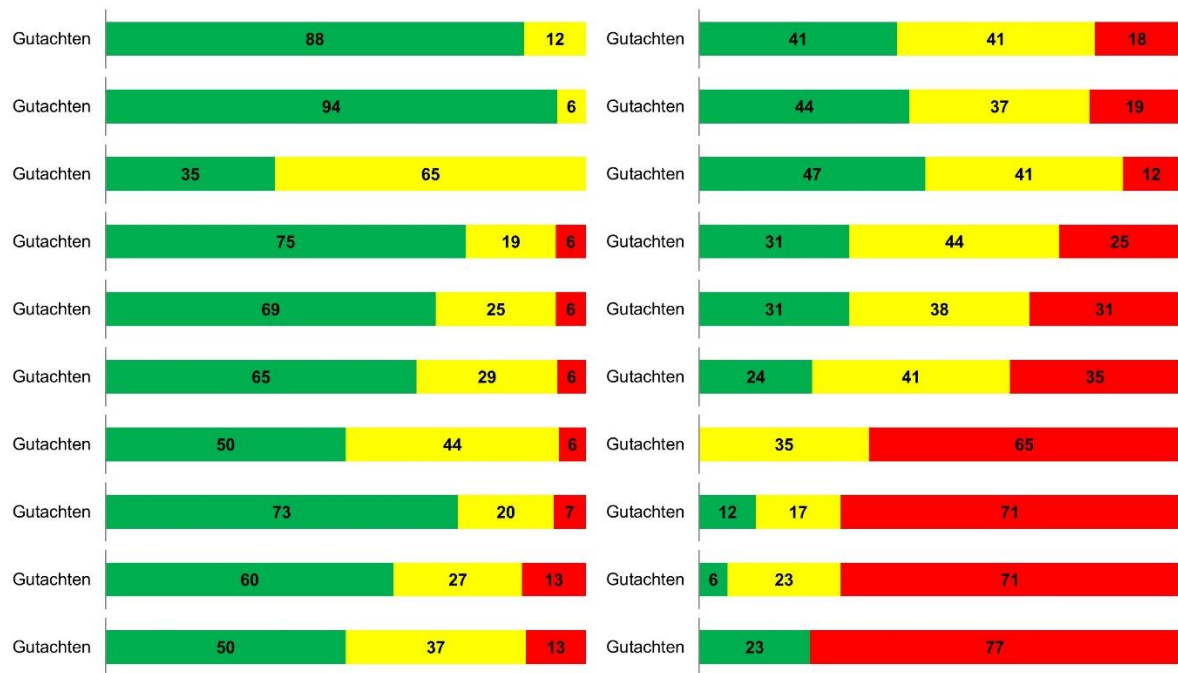


Abbildung 14: Varianz der Ampelbewertungen (übergeordnetes Kriterium) für die 20 von allen Peers beurteilten Gutachten (in Prozent; modifiziert nach Strahl et al., 2016, S.158).

Daraus ableitend lässt sich unter der gutachtenspezifischen Betrachtung der prozentualen Übereinstimmung der Beurteilungen des übergeordneten Kriteriums ebenfalls eine auffällige Varianz der Peer-Urteile beschreiben. Die Range der Peer-Übereinstimmung schwankte zwischen 87,5 % und 29,2 % (Tab. 27). Demnach streut unabhängig von der Verteilung der Ampelbewertung (Abb. 14) auch die prozentuale Übereinstimmung der Peers bei der Beurteilung des übergeordneten Kriteriums stark.

Tabelle 27: Gutachtenspezifische prozentuale Übereinstimmung des übergeordneten Kriteriums (325 Reviews).

Gutachtennummer	Prozentuale Übereinstimmung	Gutachtennummer	Prozentuale Übereinstimmung
#1	87,5 %	#11	47,8 %
#2	76,7 %	#12	41,8 %
#3	61,5 %	#13	41,0 %
#4	57,5 %	#14	36,8 %
#5	55,2 %	#15	36,7 %
#6	52,9 %	#16	33,1 %
#7	51,5 %	#17	32,5 %
#8	51,5 %	#18	30,9 %
#9	51,5 %	#19	30,8 %
#10	50,8 %	#20	29,2 %

4.3.4 Beurteilungsunterschiede zwischen den Peers

Da alle Peers dieselben Gutachten beurteilt haben besteht die Möglichkeit die interindividuelle Beurteilungsunterschiede deskriptiv darzustellen. Die beschriebene Heterogenität der Reliabilitätskoeffizienten zeigt sich auch zwischen den Peers, wie Abbildung 15 veranschaulicht. In Bezug auf das übergeordnete Kriterium beschreibt die Analyse die interindividuellen Beurteilungsmaßstäbe und damit die Beurteilerstrenge der teilnehmenden Peers. Der „mildeste“ Peer (links oben) vergab bei 83 % aller beurteilten Gutachten „grün“, 17 % „gelb“ und kein Mal „rot“, während hingegen der „strengste“ Peer (rechts unten) 44 % seiner Gutachten mit „rot“, 33 % „gelb“ und 22 % „grün“ beurteilte.



Abbildung 15: Varianz der Peerbeurteilungen des übergeordneten Kriteriums (in Prozent)³.

Daraus ableitend lässt sich unter der gutachtenspezifischen Betrachtung der prozentualen Übereinstimmung der Beurteilungen des übergeordneten Kriteriums ebenfalls eine auffällige Varianz der Beurteilungen beschreiben.

4.3.5 Qualitative Inhaltsanalyse der Einzelkriterien

Das herausgearbeitete Codierungssystem zur Systematisierung der freitextlichen Mängeldokumentation beinhaltet sowohl prüffragenspezifische als auch -übergreifende Codes. Unter der Annahme, dass vorhandene Mängel entlang der hinterlegten Mängelkategorien einer Prüffrage bewertet und dementsprechend dokumentiert werden, wurden prüffragenspezifische Codes definiert, die inhaltlich den Mängelstufen der Prüffragen zuordenbar sind. Zum anderen wurden aus dem Ausgangsmaterial Codes identifiziert, die über die prüffragenspezifischen Begründungskategorien hinausgehen.

³ von den 19 Peers hat ein Peer lediglich ein Review erstellt. Dieses wird an dieser Stelle nicht berichtet.

Im Anhang 6 werden für alle Prüffragen der Einzelkriterien die 77 prüffragenspezifischen Codes tabellarisch aufgelistet und inhaltlich erläutert. Die Codes bilden im Wesentlichen die in den Bewertungsmatrizen hinterlegten Beurteilungskriterien des Prüffragenkatalogs ab. Zusätzlich wurden während der Analyse innerhalb der Einzelkriterien 20 weitere prüffragenübergreifende Codes identifiziert (Anhang 7).

Prüffrage A1. Die Prüffrage A1 erfordert die adäquate Beurteilung des Anforderungsprofils für sozialmedizinische Gutachten gemäß DRV-Schrift Band 21, ohne Einbezug der inhaltlichen Vollständigkeit oder Nachvollziehbarkeit der Darstellung. Die Begründungen der festgestellten Mängel ergaben sich überwiegend aus den im Manual hinterlegten Qualitätsaspekten. Teilweise wurden die Einschränkungen des Manuals nicht beachtet und inhaltliche Komponenten beurteilt (n = 10). Darüber hinaus waren bei Prüffrage A1 auch zu differenzierte Begründungen zu erkennen (n = 10), die sich z. B. auf spezielle Unterpunkte in der Anamnese richteten, die laut Manual nicht zu beurteilen waren. Dagegen wurden indirekte Begründungen (n = 7), die sich nicht auf die Mängelkategorien des Manuals beziehen und zwischen den Kategorien nicht trennscharf differenzieren oder pauschale Begründungen (n = 1) nur vereinzelt verwendet (Tab. 28).

Tabelle 28: Codierung der Prüffrage A1.

Prüffrage A1 bezieht sich auf das rein formale Vorhandensein der Gliederungsvorgabe der Überschriften für sozialmedizinische Gutachten.		
Code	n	Ankerbeispiel
A1L	20	„Reihenfolge der Gliederung stimmt nicht, dazwischen plötzlich Befundmitteilung“ (128 ⁴ / zm ⁵)
A1D	81	„Der Gliederungspunkt ‚Sozialmedizinische Leistungsbeurteilung‘ wird nicht per Überschrift klar abgegrenzt“ (41 / zi)
A1G	28	„Die Überschriften Anamnese und Untersuchungsbefunde fehlen“ (267 / zp)
X3	1	„Im Text kommt die sozialmed. LB vor dem Gliederungspunkt“ (191 / zg)
X4	10	„Keine Anamnese im engeren Sinne, lediglich ‚angedeutete Anamnese‘ über/unter Berufung auf Fremdbefunde“ (145 / zx)
X6	2	„Vorbefunde in der Vorgeschichte, ohne eigentliche Krankheitsgeschichte“ (71 / zo)
X7	12	[Keine Textangabe]
X10	10	„Aktuelle Anamnese wurde unter spezieller Anamnese aufgeführt, nicht als eigener Punkt“ (177 / zm)
X11	2	„Aufbau insgesamt sehr diffus und nur schwer nachvollziehbar, da keine Grundlinie“ (18 / zo)
X13	1	„Z. B. letzte T.“ (172 / zf)
X14	7	„Epikrise und Leistungsbeurteilung werden in einem Kapitel zusammengefasst“ (155 / zq)
X15	7	„Keine DRV konforme Gliederung“ (184 / zo)
X19	1	„Epikrise partiell vorhanden, da 2 Seiten des GA nicht zugesandt wurden“ (222 / zc)
X20	11	„Rein formal m.E. keine Mängel“ (51 / zb)

⁴ Interne Gutachten-ID

⁵ anonymisierter Peer-Code

Prüffrage A2. Die Prüffrage A2 erfordert ausschließlich die Überprüfung der formalen Verwendung des einheitlichen Deck- und Schlussblattes. Die Begründungen der festgestellten Mängel ergaben sich überwiegend aus den im Manual hinterlegten Qualitätsaspekten. Der Fall, dass lediglich das einheitliche Deckblatt der DRV fehlte, kam in keinem der beurteilten Gutachten vor. Die Mängelkategorien erscheinen trennscharf, da wenige prüffragenübergreifende Codes vergeben wurden ($n = 15$). Auffällig bei Prüffrage A2 ist, dass 31 Peers einen der im Manual definierten Mängel codierten, die festgestellten Qualitätsdefizite jedoch nicht schriftlich dokumentierten (Tab. 29).

Tabelle 29: Codierung der Prüffrage A2.

Prüffrage A2 bezieht sich auf die Verwendung und das Vorhandensein des einheitlichen Deck- und Schlussblattes.		
Code	n	Ankerbeispiel
A2L1	9	„Das Deckblatt ist beigelegt. Es ist aber gegenüber dem einheitlichen Formularsatz abgewandelt.“ (128 / zp)
A2L2	29	„Das Schlussblatt ist leicht abgewandelt, es fehlt die Möglichkeit, eine konkrete Befristung einzutragen“ (51 / zq)
A2D1	0	---
A2D2	10	„Schlussblatt fehlt“ (282 / zf)
A2G	72	„Es fehlen Deck- und Schlussblatt“ (76 / zm)
X4	1	„Nicht komplett ausgefüllt, vor allem keine Krankheitszeiten, auch nicht aus dem GA ableitbar“ (53 / zo)
X7	31	[Keine Textangabe]
X9	2	„Es fehlt ein abgrenzbarer Abschnitt, Epikrise“ (128 / zi)
X11	5	„Leichte Formabweichungen“ (63 / ze)
X13	1	„s. Text oben“ (76 / zp)
X14	2	„Formal kein Mangel. Das Schlussblatt ist jedoch modifiziert“ (129 / zq)
X20	4	„Es wird darauf aufmerksam gemacht, daß die „letzten 2 Blätter“ nicht bearbeitet wurden (keine Kreuze ‚gesetzt‘ etc) → einheitlicher Formularsatz ansonsten vorhanden“ (41 / zc)

Prüffrage B1. Die Prüffrage B1 erfordert die adäquate Beurteilung der Verständlichkeit in Form der Unmissverständlichkeit der sprachlichen Ausdrucksweise ohne Berücksichtigung des persönlichen Schreibstils des Gutachters. Die Begründungen der festgestellten Mängel ergaben sich überwiegend aus den im Manual hinterlegten Qualitätsaspekten. Die qualitative Inhaltsanalyse zeigte jedoch auch, dass bei dieser Prüffrage die prüffragenspezifische Kategorie „gravierende Mängel“ von den Peers nie verwendet wurde. Gutachtenbeispiele ($n = 33$) geben Einblick in die sprachliche Ausdrucksweise des Gutachters, sind jedoch nicht in der Lage die Mängelkategorie ausreichend zu differenzieren. Ferner wurde mehrfach irrtümlich die Nutzung von Fremdwörtern und Abkürzungen als Fehler angemerkt ($n = 10$), obgleich dies erst Prüfungsgegenstand der Prüffrage B2 ist. Auch wurden in den Mängelbegründungen z.T. inhaltliche Widersprüche innerhalb des Gutachtens dokumentiert ($n = 7$), was laut Manual erst in der Ampelbewertung Gegenstand der Beurteilung sein soll (Tab. 30).

Tabelle 30: Codierung der Prüffrage B1.

Prüffrage B1 überprüft die Sprache, Grammatik und Satzkonstruktion im gesamten Gutachten, um unnötige Längen und komplizierte Formulierungen zu vermeiden.		
Code	n	Ankerbeispiel
B1L1	66	„Unnötige Längen mit Wiederholungen, insb. In der Epikrise“ (208 / zc)
B1L2	8	„Stichwortartige Diktion. Grammatisch problematische Formulierungen in der Epikrise (‚Er dennoch gemäß Dr.____ vollschichtig arbeiten kann‘, ‚Der psychische Befund bietet ...einen Versicherten...‘)“ (33 / zi)
B1D	12	„[...] missverständliche/nicht eindeutige Äußerungen (unter Epikrise wie z. B. ‚Oberlappenresektion der linken Lunge wegen eines Tumors‘)“ (48 / zx)
B1G	0	---
X3	1	„Erhebliche Redundanzen und schwer verständliche Aussagen“ (37 / ze)
X6	2	„Die Anamnese erscheint unübersichtlich“ (32 / zu)
X7	5	[Keine Textangabe]
X8	1	„Siehe A1, schwer zu lesen“ (18 / zo)
X9	10	„Viele Fremdwörter, viele nicht erklärte Abkürzungen“ (100 / zo)
X11	2	„Etwas durcheinander, keine Reihenfolge“ (157 / zo)
X12	7	„Inhaltlich liegen Widersprüche vor, bei der Diagnose depressive Anpassungsstörung, in der Epikrise dann schwere depressive Episode (Anpassungsstörung nicht mit depressiver Episode gleichzusetzen), Parkinson-Syndrom in der Diagnose, in der Epikrise plötzlich Verdacht auf Tremor rechtes Bein, ohne dass dies i neurologischen Befund beschrieben wird“ (201 / zm)
X13	3	„Einige ‚Ungenauigkeiten‘ wie zum Beispiel unter ‚jetzige Beschwerden‘: Kopfschmerzen, Schwindel, Kreislaufstörungen usw.“ (34 / zx)
X14	2	„Blatt 7, letzter Absatz: „Rotationseinschränkungen de HWS – motorische Defizite liegen nicht vor“. Ich verstehe was gemeint ist, sprachlich leicht missverständlich, insgesamt aus meiner Sicht kein Mangel“ (242 / zq)
X15	1	„In sich verschachtelte Sätze mit schwer/kaum sich erschließenden Aussagen, nicht verständliche Darstellung der Soz.med. Beurteilung. Überflüssige Wiederholungen in der Epikrise.“ (74 / zx)
X16	33	„Schwammige Ausdrucksweise ‚wohl chronische Bronchitis‘, ‚Bier‘ (Menge?)“ (88 / ze)
X18	1	„Sehr allgemeine Darstellungen“ [...] (92 / zx)
X20	9	„Ausführlich (gut), nicht redundant“ (132 / zg)

Prüffrage B2. Die Prüffrage B2 erfordert die adäquate Beurteilung der allgemeinverständlichen Erklärung von Fachausdrücken und Abkürzungen in der Diagnose und in der Epikrise. Die Begründungen der festgestellten Mängel ergaben sich überwiegend aus den im Manual hinterlegten Qualitätsaspekten und bezogen sich unmittelbar auf die festgelegte Bewertungscharakteristik mit mengenmäßiger Steigerung (keine, vereinzelt, mehrfach, überwiegend). Vielfach verwendeten die Peers konkrete Beispiele (n = 147), um die gewählten Mängelkategorien zu begründen. Problematisch an diesem Vorgehen ist, dass die der Prüffrage hinterlegte Bewertungscharakteristik nicht trennscharf wiedergegeben werden kann (Tab. 31). In der Folge beurteilten die Peers trotz der Dokumentation derselben Mängel unter Umständen unterschiedliche Mängelkategorien (Tab. 32).

Tabelle 31: Codierung der Prüffrage B2.

Prüffrage B2 überprüft, inwieweit Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt wurden.		
Code	n	Ankerbeispiel
B2L1	0	---
B2L2	4	„Erläuterung eines nicht allgemein verständlichen Fachbegriffs: OP nach Janetta?“ (141 / zp)
B2D1	21	„Wiederkehrende Abkürzungen ohne Erklärung“ (41 / zo)
B2D2	45	„Keine Erläuterung der Diagnosen wie Hypochondrie, Wurzelreizsyndrom L5 warum? Bandscheibe oder Forameneinengung?“ (48 / zs)
B2G1	0	---
B2G2	11	„Es erfolgt überwiegend keine Erläuterung nicht allgemein verständlicher Fachbegriffe“ (128 / zg)
X6	6	„Sehr vereinfachende Übersetzung der med. Diagnosen unter 3.1 die Leiden werden nur ansatzweise erläutert“ (264 / zx)
X7	5	[Keine Textangabe]
X8	9	„Siehe B1“ (92 / zd)
X9	2	„Unter 3. Diagnosen: die erste und zweite Syndrombezeichnung sind keine Diagnosen_ die übrigen Diagnosen ohne Funktionsdiagnosen“ (59 / zx)
X11	3	„Es wird med. Fachsprache verwendet“ (166 / zt)
X13	1	„B2“ (97 / zo)
X15	2	„Allgemeine Verständlichkeit in den Diagnosen nicht gegeben“ (64 / zm)
X16	147	„Epikrise: SKID, DSM, Minussymptomatik“ (199 / ze)
X20	10	„Wahrscheinlich sind AZ, EZ, BMI, BWS allgemein bekannt“ (95 / zt)

Tabelle 32: Beispiel für gleiche Mängelbegründung bei unterschiedlicher Peer-Beurteilung.

Gutachten	Peer	Mängelkategorie	Freitextliche Mängelbegründung
#1	ze	3	„Tremor, Rigor“
#1	zi	2	„Tremor“, „Rigor“?

Ferner wurden je nach subjektiver Wahrnehmung des beurteilenden Peers unterschiedliche Fachausdrücke oder Abkürzungen im gleichen Gutachten angemerkt, was zu Verzerrungen in der Beurteilung der Prüffragen führen kann (Tab. 33).

Tabelle 33: Beispiel für gleiche Mängelbegründung bei unterschiedlicher Peer-Beurteilung.

Gutachten	Peer	Mängelkategorie	Freitextliche Mängelbegründung
#2	zt	3	„Lasègue, Versuch, Tinel'sche Zeichen im Karpaltunnel, NAP“
#2	zi	2	„Hypochondrisch“? „Wurzelläsion“?

Prüffrage B3. Die Prüffrage B3 erfordert die adäquate Beurteilung der richtigen Anwendung sozialmedizinischer Terminologie gemäß den Definitionen des Sozialmedizinischen Glossars (DRV-Schrift Band 81). Die Begründungen der festgestellten Mängel ergaben sich nur teilweise aus den im Manual hinterlegten Qualitätsaspekten. Auch bei der Prüffrage B3 verwenden die Peers stattdessen vielfach konkrete Beispiele (n = 75), um die gewählten Mängelkategorien zu begründen („Kur“, „vollschiebig“, „ortsüblicher Anmarschweg“, „Wegefähigkeit“ etc.). Zusätzlich wurde vier Mal ein neuer Gesichtspunkt bei den Mängelbegründungen kodiert, der über die Ausprägungen der ursprünglichen

Mängelkategorien hinausgeht. Dabei geht es jeweils um die inhaltliche Komponente, wenn der Gutachtentext und die Kreuze auf den Formblättern nicht übereinstimmen (Tab. 34).

Tabelle 34: Codierung der Prüffrage B3.

Prüffrage B3 untersucht, inwieweit die sozialmedizinische Terminologie, gemäß der Definition im Sozialmedizinischen Glossar der Deutschen Rentenversicherung, richtig angewendet wird.		
Code	n	Ankerbeispiel
B3L	5	„Epikrise: ‚Erwerbsminderungsrente‘ sollte ‚anerkannt werden‘. Unbenommen der weiter unten abzuhandelnden Frage, ob präjudizierende Aussagen gemacht werden, meine ich, dass hier auch ein sozialmedizinischer Terminus nicht korrekt angewandt wird - besser wäre es gewesen, von einer Einschränkung der quantitativen (unter dreistündigen) Leistungsfähigkeit zu reden.“ (108 / zq)
B3D	23	„Dreimal ‚zeitweise‘ geht nicht (Arbeitshaltungen)“ (41 / e)
B3G	1	„akute AU..., 3 bis unter 6 stündigen Zeitumfang bei Rücksicht auf die Restgesundheit [...] ein 6- und mehrstündiges Leistungsvermögen auf dem allgemeinen Arbeitsmarkt“ unter bestimmten Umständen????? Auf Schlussblatt Teil 1 dann ‚mehr als 6h‘ auf dem allgemeinen Arbeitsmarkt.“ (74 / zx)
X1	4	„Leistungen zur Teilhabe am Arbeitsleben (LTA) werden empfohlen -> siehe B4, aber unter Punkt 7. mit ‚NEIN‘ angekreuzt!“ (166 / zc)
X6	4	„Arbeitsfähigkeit interessiert nur am Rande. Wichtig wäre Angabe zur Erwerbsminderung.“ (78 / zz)
X7	7	[Keine Textangabe]
X9	2	„In der sozialmedizinischen Beurteilung fehlen Angaben zum quantitativen Leistungsbild, es werden lediglich (wenige) Aussagen zu den qualitativen Einschränkungen getroffen. Im Schlussblatt sind sodann die quantitativen Leistungskriterien angekreuzt.“ (48 / zq)
X12	3	„Die Ausführung im Leistungsbild, dass ‚eine Reha-Maßnahme mit Belastungserprobung ab September möglich und sinnvoll wäre“ ist interpretationsbedürftig, zumal im Schlussblatt eine medizinische Reha befürwortet wird, Leistungen zur Teilhabe am Arbeitsleben jedoch nicht. Die abschließende Empfehlung der erneuten Leistungsbeurteilung in ca. 2 bis 3 Jahren ist ebenfalls erklärungsbedürftig, zumal im Schlussblatt unter Punkt 5 kein voraussichtliches Ende der Leistungsminderung angegeben wird. (186 / zw)
X13	1	„Leistungsbeurteilung nur auf Umwegen verständlich (S. 10, 3. Zeile)“ (37 / ze)
X14	2	„Wegefähigkeit statt Gehstrecke, es geht auch nicht um Altersruhegeld, es geht auch nicht um Arbeitsfähigkeit, sondern um Leistungsminderung, Berufs- und Erwerbsunfähigkeit sind Ausdrücke aus dem alten Rentenrecht aus 2000, zudem juristische Ausdrücke“ (128 / zm)
X15	1	„Arbeitshaltung nicht vollständig, Minijob als relevante letzte Tätigkeit“ (117 / zs)
X16	75	„vollschichtig“ (115 / ze)
X18	2	„Der Rentenversicherungsträger kann nicht einen Rentenantrag in einen Reha-Antrag ‚umwandeln‘. Die sozialmedizinischen Schlussfolgerungen sind aber unmissverständlich dahingehend, dass eine erhebliche Gefährdung der Erwerbsfähigkeit vorliegt und somit Bedarf für eine psychosomatische Reha besteht.“ (115 / zw)
X20	6	„Da die Leistungsfähigkeit für die letzte Tätigkeit (in diesem Fall als Reinigungsfrau) nicht bewertet werden soll, wird auch die etwas unpräzise Formulierung im letzten Absatz auf Blatt 5 hier nicht gewertet.“ (229 / zq)

Prüffrage B4. Die Prüffrage B4 erfordert die Beurteilung, ob sozialrechtliche Ausführungen und Folgerungen innerhalb des Gutachtens unterlassen wurden. Die Begründungen der festgestellten Mängel ergaben sich häufig aus den im Manual hinterlegten Qualitätsaspekten. Leichte Mängel wurden jedoch in keinem der Reviews codiert. Ferner finden sich neben einige Verweisen auf andere Prüffragen (n = 6) und Beispielen (n = 6) auch Freitexte (n = 4), die nicht in der Lage sind, die gewählte Mängelkategorie adäquat im Sinne des Manuals zu begründen (Tab. 35).

Tabelle 35: Codierung der Prüffrage B4.

Prüffrage B4 untersucht, inwieweit die sozialmedizinische Terminologie, gemäß der Definition im Sozialmedizinischen Glossar der Deutschen Rentenversicherung, richtig angewendet wird.		
Code	n	Ankerbeispiel
B4L	0	---
B4D	31	„Begriff ‚Wegefähigkeit‘ wird verwendet“ (40 / zg)
B4G	24	„Die Gutachterin spricht von Arbeitsunfähigkeit [...], berücksichtigt aber nicht, dass es sich um einen Zeitraum handelt, der bereits deutlich über sechs Monaten liegt.“ (248 / zg)
X6	4	„Es erfolgt keine Auseinandersetzung mit ‚sozialrechtlichen Folgen.‘“ (264 / zx)
X7	15	[Keine Textangabe]
X8	6	„S. B2. Aber ‚zum Glück‘ gibt’s ja das Schlussblatt...“ (282 / zi)
X9	1	„Warum sind die beschriebenen psych. Leiden bedeutsam? Unkritische Übernahme der somatischen Diagnosen.“ (179 / zx)
X12	2	„Minimale Anamnese, anschließend mit subjektiver Mischung ‚Beschwerden sind glaubhaft ‚Schlussfolgerungen nicht ableitbar.‘“ (100 / zo)
X13	4	„Seite 8, Abs. 4“ (115 / zq)
X15	1	„... rentenrelevante Verbesserung...“ (160 / zc)
X16	6	„Seite 6, Pkt.5: Der Begriff der Verweisbarkeit ist meines Erachtens ein juristischer Begriff und kein medizinischer.“ (28 / zq)
X18	2	„Dem GA fehlen Klarheit und Übersichtlichkeit. Die vorgeschlagene med. Reha sollte gemäß dem Grundsatz als Eil-Reha vor Abschluss des Rentenverfahrens erfolgen. Erst dann kann über die ja ebenfalls vorgeschlagene berufliche Reha entschieden werden. Ganz falsch wäre aber die Gewährung einer Zeitrente. Die Reha-Maßnahmen könnten dann nämlich erst nach Ablauf der Rente beginnen.“ (84 / zz)
X20	1	„Der Gutachter nimmt Stellung zur Rentenempfehlung der Tagesklinik, was zulässig und erforderlich ist.“ (232 / zt)

Prüffrage C1. Die Prüffrage C1 erfordert die adäquate Beurteilung der Erkennbarkeit von medizinischen Informationen im sozialmedizinischen Gutachten. Die Begründungen der festgestellten Mängel ergaben sich lediglich teilweise aus den im Manual hinterlegten Qualitätsaspekten und bezogen sich unmittelbar auf die sozialmedizinische Relevanz und quantitativ-mengenmäßige Steigerung. Häufiger wurden freitextliche Begründungen gegeben, bei denen augenscheinlich nicht klar ersichtlich ist, ob die Herkunft und/oder zeitliche Einordnung einer medizinischen Information für die Leistungsbeurteilung relevant ist oder nicht (n = 133). Die Betrachtung der freitextlichen Mängeldokumentation ohne Bezugnahme zur Mängelkategorie lässt jedoch keine substanzielle Differenzierung zwischen den Kategorien zu. Lediglich unter gleichzeitiger Berücksichtigung der gewählten

Mängelkategorie ließe sich unter Umständen ein Rückschluss auf die Richtigkeit der Begründung ziehen. Analog verhielt es sich mit den Codes der Kategorie „Beispiele“ (Code X16, n = 41). Durch die reine Aufzählung von Beispielen war es nicht möglich anhand der Mängeldokumentation nachzuvollziehen, zu welcher Mängelkategorie die freitextlichen Begründungen zugeordnet werden sollten. Folglich sind diese Begründungen nicht hinreichend trennscharf und bedürfen bei Anwendung grundsätzlich eines Zusatzhinweises zur sozialmedizinischen Relevanz (Tab. 36).

Tabelle 36: Codierung der Prüffrage C1.

Prüffrage C1 untersucht, inwieweit die Herkunft von medizinischen Informationen im Gutachten erkennbar ist.		
Code	N	Ankerbeispiel
C1L1	2	„Bei den aufgeführten ärztlichen Gutachten wird die Fachrichtung des Gutachters nicht ersichtlich.“ (57 / zm)
C1L2	0	---
C1D1	12	„Mehrere Quellenangaben fehlen bei irrelevanten med. Angaben“ (169 / zc)
C1D2	0	---
C1G1	9	„Für die führende Diagnose Befund der Zweigefäß-Erkrankung fehlt der in der Epikrise beiläufig erwähnte koronarangiographische Befund!“ (228 / zw)
C1G2	3	„Die zeitliche Einordnung (Datum) fehlt bei mindestens einer medizinischen Information, die für die sozialmedizinische Leistungsbeurteilung relevant ist.“ (48 / zg)
X5	133	„Quelle und Herkunft der radiologischen Befunde im Gutachten nicht benannt, aber in der Anlage dokumentiert.“ (55 / zx)
X6	4	„Die im Anhang aufgeführte Kurznachricht über eine 1-tägige Krankenhausbehandlung wird im Gutachten nicht erwähnt.“ (271 / zi)
X7	10	[Keine Textangabe]
X9	12	„Laborwerte werden nicht explizit mitgeteilt. Auch nicht, ob das Routinelabor unauffällig ist.“ (76 / zd)
X10	1	„Quelle wurde benannt aber ohne Namen, nur Fachrichtung und Datum.“ (63 / zs)
X11	6	„Hier wird lediglich auf die Akte verwiesen“ (264 / zo)
X13	3	„Befunde unter 1.1.2“ (223 / zv)
X14	1	„Woher die Ergebnisse der nicht notwendigen Lungenfunktionsuntersuchung (Punkt 7) ist unklar, - aber auch unwichtig“ (232 / zt)
X15	7	„Auf Bl. 3 wird angegeben, keine Begutachtung, in Epikrise erfährt man dann, dass Vers. durch Arbeitsagentur begutachtet wurde, auch wann und mit welchem Ergebnis, aber nicht aufgrund welcher Erkrankung, auch wird nicht ersichtlich, ob es sich um eine fachspezifische Gutachterin handelt.“ (217 / zm)
X16	41	„Fehlt: Herkunft von Ruhe-EKG, Ergometrie, Rö-Hüfte, Bodyplethysmographie, Spirometrie, Laborwerte, Diagnose Anteriorinfarkt.“ (41 / ze)
X18	2	„Der Verweis auf Berichte und Befunde, die in der Akte/Anlage vorliegen (aber nicht jetzt!!) ist ärgerlich (Reha-EB, BB von NP und Orth, CT)“ (89 / zt)
X19	4	„Das liegt aber z.T. an der teilweise hirnrissigen ‚Schwärzungs-Praxis‘. Dadurch ist oft Art und Herkunft der med. Inf. nicht mehr nachvollziehbar. M.E. sollte in Zukunft auf Schwärzungen weitgehend verzichtet werden.“ (134 / zz)
X20	18	„Formal keine Mängel. Es bleibt allerdings etwas unklar, woher die Informationen im Absatz ‚Operationen‘ auf Seite 2 stammen, nämlich ob es anamnestische Angaben sind, oder dies eine Zusammenfassung einer evtl. vorliegenden medizinischen Akte ist.“ (282 / zq)

Prüffrage C2. Die Prüffrage C2 erfordert die adäquate Beurteilung von Verfahren, Methoden und Instrumenten selbst erhobener Befunde. Die Begründungen der festgestellten Mängel ergaben sich wiederum nur selten aus den im Manual hinterlegten Qualitätsaspekten (n = 5). Ähnlich wie in der Prüffrage C1 begründeten die Peers die Wahl der Mängelkategorie primär durch Beispiele (n = 35) oder Äußerungen, bei denen die Angaben zur sozialmedizinischen Relevanz der medizinischen Information fehlten (n = 42). Positiv sei anzumerken, dass lediglich zweimal gegen die Anweisung im Manual verstoßen wurde, einen Verweis auf Befunde in der Anlage nicht mit einem Mangel zu kodieren (Tab. 37).

Tabelle 37: Codierung der Prüffrage C2.

Prüffrage C2 untersucht inwieweit dargestellt wird, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben wurden.		
Code	n	Ankerbeispiel
C2L	1	„Es wurde an Tests ausschließlich der SKT durchgeführt. Es wird weder gesagt, was die Abkürzung bedeutet, noch wird etwas zum Verfahren gesagt. Es liegt nur ein leichter Mangel vor, da dieser Test nicht wirklich sozialmedizinisch relevant ist.“ (51 / zw)
C2D	0	---
C2G	4	„Die Messverfahren zu den relevanten Untersuchungen werden nur angedeutet, z. B. Doppler-Sono, Labor, Fremdbefunde“ (100 / zx)
X4	2	„Nicht beschrieben aber mit/in der Anlage dokumentiert“ (55 / zx)
X5	42	„Bedingungen der Untersuchungsdurchführung bzgl. des Testes fehlen, wie die Lernbehinderung festgestellt wurde, wird nicht erkennbar.“ (51 / zm)
X6	4	„Wertigkeit SKID-II, SFSS“ (199 / ze)
X7	13	[Keine Textangabe]
X8	6	„entspr. C1“ (173 / ze)
X9	2	„bzgl. Vorerkrankungen sind die Quellen nicht ersichtlich“ (72 / zf)
X11	2	„Die Testpsychologie sollte standardisiert beschrieben werden“ (115 / zt)
X12	1	„[...] Diagnose Nummer 5 ist nach meinem Dafürhalten nicht nachvollziehbar. Weder finden sich entsprechende Angaben in den dokumentierten Vorberichten, noch ist nachvollziehbar, warum der Gutachter ‚ASR-Ausfall re.‘ präzisiert, entgegen seinem eigenen körperlichen Untersuchungsbefund 2.9.: ‚MER seitengleich auslösbar‘“. (107 / zq)
X13	2	„Es gibt keine.“ (263 / zf)
X14	2	„es fehlt eine orientierende allg.-internistische Untersuchung“ (48 / zt)
X15	11	„Die durchgeführten Untersuchungen bzw. Ergebnisse werden erst in der Epikrise erwähnt (Bel.-EKG, Lungenfunktion)“ (41 / zg)
X16	35	Lungenfunktion (78 / zz)
X17	1	„Hier stellt sich die Frage, bei wem der Mangel anzusiedeln ist: Beim Gutachter oder bei den Projekt-Leitern von ‚EvaBegut‘ Auf Bl. 6 des Gutachtens findet sich der Hinweis: ‚Sonstige Befunde: Labor, EKG, siehe Anlagen‘. Diese werden hier nicht vorgelegt! Insbesondere das Labor aber wäre entscheidend. Da sich im Text keine konkreten Laborwerte finden lassen, ist der Prüfarzt auf diese Anlagen quasi angewiesen. Wie gesagt, sie wird ihm vorenthalten. Wer hat Schuld? Im Zweifelsfall der Gutachter, deswegen: ‚gravierender Mangel‘“. (155 / zq)
X19	1	„Leider wieder wie so oft ein unvollständig eingereichtes GA. Selbstverständlich sind Messblätter, Laborbefunde und andere techn. Unterlagen mit einzureichen. Sonst sind Fragen wie c1, c2, c3, d2, d5, e1, f1, f2 im Grunde unsinnig, da teilweise unvollständige GA vorliegen.“ (231 / zz)
X20	23	„nicht beurteilbar, Verweis auf Anlage bei LUFU“ (147 / zg)

Prüffrage C3. Die Prüffrage C3 erfordert die adäquate Beurteilung der Darstellung von selbst erhobenen apparativ-technischen Messwerten, Referenzbereichen und Graduierungen, die Grundlage der Beurteilung sind. Verweise auf den Anhang wurden wiederum nicht als Mangel kodiert. Die Begründungen der festgestellten Mängel ergaben sich nur selten aus den im Manual hinterlegten Qualitätsaspekten. Die Peers begründeten die Wahl der Mängelkategorie primär durch Beispiele (n = 92) oder Äußerungen, bei denen die Angaben zur sozialmedizinischen Relevanz der medizinischen Information fehlten (n = 25). Ferner wurden zur Beantwortung dieser Prüffrage häufig Verweise auf vorangegangenen Prüffragen (n = 37), vor allem auf die Prüffrage C2, verwendet (Tab. 38).

Tabelle 38: Codierung der Prüffrage C3.

Prüffrage C3 überprüft, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind.		
Code	n	Ankerbeispiel
C3L	0	---
C3D	0	---
C3G	2	„Mehrere relevante Parameter werden nicht anhand von Referenzbereichen bzw. auch Messwerten dargestellt-> Bsp.: Blutgasanalyse, Diabetes Mellitus, Doppler-Sono, Lungenfunktion, etc.“ (65 / zc)
X4	4	„Seite 6: ‚Bewegung im rechten Schultergelenk schmerzhaft eingeschränkt‘ - hier fehlt die Quantifizierung (Neutral-Null Methode).“ (246 / zq)
X5	25	„Die psychologischen Testungen wurden benannt aber ohne Referenzwerte dargestellt“ (26 / zs)
X6	4	„Von einer Fachärztin für Allgemeinmedizin ist eine eigene Beurteilung des am 07.06.2011 abgeleiteten EKG zu erwarten.“ (271 / zw)
X7	28	[Keine Textangabe]
X8	37	„s. C2“ (162 / zb)
X11	1	„die techn. Befunde sind nicht ‚transparent‘“ (77 / ze)
X12	2	„Intelligenz sei im Normbereich...dann aber doch leicht gemindert“ (122 / zf)
X13	1	„Überprüfung des erhöhten Blutdruck!“ (193 / zx)
X15	6	„Die Aussage, dass sich bei der Lungenfunktion um einen Normalbefunde handelt, ist unzureichend, zumal die Erstdiagnose sich vorrangig auf diesen Befund stützt.“ (104 / zw)
X16	92	„für COPD, für BZ-Werte!“ (53 / zq)
X17	1	„M.E. ist diese Frage überflüssig“ (48 / zx)
X18	2	„Beurteilung der psychischen Leiden m.E. unzureichend“ (48 / zx)
X20	33	„Entfällt“ (216 / zg)

Prüffrage D1. Die Prüffrage D1 erfordert die adäquate Beurteilung der Vollständigkeit der Anamnese. Die Begründungen der festgestellten Mängel ergaben sich sehr häufig aus den im Manual hinterlegten Qualitätsaspekten (n = 214). In ihren freitextlichen Antworten gaben einige Peers (n = 3) auch ausführliche Begründungen ab, die mehr Aspekte enthielten, als ursprünglich von der Prüffrage zu erwarten waren (z. B. aktuelle Medikamentendosis, Umgang mit unstrukturierter Anamnese). Ferner konnten neben diesen „neuen Aspekten“ bei der Beurteilung der Prüffrage auch zu differenzierte Mängelbegründungen (n = 12) und zu pauschale Begründungen (n = 17) kodiert werden (Tab. 39).

Tabelle 39: Codierung der Prüffrage D1.

Prüffrage D1 überprüft, inwieweit die Anamnese vollständig dargestellt wird.		
Code	n	Ankerbeispiel
D1L	121	„Behandler bzw. Fachrichtungen fehlen“ (41 / zg) „Anamnese unvollständig und dürftig. Keine aktuellen Therapien benannt.“ (128 / zs) „Der alleinige Hinweis auf einen Reha-Bericht der mehrere Jahre alt ist ersetzt nicht die Anamnese mit OP und Erkrankungsschilderung, Lebensumstände, laufende Therapien werden nicht aufgeführt“ (41 / zs)
D1D	74	
D1G	19	
X1	1	„Bei aktenkundigem Alkoholabusus mit der sozialmedizinischen Relevanz der Non-Compliance sollte zumindest ansatzweise eine Suchtananamnese erhoben werden.“ (251 / zw)
X2	3	„Hinweisen möchte ich zunächst darauf, dass die Anamnese etwas ‚unstrukturiert‘ ist (vergleiche Blatt 2): In 1.1.b vermischen sich Daten der eigenen Krankheits-Vorgeschichte mit biografischen Daten und Angaben zu den jetzigen Krankheitserscheinungen und Beschwerden. Zum Teil werden die Daten dann in abgewandelter Form in den entsprechenden nachfolgenden Abschnitten wiederholt. Auch fällt auf, dass kein eigener Abschnitt für die in neuropsychiatrischen Gutachten erforderliche Erhebung des Tagesablaufs vorgesehen ist. Elemente zu diesen Themen-Bereich finden sich gleichwohl unter 1.1.b, jedoch nur oberflächlich skizziert, nicht systematisch erhoben. Formal somit kein deutlicher und kein gravierender Mangel (dies mit Bezug auf die obigen Vorgaben), aus meiner Sicht dennoch zumindest leichter Mangel.“ (229 / zq)
X3	2	„Anamnese etwas verworren. Wichtige Aspekte der Eigen-Anamnese bei Arb.-An. oder erst in der Epikrise. (Stat. psychiatr. Beh. und kardiolog. Diagnostik in der Vorgeschichte.)“ (107 / zt)
X6	5	„Bei einzelnen Anamnesepunkten erfolgen Querverweise- etwas unübersichtlich.“ (115 / zc)
X7	10	[Keine Textangabe]
X8	4	„S. A1, keine I-Anamnese, P gar nicht“ (71 / zo)
X9	5	„Therapien kein eigener Gliederungspunkt, lediglich Medikation.“ (196 / zc)
X10	12	„Keine O oder P Anamnese, keine Familienanamnese, keine Vorgeschichte, Internistisch nur Nieren“ (79 / zo)
X11	17	„Sehr kurze Anamnese, anscheinend aber keine weiteren Erkrankungen?“ (115 / zo)
X13	1	„Keine Mitteilungen zur Genese der ‚DCM‘“ (263 / zd)
X14	3	„allerdings gravierend: zu ungenaue Ausführungen zur psychischen Beeinträchtigung, Tagesablauf?“ (76 / zf)
X15	11	„Die Anamnese der psychischen wird anteilig indirekt z. B. unter ‚Epikrise‘ dargestellt.“ (157 / zx)
X17	2	„Gemäß obigem Formalismus: Keine Mängel. Ich vermisste bei einem psychiatrischen Gutachten dennoch die detaillierte Erhebung des Tagesablaufs.“ (139 / zq)
X18	8	„Die Strukturierung/ Gliederung der Anamnese ist ‚chaotisch‘/ quasi nicht vorhanden. Entsprechende Infos liegen größtenteils schon vor, sind jedoch oft in anderen Kontexten fließend miteinander verbunden- Beurteilung?“ (91 / zc)
X19	2	„Die Fachrichtung der behandelnden Ärzte wurde schlicht und einfach geschwärzt - dies dem Gutachter anzulasten, wäre verwegen! Ich könnte im Zweifelsfall natürlich auch gegen den Angeklagten entscheiden - auch das ist verwegen! In jedem Fall handelt es sich formal um leichte Mängel.“ (143 / zq)
X20	17	„Bei angegebener depressiver Stimmungslage wären natürlich erweiterte anamnestische Erhebungen wünschenswert gewesen, mit Rücksicht auf die Fachdisziplin des Gutachters jedoch formal kein Mangel.“ (282 / zq)

Prüffrage D2. Die Prüffrage D2 erfordert die adäquate Beurteilung der Dokumentation der zugrunde gelegten Befunde. Die Begründungen der festgestellten Mängel ergaben sich nur teilweise aus den im Manual hinterlegten Qualitätsaspekten. Aufgrund der Prüffragencharakteristik der Bewertungsmatrix wurden die Mängelstufen im Wesentlichen durch das prüffragenspezifische Charakteristikum (es fehlt ein/mehrere/überwiegende Befund[e]) der zu bewertenden Aspekte definiert. Häufig konnte die Mängelbegründung nicht eindeutig der jeweiligen Mängelstufe zugeordnet werden, da ein Verweis auf die Prüffragencharakteristik fehlte (n = 33). Ferner wurde in 30 Anmerkungen lediglich ein Verweis auf eine andere Prüffrage aufgeführt, meistens war dies der Verweis auf Prüffrage C1 (Tab. 40).

Tabelle 40: Codierung der Prüffrage D2.

Prüffrage D2 überprüft, inwieweit die bei der Beantwortung der sozialmedizinischen Fragestellung zugrunde gelegten Befunde dokumentiert sind.		
Code	n	Ankerbeispiel
D2L	31	„Kernaussagen aus Reha-EB v. 20.11.2004 fehlen (hieraus offensichtlich Ableitung der Inf. über den Schlaganfall)“ (41 / ze)
D2D	3	„Nähere Beschreibung der externen Befunde nur in einem sehr eingeschränkten Maße“ (179 / zx)
D2G	9	„Zur Hauptdiagnose fehlt die Befunddokumentation komplett, auch sonst wird nicht viel dokumentiert.“ (100 / zg)
X3	1	„Kernaussage der Vorbefunde nicht dokumentiert.“ (100 / zp)
X5	33	„Die ‚Kernaussagen‘ der aufgeführten Unterlagen (S. 2 des Gutachtens) fehlen.“ (162 / zn)
X6	11	„Gutachten d. Arbeitsagentur lag offenbar nicht vor“ (179 / zm)
X7	15	[Keine Textangabe]
X8	30	„s. C 1“ (160 / zv)
X9	2	„Ärzte werden keine Benannt, keine Quellenangaben.“ (204 / zs)
X10	1	„Bei der Befunderhebung wird auf Vorgutachten verwiesen, wobei nicht klar wird von wann dieses Vorgutachten stammt und durch wen in wessen Auftrag es erstattet wurde“ (23 / zw)
X11	5	„unspezifischer, sehr kurzer körperlicher-unzureichender-Untersuchungsbefund“ (152 / zo)
X12	2	„Der psychische Befund ist zu knapp, insbesondere kann hieraus keine schwere depressive Episode abgeleitet werden.“ (179 / zw)
X13	2	„...Entwöhnung mit Entlassungsform als arbeitsfähig.“ (164 / zc)
X14	3	„nicht entscheidbar, die Befunde befinden sich nicht in einer vorliegenden Anlage“ (284 / zt)
X15	6	„Minimalistische Untersuchung, keine Diagnostik erwähnt oder durchgeführt auf O od. P Gebiet, aber es tauchen plötzlich schwerwiegende Diagnosen und Verdächtigungen auf.“ (193 / zo)
X16	31	„Tagesklinische Behandlung“ (76 / zv)
X18	6	„Gutachterlicherseits wird von einer suboptimalen Blutdruckeinstellung ausgegangen, die Annahme stützt sich allerdings auf die anamn. Angabe, wohingegen die objektiv erhobenen Befunde dem widersprechen. [...] hieraus sind indirekt qualitative Leistungseinschränkungen ableitbar“ (223 / zq)
X20	14	„Rein formal keine Mängel, da offensichtlich keine Fremdbefunde im engeren Sinn vorliegen, auch keine eigenen Rö.-Aufnahmen etwa angefertigt werden - das Gutachten stützt sich lediglich auf Anamnese und körperlichen Untersuchungsbefund.“ (78 / zq)

Prüffrage D3. Die Prüffrage D3 erfordert die Beurteilung der Formulierung der Diagnosen mit einhergehenden Funktionseinschränkungen (Funktionsdiagnosen nach DRV-Schrift 21). Die Begründungen der festgestellten Mängel ergaben sich sehr häufig aus den im Manual hinterlegten Qualitätsaspekten. Ferner wurde bei 308 der 771 Reviews dokumentiert, dass die Funktionsdiagnosen (teilweise) fehlen. Es wurde als Mangel angegeben, dass Diagnosen nicht nach Schweregrad geordnet wurden. Dieser Aspekt wird jedoch nicht in den zu beurteilenden Mängelkategorien aufgeführt (n = 5). Darüber hinaus fanden sich Gutachtenfälle, bei denen nicht alle Diagnosen, meist nur die Hauptdiagnose, als Funktionsdiagnosen formuliert wurden. Die Peers beurteilten diese Fallkonstellation sehr heterogen. Während einige Peers deutliche Mängel vergeben, beurteilen andere (n = 6) das gleiche Gutachten mit „kein Mangel“ (Tab. 41).

Tabelle 41: Codierung der Prüffrage D3.

Prüffrage D3 überprüft, inwieweit die Diagnosen dargestellt und mit entsprechenden Funktionseinschränkungen (Funktionsdiagnosen) formuliert wurden.		
Code	n	Ankerbeispiel
D3L	29	„Formal deutliche Mängel, da nur zwei Diagnosen ICD-codiert, überwiegend ‚reine Diagnosen‘.“ (49 / zx)
D3D1	2	„Nur die nackten med. Diagnosebezeichnungen“ (100 / zf)
D3D2	308	„Die sich aus der Diagnose ergebene Funktionseinschränkung fehlt.“ (49 / zs)
D3G	12	„ICD-Verschlüsselung fehlt, keine bzw. unzureichende Funktionsdiagnostik.“ (41 / zm)
X3	5	„Keine Auflistung der Störungsbilder bezüglich ihres sozialmedizinischen Schweregrades/Wertigkeit“ (109 / zi)
X4	7	„falsche Kodierung der Hauptdiagnose - kardiologische Befunde werden nicht im Rahmen von Diagnosen dokumentiert (125 Watt Belastung z. B.)“ (119 / zt)
X6	3	„Sämtliche somatischen Diagnosen wurden nicht benannt und entsprechend gewürdigt, aber auf Grund der ausgeprägten nervlichen Erkrankung Leistungsbild nachvollziehbar aufgehoben.“ (6 / zo)
X7	37	[Keine Textangabe]
X8	4	„Kritik wie unter B2“ (59 / zx)
X9	7	„Grenzwertige sprachliche Formulierung [...]“ (152 / zx)
X10	1	„Allgemeinverständliche Übersetzung wäre hier verständlicher“ (57 / zs)
X11	4	„Allenfalls ansatzweise Funktionsdiagnosen“ (170 / zg)
X13	7	„ICDs fehlen im freien Text“ (113 / ze)
X14	6	„Zumindest teilweise Funktionsdiagnosen“ (138 / zg)
X15	1	„Während die Diagnose einer unter Behandlung gut kompensierten Angststörung als Funktionsdiagnose gewertet werden kann, müssen die orthopädischen Diagnosen der Abnützungserscheinungen bemängelt werden: Vorschlag: Abnützungserscheinungen mit Bewegungs- bzw. Belastungseinschränkung?“ (206 / zw)
X17	4	„Meine ‚Lieblings-Frage‘, da hier nach Äpfeln und Birnen gleichzeitig gefragt wird! Wir haben: Keine ICD-Verschlüsselung. Eine Funktionsdiagnose für Diagnose Nummer 1 (von 3). Eine vollkommen diffuse Diagnose Nummer 3 (für die mangels Eindeutigkeit durchaus das Kriterium einer unzureichenden sprachlichen Formulierung zutreffen würde). Was bitte soll ich ankreuzen? Im Zweifel gegen den Angeklagten!“ (108 / zq)
X18	2	„Diagnosen auf Seite 8, ICD-Verschlüsselung auf Schlussblatt Teil 2., 3. Diagnose ‚Laborauffälligkeiten‘ zu ungenau da eindeutige ‚diabetische Stoffwechsellaage und Blutarmut‘“ (94 / zx)
X20	10	Diagnosen nur auf Schlussblatt aufgeführt, dort aber als Funktionsdiagnose, daher nicht als Mangel eingestuft (170 / zm)

Prüffrage D4. Die Prüffrage D4 erfordert die Erfassung der vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen in der Epikrise. Die Begründungen der festgestellten Mängel ergaben sich nur teilweise aus den im Manual hinterlegten Qualitätsaspekten. Neben reinen Beispielen ($n = 119$) wurden auch bei Prüffrage D4 Begründungen gegeben, bei denen nicht ersichtlich wurde, ob die Auseinandersetzung für sozialmedizinisch relevante oder nicht relevante Beschwerden erfolgte ($n = 13$). Da die Bewertungscharakteristik der Prüffrage auf dieser Differenzierung beruht, konnten diese Begründungen nicht eindeutig einer Mängelkategorie zugeordnet werden und dienen daher nicht als ausreichende Begründung. Ferner hat die Prüffrage einen relativ hohen Anteil an fehlenden Mängelbegründungen ($n = 38$). Tabelle 42 zeigt eine Übersicht über die Codierungen dieser Prüffrage.

Tabelle 42: Codierung der Prüffrage D4.

Prüffrage D4 untersucht, inwieweit die vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen in der Epikrise erfasst wurden.		
Code	n	Ankerbeispiel
D4L	8	„Epikrise setzt sich nicht mit allen vom Versicherten geltend gemachten Beschwerden auseinander (LWS fehlt, ist sozialmed. nicht relevant“ (201 / zg)
D4D	2	„Für einzelne, vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen, die sozialmedizinisch relevant sind, fehlt die Auseinandersetzung (LWS Knie)“ (242 / zg)
D4G	50	„Es erfolgt keine Auseinandersetzung mit den Beschwerden, die die Versicherte vorbringt“ (179 / zg)
X1	3	„Man kann nicht sagen, dass eine Auseinandersetzung komplett fehlt, jedoch erfolgt diese sehr pauschal und nicht bezogen auf die Beschwerden/Symptome im Detail. Auf Symptome aus fachfremden wird nicht eingegangen“ (282 / zq)
X5	13	„Eine Auseinandersetzung erfolgt lediglich für die depressive Symptomatik, sowie für die Anämie. Viele geltend gemachte Beschwerden sind unspezifisch und wären grundsätzlich der depressiven Störung zuzuordnen“ (179 / zq)
X6	3	„führende Diagnose entspricht nicht den Hauptbeschwerden“ (41 / zu)
X7	38	[Keine Textangabe]
X8	4	„siehe D3“ (53 / zo)
X9	1	„Keine ausreichende Anamnese“ (41 / zo)
X10	2	„Bzgl. der ‚geltend gemachten Beschwerden‘: Kein Mangel. Aus der ‚medizinischen Anamnese‘ lassen sich jedoch mehrere internistische Erkrankungen ableiten, welche in keinsten Weise abgehandelt werden. Es ist jedoch auch andererseits nicht ersichtlich, ob und inwiefern diese internistischen Erkrankungen aus der Vorgeschichte zum Zeitpunkt der Begutachtung Funktionsdefizite generieren.“ (128 / zq)
X11	1	„nur angedeutet“ (71 / zo)
X12	3	„Aus fehlender Anamnese und minimaler körperlicher Untersuchung ist das Ergebnis nicht ableitbar“ (48 / zo)
X13	1	„da konkrete Beschwerden in Bezug auf ES und Atmung nicht erfragt werden“ (284 / zf)
X15	6	„Beschwerden werden nicht erläutert, sondern pauschal abgegolten, wie: es findet sich kein organisches Korrelat für geschilderten Beschwerden“ (57 / zs)
X16	119	„Schwindel, Kollaps, Kopfschmerzen, Kribbelgefühlsstörungen, u.a.“ (34 / zm)
X18	9	„Wo leitet der Gutachter ab, dass die Versicherte in ihren Aktivitäten des täglichen Lebens unter Medikation nicht kompromittiert ist?“ (206 / zi)
X19	2	„kann hier nicht beurteilt werden, da die entscheidenden Seiten fehlen- somit keine Mängel?“ (222 / zc)
X20	10	„auch wenn nicht konkret auf die Beschwerden eingegangen wird, wird zu allen betroffenen Organsystemen“ (34 / zg)

Prüffrage D5. Die Prüffrage D5 erwartet die Auseinandersetzung mit der medizinischen Relevanz von Fremdbefunden und selbst erhobenen Befunden in der Epikrise. Die Begründungen der festgestellten Mängel ergaben sich lediglich teilweise aus den Manual. Infolge der prüffragenspezifischen Charakterisierung der Bewertungsmatrix begründeten die Peers, analog zur vorangegangenen Prüffrage D4, die gewählten Mängelkategorien vorrangig mit Beispielen (n = 115), bei denen aus den Antworten nicht ersichtlich wurde, ob die Auseinandersetzung für sozialmedizinisch relevante oder nicht relevante Beschwerden erfolgte. Ferner stellten 34 Freitexte lediglich Verweise auf bereits beantwortete Prüffrage dar (z. B. D4). Auch bei Prüffrage D5 gab es einen hohen Anteil (n = 33) an fehlenden Mängelbegründungen (Tab. 43).

Tabelle 43: Codierung der Prüffrage D5.

Prüffrage D5 überprüft, inwieweit die im Gutachten aufgeführten Befunde in der Epikrise erfasst sind.		
Code	n	Ankerbeispiel
D5L	1	„keine Auseinandersetzung mit den erhobenen Befunden bzgl. Wirbelsäule, untere Extremität, dies fällt allerdings auch eher ins orthopädische Fachgebiet und da internistische Leiden objektiv im Vordergrund stehen nicht so relevant“ (49 / zm)
D5D	1	„Für einzelne im Gutachten aufgeführte und sozialmedizinisch relevante Befunde fehlt die Auseinandersetzung, z. B. bzgl. pAVK, LWS“ (167 / zp)
D5G	37	„keine Auseinandersetzung mit Befunden“ (107 / zg)
X5	9	„eher Aufzählung der auffälligen Befunde ohne hinreichende Auseinandersetzung“ (51 / q)
X6	1	„da nicht erfasst, können sie auch nicht ausgewertet werden“ (242 / zo)
X7	33	[Keine Textangabe]
X8	34	„siehe D4“ (71 / zc)
X11	2	„Epikritische Auseinandersetzung mit allen Leiden völlig unzureichend“ (55 / zx)
X12	1	„Die Diagnosen und die Diskussion in der Epikrise sind nicht aus der Befunderhebung und der Untersuchung ableitbar“ (193 / zo)
X13	2	„Angeblicher Bluthochdruck_ Ursache und Art der Anämie (bekannt schon seit 3/10, Ga-Untersuchung 8/10)“ (179 / zx)
X14	1	„Nur Auseinandersetzung mit den neuro/psych. Befunden (neuro/psych. Gutachten!)“ (198 / zx)
X15	3	„Es erfolgt lediglich eine Auseinandersetzung des Gutachters mit der psychischen Problematik ‚im Eildurchlauf‘, allein basierend auf kurzer Anamnese ohne Berücksichtigung weiterer Kontextaspekt_ so ergibt sich eine rasch formulierte ‚Eildiagnose: Anpassungsstörung‘.“ (109 / zi)
X16	115	„Regelrechter Reflexstatus bei Diagnose iatrogene Nervenläsion“ (107 / zs)
X17	1	[Bewertung: keine Mängel] „... bei rein formaler Betrachtung!“ (129 / zq)
X18	2	„Wieso besteht eine suboptimale Blutdruckeinstellung? m.E. gute Blutdruckeinstellung!“ (223 / zx)
X19	1	„??? Seite 12 (Epikrise) fehlt“ (102 / zp)
X20	14	„Ergebnis der Reha-maßnahme nicht bekannt, Gutachter setzt sich nicht mit Reha-Bericht auseinander, allerdings Reha-Bericht alt, Ergebnis ebenfalls nicht bekannt: Mangel? habe dies als nicht Mangel da Reha 7 Jahr her ist, eingestuft“ (41 / zm)

Prüffrage D6. Die Prüffrage D6 erfordert die Prüfung, inwieweit die im Gutachten abgeleiteten Funktionsstörungen auch sozialmedizinisch bewertet werden. Die Prüffrage

bezieht sich auf die Vollständigkeit, nicht aber auf die Richtigkeit der sozialmedizinischen Bewertung. Die Begründungen der Mängel ergaben sich auch an dieser Stelle nur teilweise aus den im Manual hinterlegten Mangelkategorien. Die Peers begründeten die festgestellten Mängel vorwiegend anhand von Beispielen (n = 57) und Verweisen (n = 41). In seltenen Fällen (n = 4) wurden die Einschränkungen des Manuals nicht beachtet und die Richtigkeit und Nachvollziehbarkeit der Leistungsbeurteilung bewertet. Auch wurde in sieben Fällen ein Mangel lediglich indirekt beantwortet, sodass eine genaue Differenzierung der Mangelkategorien anhand der freitextlichen Beurteilung nicht möglich war. Darüber hinaus fehlten in 33 Fällen die freitextlichen Mängelbegründungen (Tab. 44).

Tabelle 44: Codierung der Prüffrage D6.

Prüffrage D6 überprüft, inwieweit die festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit im Erwerbsleben bewertet werden.		
Code	n	Ankerbeispiel
D6L	1	„aufgrund der vorliegenden Befunde“ ist die zuletzt ausgeübte Tätigkeit nicht mehr verrichtbar. Ich hätte mir gewünscht, dass bei dem folgenden qualitativen Einschränkungen eine Zuordnung erfolgt, ob dies a) wegen der Gonarthrose b) wegen des Diabetes oder c) wegen der COPD erforderlich ist“ (92 / zw)
D6D	1	„Die Funktionsstörungen auf Seiten des orthopädischen Fachgebietes sind klar benannt. Im Hinblick auf den als übermäßig eingestuften Alkoholgebrauch muss aus sozialmedizinischer Sicht eine erhebliche Gefährdung der Erwerbsfähigkeit festgestellt werden“ (251 / zw)
D6G	37	„Eine funktionsgebundene Betrachtungsweise erfolgt in der Epikrise weder im Hinblick auf die ortho. Diagnose noch auf die psych. Situation“ (35 / zw)
X1	4	„Bewertung nur allgemein, es wird nicht deutlich, wegen genau welchen Erkrankungen der Vers. in der Leistungsfähigkeit eingeschränkt ist“ (34 / zv)
X4	4	„Nicht nachvollziehbare Leistungsfähigkeitsbeurteilung auf allg. AM“ (41 / zd)
X5	10	„Die Funktionsstörungen werden nicht einzeln bewertet, es erfolgt eine summarische Beurteilung“ (124 / zu)
X6	5	„Bewertung tw. auf neurolog. Zusatzgutachten verschoben“ (154 / zf)
X7	36	[Keine Textangabe]
X8	41	„s.D4“ (166 / zf)
X10	2	„falls wirklich ‚Schwindel‘ und ‚Inkontinenz‘ geltend gemacht wird ‚gehört die Problematik vertieft‘ - der ‚vorgealterte Aspekt‘, dass ‚blasse Hautcolorit‘, ‚Nykturie 3x‘ sind allgemeinmedizinisch nicht unwesentlich, aber auch unspezifisch“ (219 / zt)
X11	5	„Nur andeutende Bewertungen“ (55 / zx)
X12	4	„Es werden aus den Diagnosen keine Funktionsstörungen abgeleitet“ (57 / zs)
X13	1	„Psyche wird diskutiert; Hauptthema Blutarmut ohne weiteren Beleg“ (179 / zs)
X14	1	„Es bestehen Gesundheitsstörungen auf nicht-orthopädischem Gebiet. Zwar wird anhand der Befunde nahegelegt, dass weitere med. Ermittlungen nicht empfohlen sind, dennoch wäre eine eindeutige Stellungnahme klarer“ (166 / zt)
X15	7	„Die psychiatrischen Funktionsstörungen werden nur rudimentär erörtert in der sozialmedizinische Leistungsbeurteilung. Dadurch ist das Gutachten angreifbar“ (199 / ze)
X16	57	„Kein Belastung-EKG bei KHK, keine hinreichende Aussagekraft zur kardialen ‚Beschwerdefreiheit‘“ (68 / zi)
X18	3	„Das Leistungsbild wurde nur unzureichend formuliert, Im Schlussblatt wurde dann noch eine falsche Angabe gemacht ‚zeitweise stehen, gehen, sitzen‘ damit werden keine 100 % erreicht und somit würde der Antragsteller Rente bekommen, obwohl eine volle Leistungsfähigkeit vorliegt“ (41 / zo)
X19	1	„??? Seite 12 (Epikrise) fehlt“ (102 / zp)
X20	13	„Keine Mängel - berücksichtigt man, dass die Ausformulierung der qualitativen Leistungseinschränkungen im Formularteil Abs. 3.“ (224 / zq)

Prüffrage D7. Die Prüffrage D7 untersucht, ob in der Epikrise eine Stellungnahme zu bisherigen Therapien und ggf. weiteren Behandlungen erfolgt. Die Begründungen der festgestellten Mängel ergaben sich überwiegend aus den im Manual hinterlegten Qualitätsaspekten. Mängel wurden häufig mit ausführlicheren Aspekten, die z. T. über die Ausprägung der Mängelkategorien hinausgehen begründet. So wurden beispielsweise im Gutachten getätigte zusätzlich Stellungnahmen z. B. nach deren Genauigkeit beurteilt. Teilweise wurden Beispiele bei der Mängeldokumentation herangezogen (n = 32), die aufgrund fehlender Begleitinformationen keine Differenzierung zwischen den Kategorien ermöglichten. Darüber hinaus sind bei Prüffrage D7 auch pauschale Begründungen kodiert worden (n = 10), die ebenfalls keine eindeutige Differenzierung oder lediglich indirekt (n = 15), unter Einbezug der Mängelkategorie, eine eindeutige Zuordnung zulassen. Erneut ist der Anteil an Peers (n = 89), die eine der definierten Mängelkategorien vergeben, jedoch die festgestellten Qualitätsdefizite nicht dokumentierten relativ hoch (Tab. 45).

Tabelle 45: Codierung der Prüffrage D7.

Prüffrage D7 untersucht, inwiefern eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen erfolgte.		
Code	n	Ankerbeispiel
D7L	56	„Keine Stellungnahme zur bisherigen Behandlung: - Antiobstruktive Therapie? - Nikotinkarenz/Entwöhnung?“ (49 / zi)
D7D	88	„keine Therapieoptionen aufgezeigt (z. B. weitere Gewichtsreduktion, Psychotherapie)“ (147 / zp)
D7G	98	„Keine detailhafte Stellungnahme zur bisherigen Therapie, auch nicht dezidiert zu offenen Therapie-Optionen - Anmerkung: Die pauschale Behauptung, dass durch Reha-Maßnahmen nichts zu erreichen sei, halte ich formal für unzureichend und inhaltlich für unzutreffend.“ (100 / zq)
X2	1	„Die Angaben zu den bisherigen Therapien sind punktuell ungenau. Auf Entgiftungsbehandlungen wird zwar hingewiesen, in der Epikrise findet sich hierzu keine Stellungnahme, auch nicht dazu, dass im orthopädischen Bereich die ambulanten Behandlungsmöglichkeiten nicht genutzt werden.“ (34 / zq)
X3	2	„durch die Qualifizierung der Hauptdiagnose als Verdachtsdiagnose distanziert sich der GA der ambulanten Beurteilung. Hierzu würde man dann auch eine Erörterung erwarten“ (193 / zt)
X6	2	„Insbesondere vor dem Hintergrund des auffälligen Lungenbefundes wären weitere Angaben sinnvoll gewesen“ (88 / ze)
X7	89	[Keine Textangabe]
X11	10	„nur sporadische Stellungnahme, subjektiv gefärbt“ (107 / zo)
X12	1	„Warum hat der Pat. Pankreatitis, es waren lediglich 2 kleine Entzündungsschübe aus den Unterlagen ableitbar, ist die Behandlungsmöglichkeit ausgeschöpft? 2. Diagnose schon PaVK, obwohl keine Gehstreckeneinschränkung usw“ (100 / zo)
X13	3	„Auseinandersetzung mit der derzeit wenig intensiven antidepressiven Therapie“ (76 / zf)
X14	3	„keine Angaben zu bisherigen Behandlungen oder Therapien, die geeignet sind, eine Besserung herbeizuführen, wie z. B. Krankengymnastik“ (242 / zm)
X15	15	„keine Stellungnahme“ (71 / zo)
X16	32	„Knie, Alkohol, Haschisch, Rauchen“ (181 / zx)
X18	6	„fachpsychiatrische Mitbehandlung findet statt. Man weiß nicht seit wann die Therapien erfolgen und ob bzw. inwiefern die Pat. hiervon profitiert. Chron. Schmerzen, werden überhaupt nicht diskutiert wird!“ 107 / zw)
X20	15	„Entfällt mangels bisheriger Therapie und mangels therap. Optionen.“ 236 / zq)

Prüffrage D8. Die Prüffrage D8 prüft, ob wesentliche Abweichungen der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber dem Vorgutachten dargelegt wurden. Liegen aus Vorgutachten keine Aussagen zur Leistungsfähigkeit vor, so führt dies nicht zu einem Mangel. Die Begründungen der festgestellten Mängel wurden nur selten aus den Mängelkategorien des Manuals abgeleitet. Generell erschien die Mängeldokumentation dieser Prüffrage erschwert. So konnte seitens der Peers nicht nachvollzogen werden, ob Abweichungen im Gutachten vorhanden waren, diese jedoch von Seiten des Gutachters nicht angegeben wurde oder ob generell keine Aussagen aus Vorgutachten vorlagen. Die Mehrheit der Peer-Gutachter (n = 120) vergab infolgedessen keine Mängel, beschrieben jedoch ihre Einwände und Bedenken. Darüber hinaus wurde 47-mal ein Mangel dokumentiert, der über die Mängeldefinition nicht abgedeckt wird, obgleich dieser im erläuternden Text der Prüffrage aufgeführt ist. So merkten die Peers an, dass eine Stellungnahme gänzlich fehlte oder die Beurteilung nicht erwähnt sei. Folglich könne keine Beurteilung der Prüffrage erfolgen (Tab. 46).

Tabelle 46: Codierung der Prüffrage D8.

Prüffrage D8 prüft, inwieweit wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber dem Vorgutachten dargelegt wurden.		
Code	n	Ankerbeispiel
D8L	2	„na ja, gegenüber Reha-EB wird quasi-quasi ein verbessertes L-bild beschrieben (Leicht bis mittelschwer vs. mittelschwer)“ (242 / zt)
D8D	5	„Die Einschätzung der Orth. Rehaklinik wird lapidar abgetan, es fehlt eine differenzierte Bewertung“ (107 / zx)
D8G	7	„Abweichung zum AA-GA (s. 1.4) werden nicht diskutiert.“ (49 / zp)
X3	47	„Nicht beurteilbar, da der Gutachter die Vorbefunde nicht reflektiert.“ (198 / zx)
X4	1	„Angeblich keine Vorgutachten oder sozialmedizinische Vorbefunde. Aber unter Epikrise Hinweis auf einen GdB40 (Rheuma und Wirbelsäule)“ (193 / zx)
X7	11	[Keine Textangabe]
X8	4	„s.o.“ (181 / zo)
X13	3	„Keine Bezugnahme auf AHB-EB 2010 (LV dort?) sowie auf BG-Gutachten“ (178 / zf)
X14	1	„Abweichungen zum Gutachten der ARGE nicht erläutert, aber auch nur der psychologische Teil übernommen eine LF nicht ersichtlich“ (39 / zs)
X15	8	„Die Gegenüberstellung der im Rahmen einer Deeskalation nunmehr gebesserten Befunde mit den in den Vorgutachten beschriebenen hätte zur Transparenz beigetragen.“ (21 / zi)
X18	2	„In der Epikrise wird festgestellt, dass die Beurteilung der orthopädischen Rehaklinik (Entlassung als arbeitsfähig!) die Depression nicht gewürdigt habe. Leider erfolgt auch im nervenfachärztlichen Gutachten keine Würdigung!“ (107 / zw)
X19	1	„Seite 9/10 (Epikrise)sind leider nicht übersandt worden (trotzdem ist das Leistungsbild nachvollziehbar hergeleitet“ (222 / zo)
X20	120	„da nicht bekannt, zu welchem Ergebnis die Reha-klinik kam (welche Leistungseinschätzung?) kann Frage nicht beantwortet werden“ (27 / zm)

Prüffrage D9. Die Prüffrage D9 prüft, inwieweit die sozialmedizinischen Fragestellungen vollständig beantwortet wurden. Die Begründungen der festgestellten Mängel ergaben sich

überwiegend aus den im Manual hinterlegten Mangelkategorien. Nur in seltenen Fällen (n = 10) wurde gegen die Einschränkungen des Manuals verstoßen. Ein neuer Gesichtspunkt, der einmal kodiert wurde, beschreibt Unstimmigkeiten zwischen den Reha-Empfehlungen im freien Text und auf dem Schlussblatt. Diese Abweichungen werden über die Mangelkategorien des Manuals nicht abgedeckt. Neben einigen wenigen zu stark differenzierten (n = 3) Mängelbegründungen, konnten auch sehr verallgemeinerte Qualitätsmangeldokumentationen (n = 8) festgestellt werden. Der Verweis auf vorangegangene Prüffragen wurde 14-mal angewendet (Tab. 47).

Tabelle 47: Codierung der Prüffrage D9.

Prüffrage D9 untersucht, inwieweit sozialmedizinische Fragestellungen vollständig beantwortet werden.		
Code	n	Ankerbeispiel
D9L	128	„Es fehlen Aussagen zum Reha-Bedarf“ (92 / zg)
D9D	72	„Ankreuzverfahren Reha‘. Beginn der Leistungsminderung fehlt“ (193 / zi)
D9G1	29	„bei qualitativen Einschränkungen genügt das Ankreuzen nicht, zumal nicht erkennbar wird, was nun eingeschränkt ist, alles oder nur z. B. Publikumsverkehr? besteht psychisch bedingte Einschränkung der Gehstrecke? können öffentliche Verkehrsmittel benutzt werden?“ (115 / zm)
D9G2	13	„Es fehlen Aussagen zur qualitativen Leistungsfähigkeit und konkrete Aussagen zur quantitativen Leistungsfähigkeit (>6 Std.)“ (48 / zp)
X1	1	„Der Gutachter empfiehlt ausdrücklich eine STATIONÄRE Reha-Maßnahme, kreuzt aber auf dem Schlussblatt die Möglichkeit ‚auch einer ambulanten Reha-Leistung‘ an.“ (115 / zi)
X3	5	„inhaltlich wurde Leistungen zur Teilhabe am Arbeitsleben empfohlen, auf dem Schlussblatt jedoch negiert“ (166 / zm)
X4	10	„Aussagen zum Beginn der Leistungsminderung werden zu beschrieben, sind aber falsch („AB Rentenanspruch?“)“ (257 / zx)
X6	2	„das qualitative L-bild hätte etwas ausführlicher beschrieben werden können: ein Hinweis auf mögliche L-minderung durch "nicht-nervenärztliche" Gesundheitsstörungen könnte nicht schaden“ (48 / zt)
X7	30	[Keine Textangabe]
X8	14	„s. D7“ (61 / ze)
X9	2	„Schlussblatt B.1. Arbeitshaltung 3x zeitweise ist mit vollschichtigem Leistungsvermögen nicht vereinbar.“ (212 / zz)
X10	3	„Die sozialmedizinische Beurteilung endet mit der Feststellung, dass weitere Rehabilitationsmaßnahmen seitens des Rentenversicherungsträgers in Verlauf nützlich eingesetzt werden könnten‘. Dies ist zu vage und nicht umsetzbar. Vorschlag: Nach weiteren gefäßchirurgischen Interventionen könnte eine Anschluss-Reha in einer kardio. Einrichtung erfolgen.“ (50 / zw)
X11	8	„keine Stellungnahme“ (107 / zo)
X12	3	„Es fehlen Aussagen zu Reha-Bedarf, Beginn und Dauer einer Leistungsminderung und es wird nicht begründet wieso eine vollschichtige Arbeitsfähigkeit der Patientin derzeit nicht möglich ist.“ (282 / zw)
X13	3	„In Epikrise wird deutliche Einschränkung für leichte Tätigkeit erwähnt, deswegen Kompromiss: deutliche Mängel‘ statt ‚gravierende Mängel‘“ (49 / zc)
X14	2	„Reha und teilhabe angekreuzt aber nicht ausgeführt“ (92 / zs)
X15	4	„keine Begründung gegen Reha, widersprüchliche Angaben zu LTA zw. Epikrise und Formularteil“ (51 / zf)
X16	2	„Gehfähigkeit?“ (282 / ze)
X18	4	„Als Beginn der Leistungsminderung wird Dezember 2010 genannt, wobei dies nicht begründet wird und sich aus dem Gutachten nicht erschließt.“ (201 / zw)
X20	33	„Da alle Fragen beantwortet wurden und das Schlussblatt vollständig ausgefüllt wurde, sind formal keine Mängel zu attestieren. Es wird jedoch keine Begründung für die Beurteilung gegeben.“ (179 / zw)

Prüffrage E1. Die Prüffrage E1 prüft, inwieweit der allgemein anerkannte Stand medizinischer Erkenntnisse eingehalten wird. Die Schlüssigkeit der Diagnose und deren Ableitung werden an dieser Stelle nicht geprüft. Die Peers vergaben im Vergleich zu anderen Prüffragen tendenziell weniger Mängel. Auffällig war, dass die prüffragenspezifischen Codes nicht vergeben wurden. Stattdessen wurden Verweise auf andere Prüffragen kodiert (n = 14), Begründungen verwendet, die tendenziell bei anderen Prüffragen Anwendung finden (n = 4) oder gegen die Einschränkungen des Manuals verstoßen (n = 2). Tabelle 48 zeigt eine Übersicht über die vergebenen Codes.

Tabelle 48: Codierung der Prüffrage E1.

Prüffrage E1 untersucht, inwieweit der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet wird.		
Code	n	Ankerbeispiel
E1L	0	---
E1D	0	---
E1G	0	---
X4	2	„schwere depressive Episode liegt nach den ICD-Kriterien nicht vor. Regelmäßiges aktives Fußballspielen im Verein spricht auch dagegen“ (105 / zf)
X6	6	„Keine "Belastungsprüfungen" des Bronchialleidens“ (55 / zx)
X7	4	[Keine Textangabe]
X8	14	„siehe D7“ (131 / zf)
X9	4	„Angaben zu bisherigen Therapien zu dürftig.“ (107 / zv)
X10	1	„Es erfolgt keine Begründung. Bzgl. Prognose und Therapie wird nicht ausreichend bei Depression Stellung genommen (immer nur kurz stat. Behandl., keine Phasenprophylaxe, fortgesetzter Alkoholkonsum wird nicht überzeugend objektiviert).“ (76 / zv)
X12	1	„s.o. Aus dem Untersuchungsbefund nicht erklärbares Leistungsbild Keine Auseinandersetzung mit dem Patienten“ (193 / zo)
X13	1	„fachfremd“ (56 / zd)
X18	3	„Gewinnung med. Information schlecht, insbesondere psychische Befund zu dürftig. Die 2. Diagnose lässt sich aufgrund der Untersuchung nicht ableiten, steht nur in einem Arztbrief, Vers. gibt keine entsprechende Beschwerden an. Auf Prognose und Therapieoptionen wird nicht eingegangen. Vorbefund von psychot. Symptomen wurde nicht hinterfragt, evtl. Medikation eher ungünstig, da Depression fördernd, falls keine psychotische Symptome vorliegend.“ (179 / zv)
X3 / X10	3	„Ursachen der somatischen Leiden werden nicht epikritisch bewertet. Es ist nicht ausgeschlossen, dass die somatischen Leiden zumindest teilweise die psychische Erkrankungen mitverursachen (zum Beispiel die Depression bei Anämie)“ (179 / zx)

Prüffrage E2. Die Prüffrage E2 soll grobe Abweichungen von der Begutachtungsliteratur, d.h. von der Rentenversicherung veröffentlichte sozialmedizinische Begutachtungsliteratur und Leitlinien, aufdecken. Auffällig ist auch bei Prüffrage E2, dass die prüffragenspezifischen Mängelkategorien nur selten vergeben wurden (n = 15). Stattdessen wurden die Prüffrage häufig indirekt durch Beispiele beantwortet (n = 85). Diese enthielten zumeist konkrete Angaben darüber, an welchen Teilen des Gutachtens etwas fehlte bzw. vom Gutachter nicht leitliniengerecht gearbeitet wurde. Infolge der

Bewertungscharakteristik kann dadurch nicht explizit differenziert werden, ob diese Abweichung von der Begutachtungsliteratur relevant oder nicht relevant ist. Eine eindeutige Begründung für die gewählte Mängelkategorie wurde daher nicht abgegeben. Vereinzelt wurden auch Verweise auf andere Prüffragen zur Begründung herangezogen (n = 16). Auch wiesen einige Peers darauf hin (n = 3), dass etwaige Abweichungen nicht dokumentiert werden können, da Kenntnisse über die entsprechende Begutachtungsliteratur anderer Fachgebiete fehlen würden. Dies schlug sich auch in Positivbemerkungen der Peers nieder, die angaben, dass Mängel nicht erkenn- bzw. beurteilbar seien (Tab. 49).

Tabelle 49: Codierung der Prüffrage E2.

Prüffrage E2 untersucht, inwieweit die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet wird.		
Code	n	Ankerbeispiel
E2L	0	---
E2D E2G	0 15	---
X5	1	„Angegebene Gehstreckenminderung so nicht aus Begutachtungsbuch ableitbar (6. Aufl., 7. Kapitel)“ (113 / ze)
X6	4	„Wünschenswert wäre eine aktuelle Rheumaserologie und ein aktuelle Bildgebung (Schulter li., Hände).“ (123 / zi)
X7	22	[Keine Textangabe]
X8	16	„s. D1“ (107 / zf)
X9	4	„Die Datierung des Versicherungsfalles ist falsch (lt. Gutachten Zeitpunkt Versicherungsfall=Zeitpunkt Rentenantrag?)“ (157 / zx)
X10	2	„Der Gutachter sieht selbst einen Reha- bzw. KH-stationären Behandlungsbedarf, lässt sich aber von der Haltung der Vers. unter Verkennung der Mitwirkungspflichten der Vers. davon abbringen. Auch werden die Möglichkeiten einer sozialpsychiatrischen Intervention, bei der ihrem Mann unterworfenen Vers. nicht diskutiert. Die prognostische Einschätzung des Gutachters ist daher "zu dünn" begründet.“ (36 / zi)
X11	2	„Sozialmedizinische Literatur ist anscheinend nicht berücksichtigt worden >>>siehe mehrere grobe Fehler im GA“ (41 / zo)
X12	2	„Der Gutachter stützt sich nicht auf den selbst erhobenen eher blanden psychopathologischen Befund, sondern legt einen nachträglich angeforderten Befundbericht seiner Hauptdiagnose zu Grunde. Dies untergräbt de facto das Neutralitätsgebot. Einige Diagnosen sind nicht nachvollziehbar, wie bereits weiter oben dargelegt (Schilddrüsenunterfunktion, Bluthochdruck). Für das allergische Asthma bronchiale finden sich keinerlei Funktionsprüfungen, es kann somit sozialmedizinisch kaum bewertet werden.“ (179 / zq)
X15	5	„Beurteilung erfolgte ‚nach Bauchgefühl‘“ (264 / zx)
X16	85	„BGA fehlt. Keine Bewertung des erhöhten MCV, keine CDT-Bestimmung.“ (26 / zi)
X17	3	„Ich kann keine erkennen, aber meine Kenntnisse der entsprechenden Fachliteratur sind natürlich sehr begrenzt“ (13 / zg)
X18	7	„Bei fehlenden Angaben zur Händigkeit wäre bei Linkshändigkeit die Beurteilung falsch.“ (160 / zv)
X20	14	„für mich nicht erkennbar/beurteilbar“ (51 / zg)

Prüffrage F1. Die Prüffrage F1 untersucht die Diagnostik und orientiert sich am Wirtschaftlichkeitsgebot des SGB V. Es soll geprüft werden, ob die eingesetzte medizinisch-technische Diagnostik und testpsychologische Untersuchungen zur Beantwortung der Fragestellung geeignet sind. Wie beim Einzelkriterium E „Medizinisch-wissenschaftliche Grundlagen“ wurden auch bei dieser Prüffrage prüffragenspezifischen Mängelkategorien selten vergeben (n = 7). Am häufigsten wurden die Prüffragen indirekt durch das Aufzeigen von Beispielen (n = 68) beantwortet. Darüber hinaus wurden einige Reviews identifiziert (n = 8), bei denen die Antworten der Prüffrage F1 mit denen von F2 verwechselt wurden. In 20 Fällen fehlte eine freitextliche Mängeldokumentation, obwohl ein Mangel beurteilt wurde (Tab. 50).

Tabelle 50: Codierung der Prüffrage F1.

Prüffrage F1 untersucht, inwieweit die Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig ist.		
Code	n	Ankerbeispiel
F1L	2	„Die Lungenfunktionsprüfung ist überflüssig. Die pathologischen Laborwerte sollten benannt werden in Zahlen. Eine Abdomensonographie wäre wünschenswert gewesen oder zumindest die Mitteilung eines Vorbefundes.“ (100 / zd)
F1D	0	---
F1G	5	„Die eingesetzte Diagnostik ist komplett überflüssig und bei der sozialmedizinischen Fragestellung auch nicht erforderlich (Belastungs-EKG).“ (209 / zd)
X6	4	„im Prinzip“ sind Anamnese und Befund ausreichend (s.a.u.)“ (200 / zt)
X7	20	[Keine Textangabe]
X8	2	„S. E2!“ (26 / zi)
X9	8	„Unzureichend, kein Labor bei Diabetes, Kein Röntgen-Aufnahme oder zumindest Auswertung der letzten RÖ-Aufnahmen vor ca 1,5 Jahren“ (92 / zo)
X13	3	„Nur mit Bezug auf die Skelettleiden“ (128 / zx)
X15	1	„Sehr umfangreiche Psychometrische Testung, die aber nicht weiter epikritisch gewürdigt wird!“ (106 / zx)
X16	68	„RÖ der WS m.E. überflüssig“ (204 / zb)
X17	1	„Frage muss entfallen“ (s.o.)“ (109 / zi)
X20	18	„Entfällt.“ (33 / zi)

Prüffrage F2. Die Prüffrage F2 prüft, inwieweit die vom Gutachter eingesetzte medizinisch-technische und testpsychologische Diagnostik zur Beantwortung der Fragestellung ausreichend ist. Wie bei anderen Prüffragen, ergaben sich die Begründungen der festgestellten Mängel sehr selten aus den im Manual hinterlegten Qualitätsaspekten. Neben einigen unkonkreten, der Prüffrage nicht entsprechenden (n = 5) oder indirekten (n = 15) Mängelbegründungen, wurde diese Prüffrage am häufigsten durch das Aufzeigen von Beispielen (n = 151) oder durch einen Verweis auf eine voran gegangene Prüffrage (n = 28) beantwortet (Tab. 51).

Tabelle 51: Codierung der Prüffrage F2.

Prüffrage F2 untersucht, inwieweit die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend ist.		
Code	n	Ankerbeispiel
F2L	5	„psychometrische Testverfahren sind angezeigt“ (48 / zp)
F2D	4	„Nachdem sich der Gutachter entschieden hat als einzigen Test den SKT anzuwenden und weiter ausgeführt hat, dass er den MSST (Mini-Mental-Test) nicht durchgeführt hat, muss man sich fragen, weshalb keine Selbstauskunftsfragebögen, beispielsweise BDI (Depressionsfragebogen) oder Persönlichkeitstests (FPI-R, Gießen-Test) zum Einsatz kamen. Auch hätte man sich im Hinblick auf die leichte Lernbehinderung orientierende kognitive Tests gewünscht, beispielsweise Wortschatztest, c.I.-Test und c.I.-Skala oder Benton-Test. Hierbei handelt es sich um klassische Testverfahren.“ (51 / zw)
F2G	2	„Diagnostik fehlt völlig!“ (131 / zz)
X6	5	„bzgl. der beklagten Luftnot hätte man noch Untersuchungen durchführen können, aber sozialmedizinisch nicht notwendig“ (52 / zm)
X7	8	[Keine Textangabe]
X8	28	„S. E2 (Zusatzgutachten)“ (179 / zi)
X9	3	„Blutdruckmessung die Umstände sind nicht erläutert, keine Kontrolluntersuchung durchgeführt“ (41 / zs)
X10	1	„nach wie vor bin ich der Meinung, dass auch ein Psychiater einen allg.-internistischen Status erheben sollte“ (105 / zt)
X13	1	„Fremdanamnese?“ (48 / zi)
X15	4	„Angaben zum Schweregrad der peripheren arteriellen Verschlusskrankheit fehlen.“ (100 / zu)
X16	151	„Rö-Knie, Blutgase“ (92 / zv)
X18	3	„Unkritische Übernahme somatischer Fremdbefunde“ (179 / zx)
X20	17	„es scheint auch Labor erfolgt zu sein, siehe aktueller TSH-Wert, aber nicht aufgeführt“ (147 / zm)

4.3.6 Qualitative Inhaltsanalyse des übergeordneten Kriteriums

Das übergeordnete Kriterium überprüft die Nachvollziehbarkeit des Gutachtens anhand der vom Gutachter durchgeführten Begutachtungsschritte Anamnese, Diagnostik, Befundung, Epikrise und sozialmedizinische Leistungsbeurteilung. Da im Manual zum Peer Review dem Ampelkriterium keine explizit ausdifferenzierten Mängelkategorien hinterlegt wurden, waren die codierten Textpassagen stark heterogen. Insgesamt wurden 21 verschiedene Codes zur inhaltlichen Verschlagwortung des übergeordneten Kriteriums identifiziert. Im Gegensatz zu den Prüffragen der Einzelkriterien, die sich vorrangig auf spezifische Gutachtenbereiche fokussieren, beschreibt das übergeordnete Kriterium eine globale Beurteilung des Gutachtens anhand der Nachvollziehbarkeit der gutachterlichen Stellungnahme. Infolgedessen ließen die Peers häufig mehrere und unterschiedliche Qualitätsaspekte in ihre Begründung einfließen. Zur Berücksichtigung dieses Beurteilungsverhaltens wurde bei der Codierung des Kriteriums eine Mehrfachcodierung a priori zugelassen. Eine Übersicht und Beschreibung der verwendeten Codes können im *Anhang 8* eingesehen werden. Den insgesamt 191 analysierten Freitextangaben wurden 355 Codes zugeordnet.

Mehr als 50 % der dokumentierten Mängel enthielten genau ein zentrales kodierbares Qualitätsdefizit. In jeweils ca. 1/5 der Reviews wurden gleichzeitig zwei bzw. drei Qualitätsaspekte je Gutachten beanstandet. Nur wenige Reviews (2,2 %) enthielten mehr als vier mehrfachcodierte Mangelkategorien zu Beschreibung der Defizite.

Tabelle 52: Codierung des übergeordneten Kriteriums.

Übergeordnetes Kriterium		untersucht abschließend die „Nachvollziehbarkeit des Gutachtens“ als Ganzes mittels einer Ampelbewertung
Code	n	Ankerbeispiel
üK1	70	„Logischer Bruch zwischen Diagnosenreihenfolge, und Begründung der Leistungsminderung (mit 3-., bzw. 2.-Diagnose)“ (76 / zb)
üK2		„Da die klinische Anamnese im Hinblick auf Beschwerdeschilderung und therapeutische Optionen ungenügend ist und auch der psychische Befund kein nachvollziehbares Bild vom Schweregrad der Erkrankung gibt, muss die Leistungsbeurteilung als unzureichend gewertet werden“ (107 / zd)
üK3	17	„Eigene Untersuchungsergebnisse fließen nicht in differenzierter Darstellung in Epikrise ein, obgleich diese Grundlage des LVs bildet“ (49 / zq)
	23	„Die Befunde unterstützen die Leistungsbeurteilung, es fehlen aber wichtige Befunde“ (92 / zo)
üK4	15	„die subjektiv angegebene Angstsymptomatik, auf die auch der Psychotherapeut hinwies, wurde nicht berücksichtigt bzw. wurde hierzu keine Stellung bezogen oder Bewertung vorgenommen“ (115 / zg)
üK5	26	„Wie dargestellt wird insbesondere in der Epikrise ein Abgleich mit der zuletzt verrichteten Tätigkeit als Maurer im Brückenbau vermisst“ (242 / zd)
üK6	15	„Mobilität hätte stringenter beschrieben werden können, zumal ein Rollator benutzt wird und ein Nachteilsausgleich G (noch) anerkannt ist“ (128 / ze)
üK7	11	„Vor allem die Handfunktion ist in Bezug auf das Erwerbsleben nicht dargestellt“ (193 / zj)
üK8	20	„Insbesondere das negative Leistungsbild in Zus. m. der gestörten Schulterfunktion hätte weitergehend erörtert werden müssen. Auch eine Bewertung der Gefähigkeit wäre bei den gegebenen Gesundheitsstörungen sinnvoll gewesen“ (223 / zc)
üK9	16	„Im Zweifelsfall könnten diese Erkrankungen durchaus auch quantitative Leistungseinschränkungen für Tätigkeiten des allgemeinen Arbeitsmarktes bedingen, was sich aus diesem Gutachten aber eben nicht sicher ableiten lässt.“ (147 / zu)
üK10	5	„offensichtlich steht die Alkoholabhängigkeit im Vordergrund, auch nach der Reihenfolge der Diagnosen, dementsprechend ist eine Entgiftungs-Entwöhnungsbehandlung notwendig.“ (76 / zj)
üK11	15	„Das Gutachten ist überwiegend schlüssig und nachvollziehbar, lediglich bzgl. der Prognose der Erkrankung hätte ich eine abweichende Beurteilung abgegeben“ (49 / zu)
üK12	3	„angeblich Dauerzustand, obwohl Besserung durch Therapie möglich ist“ (282 / zb)
üK13	13	„Wie bereits oben erwähnt, müsste der Prüfarzt m. E. unbedingt ein internistisches Zusatzgutachten anfordern, zum Ausschluss relevanter somatischer Funktionsdefizite“ (48 / zu)
üK14	8	„Gliederungsfolge wurde nicht eingehalten“ (201 / zg)
üK15	13	„Sehr oberflächliches GA. GA hat sich mit Vers. in keinsten Weise auseinandergesetzt!“ (107 / zq)
üK16	5	„Ein knappes Gutachten mit deutlichen Sprüngen die nicht mehr verfolgt werden können“ (242 / zm)
üK17	9	„die internistischen Leiden stehen im Vordergrund!“ (92 / ze)
üK18	5	„S. E2“ (92 / zn)
üK19	21	„Die in den Einzelkriterien dok. Mängel mindern die Schlüssigkeit u. Aussagekraft dieses Gutachtens nicht“ (49 / zc)
üK20	25	
üK21	20	[Keine Textangabe]

Den allgemeinen Beschreibungen des übergeordneten Kriteriums folgend, wurden dokumentierte Qualitätsdefizite am häufigsten mit einem Bruch in der Argumentationskette des Gutachtens begründet (n = 70). Je nach Peer wurde das Qualitätsdefizit mit einem Beispiel hinterlegt oder nicht weiter begründet. Zudem wurden öfter keine, dem Kriterium entsprechende freitextlichen Mängelbegründungen verfasst, sondern ausschließlich auf vorangegangene Prüffragen der Einzelkriterien verwiesen (n = 21). Hierbei wurde am häufigsten ein Rückgriff auf die Prüffrage E2 vorgenommen (n = 16), welche grobe Abweichungen von der Begutachtungsliteratur prüft. Mängel wurden weiterhin mit einer unzureichenden Bewertung von Vorbefunden (n = 26), einer unzureichenden oder fehlenden Darstellung eigener Untersuchungsergebnisse (n = 23) oder einer unzureichenden oder fehlenden Darstellung der Angaben zur quantitativen und qualitativen Leistungsfähigkeit (n = 41) dokumentiert.

Die Analyse zeigte, dass etwa 66 % der dokumentierten Qualitätsdefizite entlang der Argumentationskette eines Gutachtens (vgl. Abb. 6) begründet wurden. Insgesamt 235 der 355 Codes bezogen sich im weitesten Sinne auf ein Element der gutachterlichen Argumentationskette innerhalb eines sozialmedizinischen Gutachtens. Fehler entstanden sowohl bei der eigentlichen Erhebung der Gutachtenbestandteile als auch bei deren Verknüpfung innerhalb des Gutachtens.

Weitere 33 % der beschriebenen Qualitätsdefizite bezogen sich entweder auf Defizite innerhalb eines Gutachtens, wie beispielsweise inhaltliche Widersprüche im Gutachten (n = 13) oder wurden durch den Peer im Review-Verfahrens selbst verursacht, beispielsweise durch die fehlende Dokumentation von Mängeln (n = 20) oder durch die Verwendung irrelevanter Begründungen (n = 5) mit denen die Prüffragen nicht beantwortet werden konnten.

4.3.7 Exemplarische Verknüpfung qualitativer und quantitativer Daten

Da alle Peers die gleichen Gutachten beurteilten, können in der vorliegenden Untersuchung die qualitativen und quantitativen Daten miteinander verknüpft werden. Es wurden zwei Konstellationen identifiziert, welche die Übereinstimmung beeinflussen können:

1. Peers wählen unterschiedliche Mangelkategorien, verwenden jedoch die gleiche Begründung (vgl. Tab. 32),
2. Peers wählen dieselbe Mangelkategorie, verwenden jedoch unterschiedliche freitextliche Begründungen (vgl. Tab. 33).

Während die Peers bei der Beurteilung der Prüffragen der Einzelkriterien auf bestimmte Gutachtenbereiche determiniert sind, beinhaltet die Beurteilung des übergeordneten

Kriteriums ein breiteres Spektrum an Begründungsmöglichkeiten. Dies gewährte den Peers einerseits viele Freiheiten bei der Beurteilung des komplexen Begutachtungsgeschehens, andererseits erhöhte sich dadurch auch die Varianz der Peer-Urteile.

Am Beispiel des übergeordneten Kriteriums wurde am Gutachten #20, das Gutachten mit der schlechtesten prozentualen Peer-Übereinstimmung (29,2 %), die Varianz der Mangeldokumentation exemplarisch dargestellt (Tab. 53).

Tabelle 53: Übersicht über die Qualitätsbewertung, dokumentierte Qualitätsdefizite und den dazugehörigen inhaltsanalytische Codes am Beispielgutachten 1.

Ampel	Freitext	Code
Grün	---	---
Grün	---	---
Grün	---	---
Grün	„Schwach grün. Leistungsbildbeschreibung plausibel“ (48 / zn)	ÜK20
Grün	---	---
Gelb	„Da der Gutachter nur eine rudimentäre klinische Untersuchung durchführt und auch die evtl. in Berichten (Aktualität?) vorliegenden Befunde nicht genannt werden, ist die Argumentation der Befundausprägung (für die untersuchten Bereiche) nachvollziehbar aber lückenhaft“ (48 / zb)	ÜK1, ÜK3 ÜK4
Gelb	„Unzulängliche bzw. fehlende qualitative Leistungsbeurteilung. Dadurch Nachfrage Leistungsabteilung vorprogrammiert. Bezogen auf den allg. Arbeitsmarkt ist die quantitative Leistungsbeurteilung aber nachvollziehbar“ (48 / zc)	ÜK9
Gelb	„Prüfarzt müsste m. E. unbedingt ein internistisches Zusatzgutachten anfordern, zum Ausschluss relevanter somatischer Funktionsdefizite, insbesondere mit Blick auf die Leber-/Lungenfunktion“ (48 / zu)	ÜK14
Gelb	„die mitgeteilte eingeschränkte psychische Belastbarkeit lässt sich auf der Befundebene nicht ableiten, zumal Gutachter von unauffälligem psychischen Befund ausgeht, warum gar keine Stressbelastung möglich sein soll, wurde nicht begründet und wird nicht eindeutig erkennbar, zum Teil fehlen Aussagen zum qualitativen Leistungsbild, Reha inhaltlich befürworten und dann auf dem nächsten Blatt zu verneinen geht nicht (evtl. Ankreuzfehler?)“ (48 / zg)	ÜK1 ÜK9 ÜK13
Gelb	„Da keine wirkliche Argumentationskette im Hinblick auf die qualitative und quantitative Leistungsbeurteilung erfolgt. Ein Nervenarzt kann jedoch in Anbetracht der klinischen Anamnese und der Befundlage die Leistungsbeurteilung im Wesentlichen nachvollziehen“ (48 / zd)	ÜK1
Gelb	„Argumentationskette unterbrochen, da keine Funktionseinschränkungen benannt, Überbrückung aber möglich, Beurteilung nachvollziehbar“ (48 / ii)	ÜK1 ÜK7 ÜK5 ÜK18
Rot	„Defizitäre Beschreibung/Berücksichtigung der Psych- und Bronchialeiden“ (48 / hh)	ÜK3
Rot	„Durch Lungenfunktionsstörung kann Erwerbsminderung bedingt sein, dies wurde aber nicht untersucht“ (48 / zo)	ÜK1 ÜK19
Rot	„Aus Anamnese, Untersuchungsbefund und aus den Unterlagen wird ein undifferenziertes, unzureichend formuliertes Leistungsbild erstellt, angreifbar, nicht nachvollziehbar“ (48 / za)	ÜK4 ÜK5 ÜK9 ÜK10
Rot	„s. F2“ (48 / zr)	
Rot	„bei bestehenden Leiden Leberwerte hoch, Lungenteilentfernung bei unbekannten TU, Palpitationen ohne aktuellem kardiologischen Befund und unzureichenden Sprachkenntnisse und erschwerter Kommunikation mit Dolmetscher, die Befunde als Hypochondrie zu werten ist fragwürdig. Hinzu kommt die mangelnde Beurteilung der qualitativen und quantitativen LF“ (48 / zm)	

Eine sowohl in der gelben als auch roten Gutachtenbewertung wiederkehrende Beurteilung war die allgemeine Aussage zur fehlenden Nachvollziehbarkeit (ÜK1). Hierbei zeigte sich am Beispiel der Peers zd und za, dass eine im Kern gleiche Begründung zu einer unterschiedlichen Beurteilung des Gutachtens geführt hat. Auch zeigte sich innerhalb einer Mangelkategorie eine große Heterogenität bezüglich der Begründungen. So fanden sich bei Mängeln, die von sozialmedizinischen Experten überbrückt werden können (gelb) $n = 7$ und bei Mängeln, die nicht mehr überbrückt werden können (rot) $n = 8$ unterschiedliche Codierungen, welche die gewählte Kategorie begründen sollen.

4.3.8 Kalibrierungsworkshop

Im Nachgang zum Review wurde ein zweitägiger Kalibrierungsworkshop mit allen beteiligten Peers durchgeführt. Die Lageabweichungen innerhalb der Bewertungen sollten aufgelöst sowie ggf. Vorschläge zur Optimierung der Formulierungen der Mangelkategorien und Prüffragen entwickelt werden. Der Kalibrierungsworkshop beinhaltete eine allgemeine Diskussion und Rückmeldungen zur Reviewphase sowie eine Kleingruppenarbeit mit dem Ziel, das Schulungsverfahren für die Peers und/oder Elemente bzw. Formulierungen des Manuals hinsichtlich einer höheren Beurteiler-Übereinstimmung zu optimieren.

Im Hinblick auf das übergeordnete Kriterium kamen die Peers zu dem Ergebnis, dass es sich einerseits um ein sehr hilfreiches, andererseits aber auch sehr komplexes Prüfkriterium handelt. Die Teilnehmer waren sich einig, dass auch eine detaillierte, schriftliche Ausarbeitung im Manual der Komplexität des Ampel-Kriteriums nicht gerecht werden kann. Änderungsvorschläge für das Manual erschienen in Bezug auf das übergeordnete Kriterium dagegen nicht ausreichend und wurden daher auch nicht ausgearbeitet.

Die Peers ($n = 16$) beurteilten die subjektive Relevanz der einzelnen Prüffragen auf einer Ordinalskala von 0 (nicht relevant) bis 10 (sehr relevant). Die Ergebnisse zeigten, dass das übergeordnete Kriterium „Nachvollziehbarkeit“, auch vor dem Hintergrund der limitierten Reliabilitätskennwerte, als wichtigstes Kriterium des Verfahrens angesehen wird (Abb. 16).

Eher formale und inhaltlich unrelevante Kriterien, wie formale Gestaltung (A) und Verständlichkeit (B), wurden tendenziell niedriger beurteilt als das Einzelkriterium Vollständigkeit (D), mit welchem inhaltliche Aspekte des Gutachtens beurteilt wurden.

In den Kleingruppen wurde deutlich, dass durch die Diskussion zwischen den Peers eine Angleichung der Peer-Urteile bei der Ampel-Bewertung durchaus möglich sei. Aus Sicht der Peers bestand zudem Einigkeit darüber, dass ein kontinuierlicher, direkter Austausch notwendig ist, um die bei den vielschichtigen Prüfkriterien bislang bestehenden Bewertungsunterschiede zu überbrücken bzw. zu verringern.

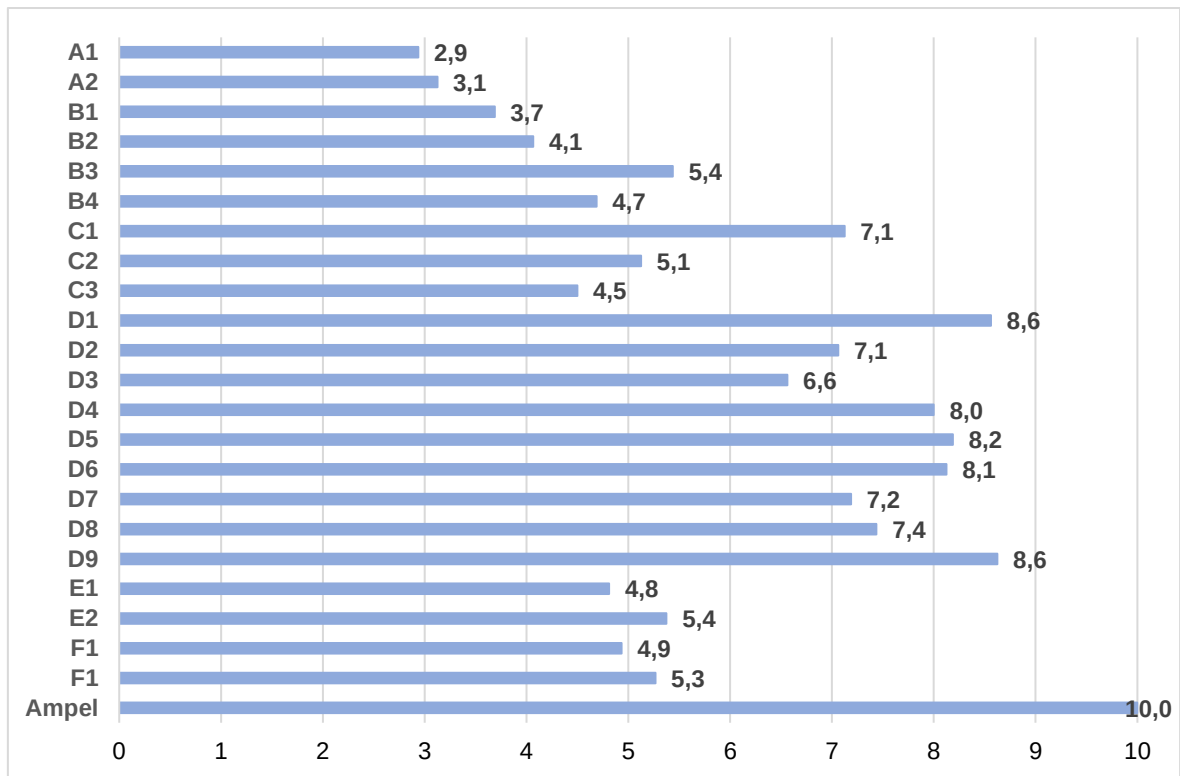


Abbildung 16: Subjektive Beurteilung der Relevanz von Prüffragen und übergeordnetem Kriterium (n = 16).

4.4 Diskussion

4.4.1 Allgemeine Diskussion

4.4.1.1 Reliabilitätsanalyse

Die Teilstudie beschreibt die Evaluation des trägerübergreifenden Peer Review-Verfahrens für die Qualitätssicherung der Sozialmedizinischen Begutachtung bei Anträgen zur EM-Rente. Das Ziel der dargelegten quantitativen Analyse bestand in der Untersuchung der Interrater-Reliabilität. Die ergänzende qualitative Inhaltsanalyse der freitextlichen Mangelbegründungen sollte potenzielle Gründe für eine mögliche Nicht-Übereinstimmung der Peers beschreiben und erste Anhaltspunkte für die Verbesserung des Manuals zum Peer Review oder der initialen Peer-Schulung liefern.

Als valider Evaluationskoeffizient wurde Kendalls **Konkordanzkoeffizient W** definiert, da dieser explizit für die Übereinstimmung mehrerer Rater auf dem Ordinalskalenniveau entwickelt wurde (Wirtz & Caspar, 2002, S. 133). Die Ergebnisse können anhand dieses Koeffizienten direkt mit der methodischen Prüfung des Peer Reviews der Entlassungsberichte im Reha-Bereich verglichen werden. Über viele Jahre hinweg wurde im Peer Review der Entlassungsberichte der Finn-Koeffizient als Reliabilitätsmaß angegeben. Dieser ist ein Übereinstimmungsmaß, welches berechnet wird, indem die beobachtete Varianz der Beurteilungen, durch die bei zufälliger Kodierung zu erwartende Varianz geteilt wird. Nach Subtraktion dieses Quotienten von 1 wird der Anteil „nichtzufälliger“ Varianz berechnet. Dieser Koeffizient ist in der neueren Literatur umstritten (Farin et al., 2003). Problematisch ist vor allem, dass der Koeffizient auf der Varianz der Ratingwerte der einzelnen Objekte basiert und die Varianz bei ordinalskalierten Daten infolge limitiert verfügbarer Streuungsmaße nicht adäquat berechnet werden kann (Wirtz & Caspar, 2002). Dementsprechend überschätzt der Finn-Koeffizient die Höhe der Interrater-Reliabilität (vgl. Farin et al., 2003, S. 326, Tab. 3) und wurde in der vorliegenden Studie nicht herangezogen. Zusätzlich wurde der **Kappa Koeffizient** als Maß für die Übereinstimmung nominalskalierten Daten berechnet. Hintergrund dieser Berechnung beruht in der Frage, ob es sich bei der Mängelabstufung der Einzelkriterien um „echte“ Ordinalskalen handelt, da primär vordefinierte, abgrenzbare, inhaltliche Mangelbegründungen der Beurteilungsskala hinterlegt sind.

Interpretation der Interrater-Reliabilität. Im Allgemeinen wird davon ausgegangen, dass eine Reliabilität ab 0,7 als befriedigend interpretiert werden kann (Farin et al., 2003; Neuderth, 2004; Wirtz & Caspar, 2002). Mit einer Interrater-Reliabilität von $r_w = 0,39$ für das übergeordnete Kriterium, erreicht das Manual zur Qualitätssicherung der Sozialmedizinischen Begutachtung diese Höhe nicht. Es sei allerdings darauf hingewiesen,

dass zahlreiche Aspekte zu nennen sind, die eine geringe Reliabilität erklären können. Die Ausprägung der Koeffizienten wird von bestimmten Störfaktoren tangiert. So bedingt die Definition des zu messenden Merkmals die Erfassbarkeit. Ein breit definiertes Merkmal, wie das übergeordnete Kriterium, welches die Nachvollziehbarkeit der sozialmedizinischen Gutachten misst, ist schwerer zu erfassen als das Merkmal „Vorhandensein von Deck- und Schlussblatt“ (Prüfrage A2). Zusätzlich bedingt auch die Anzahl der zu beurteilenden Skalenkategorien die Höhe der Übereinstimmung. Wenige Skalenkategorien können die Höhe der Koeffizienten verringern. Weitere Einflussfaktoren sind die Beurteilungsmaßstäbe und das heterogene Beurteilerverhalten der Peers selbst (vgl. Abbildung 15). Ferner bedingen die Form der Datenverteilung als auch unbesetzte Zellen in der Datenmatrix die Höhe der Konkordanz-Koeffizienten (Wirtz & Caspar, 2002, S. 59).

Vor dem Hintergrund der o.g. Beeinflussungsfaktoren orientiert sich die Interpretation der berechneten Reliabilitätswerte nicht an den allgemeinen Empfehlungen aus der Methodenlehre, sondern an den in vergleichbaren Kontexten erhobenen Reliabilitätswerten (Farin et al., 2003; Goldman, 1994). Als kritische Orientierungsmarke wurde in der vorliegenden Studie eine Reliabilität von 0,4 definiert. Auch diese Zielmarke konnte das übergeordnete Kriterium als wesentliches Qualitätsmerkmal des Peer Review-Verfahrens nicht überschreiten.

Im Folgenden wird die Reliabilität des Manuals zum Peer Review-Verfahren zur Qualitätssicherung der Sozialmedizinischen Begutachtung (übergeordnetes Kriterium) direkt mit dem Koeffizienten **Kendalls Konkordanzkoeffizient W** aus der methodischen Prüfung der Erhebungsrunde 2000/2001 in den somatischen Indikationen aus dem Peer Review der Entlassungsberichte im Reha-Bereich verglichen (Farin et al., 2003). In der methodischen Prüfung von Farin et al. (2003) schwankt Kendalls Konkordanzkoeffizient W für die Mängelkategorien im Checklistenbereich der sozialmedizinischen Stellungnahme je nach Indikation zwischen 0,15 und 0,60. Da das Peer Review der Sozialmedizinischen Begutachtung als indikationsübergreifendes Verfahren konzipiert wurde, wird an dieser Stelle ausschließlich die Übereinstimmungsreliabilität mit den entsprechenden am Verfahren teilnehmenden Indikationen (Orthopädie $r_w = 0,35$ | Neurologie $r_w = 0,3$) des Peer Reviews der Entlassungsberichte verglichen. Das übergeordnete Kriterium weist gegenüber diesen beiden im Jahr 2003 evaluierten Indikationen im Peer Review der Reha-Entlassungsberichte eine gleichwertige Reliabilität auf. Es sei angemerkt, dass eine direkte Vergleichbarkeit aufgrund der Unterschiedlichkeit der Verfahren nicht möglich ist.

Cohens Kappa erreichte Werte zwischen -0,03 und 0,76, und liegt für die Einzelkriterien im Mittel bei 0,2 (Median: 0,16) und für das übergeordnete Kriterium bei 0,2. Im internationalen Vergleich liegt die Höhe der Koeffizienten damit unter den

Reliabilitätskennzahlen vergleichbarer Studien. So zeigte die Metaanalyse von Goldman (1994) Kappa-Koeffizienten im Bereich um 0,3.

Die Ergebnisse in Tabelle 25 zeigen, dass trotz mäßiger oder guter prozentualer Übereinstimmung die berechnete Interrater-Reliabilität gering ist. Nach Wirtz und Caspar (2002, S. 246) lässt sich dieses Phänomen durch die mangelnde Varianz zwischen den Untersuchungsobjekten erklären. Die Reliabilität kann nur berechnet werden, wenn die ganze Breite der Ratingskalen ausgenutzt wird. Tritt bei den Reviewobjekten ein Deckeneffekt auf, durch den die Peers eine Mangelkategorie präferieren, so entsteht in der Datenmatrix eine Konstante, die nicht für die Berechnung der Reliabilität herangezogen werden kann. Die dadurch entstehenden ungleich verteilten Randsummen führen folglich dazu, dass die Reliabilität gering ausfällt (Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990). Wirtz und Caspar (2002, S. 246) empfehlen zur Erhöhung der Varianz unähnliche Review-Objekte in die Untersuchung zu integrieren. Diese Empfehlung ist im Rahmen der Qualitätssicherung anonym gezogener Gutachten nicht umzusetzen. Wenn das Spektrum der ordinal skalierten Ratingskala nicht vollständig ausgenutzt wird, sollten stattdessen die Abstufungen mit Ankerbeispielen beschrieben werden. Da den Kategorien der Einzelkriterien bereits prüffragenspezifischen Mängeldefinitionen hinterlegt wurden, könnte zur Erhöhung der Reliabilität nur noch deren Präzisierung im Rahmen einer Manualüberarbeitung beitragen.

Indikationsspezifische Auswertungen. Das Peer Review-Verfahren wurde im Sinne einer indikationsübergreifenden Qualitätssicherungsmethode entwickelt. Ein Sozialmediziner sollte in der Lage sein, sozialmedizinische Gutachten jeder medizinischen Fachrichtung zu lesen, zu verstehen und zu bewerten. Die qualitative Inhaltsanalyse bestätigt, dass nur in sehr seltenen Fällen angegeben wurde ($n = 3$) über keine entsprechenden Kenntnisse der Fachliteratur zu verfügen (Prüffrage E2).

In der vorliegenden Arbeit wurde auf eine indikationsspezifische Auswertung verzichtet, da diese bereits im Rahmen der Projektberichterstattung von Strahl und Kollegen (Strahl, Gerlich, Wolf & Vogel, 2012) am Gesamtdatensatz vorgenommen wurde. Die Auswertung untersuchte die Reliabilität, wenn die fachspezifischen Gutachten durch Peers mit gleicher Facharztausbildung beurteilt wurden. Die Koeffizienten zeigten ähnliche Kennwerte wie bei der indikationsübergreifenden Reliabilitätsberechnung:

- Orthopädie Einzelkriterien: $\bar{X}_{\text{Kendalls } W} = 0,38$; $\text{Median}_{\text{Kendalls } W} = 0,35$
 - Orthopädie übergeordnetes Kriterium: $r_{\text{Kendalls } W} = 0,18$
 - Innere/Allgemeinmedizin. Einzelkriterien: $\bar{X}_{\text{Kendalls } W} = 0,3$; $\text{Median}_{\text{Kendalls } W} = 0,26$
 - Innere/Allgemeinmedizin übergeordnetes Kriterium: $r_{\text{Kendalls } W} = 0,45$
 - Neurologie/Psychiatrie Einzelkriterien: $\bar{X}_{\text{Kendalls } W} = 0,36$; $\text{Median}_{\text{Kendalls } W} = 0,38$
 - Neurologie/Psychiatrie übergeordnetes Kriterium: $r_{\text{Kendalls } W} = 0,35$
-
- Indikationsübergreifend Einzelkriterien: $\bar{X}_{\text{Kendalls } W} = 0,28$; $\text{Median}_{\text{Kendalls } W} = 0,28$
 - Indikationsübergreifend übergeordnetes Kriterium: $r_{\text{Kendalls } W} = 0,39$

Die Auswahl der einzelnen Indikationen reduzierte jedoch stark die Fälle, die in die Analyse eingingen, sodass einzelne Koeffizienten nicht berechnet werden konnten oder auf nur sehr wenigen Fällen beruhten. Eine Interpretation war daher nur bedingt möglich. Beim Vergleich des übergeordneten Kriteriums kann dennoch festgehalten werden, dass die Orthopädie einen geringeren und die Innere Medizin/Allgemeinmedizin einen leicht höheren Reliabilitätswert als die indikationsübergreifende Auswertung aufweist.

Sowohl für weitere Analysen, als auch für die Umsetzung in der Routine wurde eine indikationsspezifische Auswertung infolge der kleinen Fallmenge die letztlich in die Reliabilitätsberechnung eingehen, nicht empfohlen (Strahl et al., 2012, S. 42).

4.4.1.2 Qualitative Inhaltsanalyse

Die qualitative Inhaltsanalyse sollte einen Beitrag dazu leisten, den Review-Prozess der Peer-Gutachter besser zu verstehen und das Gesamtverfahren weiterzuentwickeln. Ziel der Analyse war es, aus den Ergebnisdaten der Pilotphase des Projektes EvaBegut die freitextlichen Begründungen zu den Mängelbewertungen der Peers zu systematisieren, um auf dieser Grundlage typische Qualitätsproblemstellungen identifizieren und charakterisieren zu können. Dazu wurden die Begründungen der Qualitätsbewertungen qualitativ-inhaltsanalytisch ausgewertet, d.h. zu den einzelnen Textteilen (Begriffen, Sätzen, ...) wurde zusätzliches Material herangetragen, das das Verständnis erweitert bzw. die Textstelle erläutert, erklärt oder ausdeutet. Mayring nennt dies die induktive Vorgehensweise der Codegenerierung (Mayring, 2003; Mayring & Fenzl, 2014). Im Ergebnis führte dieses Vorgehen dazu, 77 prüffragenspezifische Codes und 20 prüffragenübergreifende Codes aus den vorliegenden freitextlichen Mängelbegründungen der Einzelkriterien ermitteln.

Um die Gutachten im Review-Prozess im Sinne des Manuals zu beurteilen, wurde erwartet, dass die freitextlichen Mängelbegründungen hauptsächlich den prüffragenspezifischen Codes, welche die einzelnen Mängelabstufungen repräsentieren, entsprechen. Bei der Dokumentation der gewählten Mängelkategorien wurden jedoch auch Begründungen abgegeben, die über die Mängelkategorien hinausgehen. Diese lassen sich einteilen in:

- prüffragenübergreifende Codes, die neue Gesichtspunkte und inhaltliche Aspekte beinhalten,
- prüffragenübergreifende Codes, die keine Beantwortung der Prüffrage zulassen, die darauf zurückzuführen sind, dass der Peer bei der Anwendung des Manuals einen Fehler begeht,
- prüffragenübergreifende Codes, welche die Fragestellung der Prüffragen nur unvollständig beantworten, da die inhaltliche Differenzierung der Antworten zu hoch oder zu niedrig sind,
- prüffragenübergreifende Codes, die lediglich indirekt eine Beantwortung der Prüffragen zulassen,
- prüffragenübergreifende Codes, die eine allgemeine Kritik am Manual, am Peer Review-Verfahren oder am Verfasser des zu beurteilenden Gutachters üben sowie
- prüffragenübergreifende Codes, die vergeben wurden, wenn ein Peer eine Mängeldokumentation vornimmt, obwohl kein Mangel in der entsprechenden Prüffrage angegeben wurde.

Mit Hilfe der qualitativen Inhaltsanalyse konnte gezeigt werden, dass die Peers bei der Beurteilung der Prüffragen mehr Qualitätsaspekte einschlossen, als vom Manual formal erwartet wurde. Zudem fanden sich auch Textstellen, die Anmerkungen, Hinweise und Kritik enthielten. Insgesamt lassen die Ergebnisse darauf schließen, dass Peer-Gutachter bei der Begründung ihrer Beurteilung nicht nur die im Manual hinterlegte Qualitätsaspekte adressieren.

Sofern die entsprechenden Mängelkategorien von den Peers manualkonform angewendet wurden, ist davon auszugehen, dass die Einzelkriterien die typischen Qualitätsdefizite der Sozialmedizinischen Begutachtung zufriedenstellend abbilden. Die freitextlichen Mängelbegründungen der Peers belegen, dass bei der Beurteilung häufig auf die prüffragenspezifischen Kategorien zurückgegriffen wurde.

Im übergeordneten Kriterium wurden Mängel in der Nachvollziehbarkeit des Gutachtens am häufigsten direkt kodiert und ggf. mit Beispielen hinterlegt (Code ÜK1, n = 70). Das Ampelkriterium bietet den beurteilenden Peers die Qualitätsdefizite eines Gutachtens in einer übergeordneten Zusammenschau zu beschreiben. Daher wurden für dieses Kriterium insgesamt 21 verschiedene Codes identifiziert. Die inhaltlichen Freiheiten des Peers bei der freitextlichen Begründung der Qualitätsdefizite erhöht jedoch gleichzeitig den Peer-Bias, da die Varianz der möglichen Beurteilungsoptionen steigt. Es ist anzunehmen, dass je nach Blickwinkel, Schwerpunktsetzung, fachlichem Hintergrundwissen und Erfahrung ein anderer Fokus auf Gutachten gelegt wird. Dies könnte sich direkt auf die Beurteilung des Kriteriums niederschlagen und zu nicht übereinstimmenden Peer-Urteilen führen.

4.4.1.3 Limitationen der Studie

Aus der Literatur sind Varianzquellen bekannt, die zu einer niedrigen Reliabilität führen können. Dies sind unter anderem ein systematischer individueller Peer-Bias oder ein systematischer Verfahrens-Bias, der sich aus der Schulung der Peers oder dem Verfahren selbst bezieht. Um Fehlerquellen zu minimieren, wird in der Literatur empfohlen, das Review-Instrument fachlich fundiert und standardisiert zu entwickeln sowie die am Peer Review partizipierten Teilnehmer mittels einer Schulung zu qualifizieren (Harris & Bratzler, 2013; Smith et al., 1997; Tuijn et al., 2012). Die individuellen Peer-Unterscheide wurden in der vorliegenden Untersuchung nachgewiesen. Die Kalibrierung der Beurteilermaßstäbe sollte durch die Schulung der Peers zu Beginn des Projektes gewährleistet werden. Die Ergebnisse belegen jedoch, dass die Schulung diesbezüglich keinen Effekt gezeigt hat. Da jedoch kein Studiendesign mit zwei Messzeitpunkten realisiert werden konnte, ist der tatsächliche Effekt der Schulung mit dem vorliegenden Studiendesign nicht evaluierbar. Die durchgeführte Schulung war weder standardisiert noch manualisiert und basierte auf keinem ausgearbeiteten Schulungscurriculum.

Die qualitative Inhaltsanalyse hat primär reduktiven Charakter und hatte nicht den Anspruch, im Sinne der „grounded theory“ neue Theorien zu erarbeiten. Folglich eignet sich die Analyse ausschließlich dazu, die Freitexte zu beschreiben (Froggatt, 2001). Die Kodierung wurde durch zwei Rater unmittelbar am Ausgangsmaterial im informellen Konsensverfahren durchgeführt. Es wurden keine raterspezifischen Aufzeichnungen der Codierung vorgenommen, infolgedessen die Intercoder-Reliabilität nicht abgebildet werden kann.

Zum organisatorischen Ablauf des Review-Verfahrens sei anzumerken, dass die Review-Phase einen Zeitraum von insgesamt fünf Monaten einnahm. Da die Bearbeitung der sozialmedizinischen Gutachten seitens der Peers berufsbegleitend vorgenommen wurde, ist nicht auszuschließen, dass einige Peers einen durch längere Pausen zwischen den einzelnen Reviews die Handhabung des Prüffragenkatalogs nicht ausreichend verstetigen konnten. Peers, welche die Möglichkeiten hatten, die Reviews geblockt im Anschluss der Schulung zu erstellen, profitieren voraussichtlich mehr von den Schulungsinhalten und verstetigen den Übungseffekt im Umgang mit den Prüffragen.

4.4.2 Beantwortung der Fragestellung

Die Evaluation des Peer Review-Verfahrens zur Qualitätssicherung der Sozialmedizinischen Begutachtung bei EM-Gutachten untersuchte die Fragestellungen, die in Tabelle 20 zusammengetragen wurden.

1. Wie hoch ist die Interrater-Reliabilität des Peer Review-Verfahrens zur Qualitätssicherung in der Sozialmedizinischen Begutachtung?

Bei der Evaluation des Prüffragenkatalogs muss zwischen den Prüffragen der Einzelkriterien und dem übergeordneten Kriterium unterschieden werden. Das übergeordnete Kriterium „Nachvollziehbarkeit der Begutachtung“ gilt als K.O.-Kriterium des Qualitätssicherungsverfahrens und entspricht dem primären Zielkriterium der Evaluation. In der Analyse muss auch zwischen Interrater-Übereinstimmung und Interrater-Reliabilität differenziert werden. Dabei sind der Reliabilitätskoeffizient Kendalls Konkordanzkoeffizient W für mehrere Rater und die prozentuale Übereinstimmung zentrale Maße der Konkordanz. Die prozentuale Übereinstimmung des übergeordneten Kriteriums beträgt 47,9 % und die Spannweite der Einzelkriterien liegt zwischen 40 % und 89 %. Während die Übereinstimmung akzeptable Werte um die 50 % aufweist, scheint die Interrater-Reliabilität von Kendalls W limitiert. Die Reliabilität beträgt $r_W = 0,39$. Damit wurde die als kritische Orientierungsmarke definierte Reliabilität von 0,4 knapp verfehlt. Im Vergleich zum bereits publizierten Datensatz zeigt sich, dass die annähernd gleiche Reliabilität bezüglich des übergeordneten Kriteriums erzielt wurde (vgl. Strahl et al., 2016).

2. Welche freitextlichen Mängelbegründungen werden genutzt, um detektierte Gutachtendefizite zu begründen?

Je nach Prüffrage wurden gutachtenspezifische und gutachtenübergreifende Mängelbegründungen detektiert (vgl. Tab 28 bis 51). Durch die Anwendung deduktiver und induktiver Verfahrensweisen der qualitativen Inhaltsanalyse wurden für jede Prüffrage charakteristische Begründungen identifiziert, codiert und quantifiziert. Die Ergebnisse belegen, dass die Peers sowohl die vom Manual vorgegebenen Qualitätsdefizite zur Dokumentation von Mängeln nutzen als auch Qualitätsaspekte für die Beurteilung heranziehen, die formal vom Manual nicht vorgesehen waren. Für das übergeordnete Kriterium wurden induktiv 21 Codes verschlagwortet, mit denen die Qualitätsdefizite begründet werden können. Die Codes sind zu 66 % dem theoretischen Modell der Begutachtungs- und Argumentationskette der Sozialmedizinischen Begutachtung zuzuordnen. In der überwiegenden Anzahl der Fälle wird das übergeordnete Kriterium entsprechend im Sinne des Manuals bearbeitet. Berücksichtigt man die Mehrfachkodierungen, ist davon auszugehen, dass mehr als $\frac{2}{3}$ aller beurteilten Gutachten manualkonform beantwortet werden.

3. Können typische Qualitätsproblemstellungen in den freitextlichen Mängelbegründungen identifiziert und systematisiert werden?

Anhand der Triangulation der Daten ließen sich zwei Peerurteilskonstellationen identifizieren, die theoretisch zu erwarten sind: Peers wählen unterschiedliche Mangelkategorien, verwenden jedoch gleiche Begründungen oder Peers wählen dieselbe Mangelkategorie, verwenden jedoch unterschiedliche Freitextbegründungen. Langfristig reduziert dies die Reliabilität des Verfahrens. Die Inkongruenz konnte sowohl innerhalb der Prüffragen der Einzelkriterien, als auch im übergeordneten Kriterium beobachtet werden.

Für die Darstellung der Problemlagen bezüglich der Einzelkriterien können primär die Code-Kategorien der prüffragenübergreifenden Codes herangezogen werden. So zeigt sich, dass einzelne Prüffragen mit Mangelkategorien, deren Bewertungscharakteristik auf einer mengenmäßigen Steigerung beruht, nicht immer eindeutig von den Peers differenziert werden können. Vor allem die Differenzierung zwischen deutlichen („mehrfach“) und gravierenden („überwiegend“) Mängeln erscheint schwierig und bedarf einer weitergehenden Standardisierung. Folglich kann bei diesen Prüffragen festgestellt werden, dass gleiche Begründungen bei unterschiedlicher Mangelvergabe dokumentiert wurden.

Die Problematik zeigt sich ebenfalls bei Prüffragen, deren Bewertungscharakteristik auf der Abstufung der sozialmedizinischen Relevanz beruht (z. B. Prüffrage C1 bis C3). Die Dokumentation der Qualitätsdefizite vollzieht sich bei diesen Prüffragen vorrangig anhand von Beispielen bzw. Aussagen, bei denen der Bezug zur sozialmedizinischen Relevanz nicht eindeutig beschrieben wird. Die Begründung wird nur in Verbindung mit der gewählten Mangelkategorie eindeutig zuordenbar. Dies zeigt sich deutlich im Einzelkriterium C: Transparenz. Die Mängelabstufungen beruhen auf der Unterscheidung zwischen zu beurteilenden sozialmedizinisch relevanten und nicht relevanten inhaltlichen Aspekten. Häufig (z. B. C1; n = 133) wurden diese Prüffragen mit Beispielen beantwortet, bei denen ein konkreter Verweis auf die Graduierungsdimension fehlt (Code X5). Die Mangelbegründung war damit unzureichend, da nicht eindeutig ersichtlich wurde, ob der kritisierte Sachverhalt sozialmedizinisch relevant ist.

Eine weitere Problematik ist die häufige Verwendung von Verweisen auf bereits beantwortete Prüffragen (z. B.: „siehe. D5“ oder „s.o.“). Verweise sollten nur dann verwendet werden, wenn diese mit weiteren Erläuterungen mit Bezug zur aktuellen Prüffrage versehen werden. Vor allem bei der Mängelbeschreibung des übergeordneten Kriteriums ist ein Verweis auf eine vorherige Prüffrage nicht ausreichend und wird der Komplexität des Kriteriums nicht gerecht. Der Verweis allein ist nicht ausreichend, um eine Prüffrage zu beantworten. Auch eine pauschale Mängelbegründung dient nicht dazu, den festgestellten Mangel adäquat zu dokumentieren.

Unvollständig und indirekte Beantwortung der Prüffragen. Prüffragen werden manchmal nur unvollständig beantwortet oder angesprochene Qualitätsdefizite müssen erst aus den Peer-Angaben erschlossen werden. Vereinzelt kommt es seitens der Peers vor, dass Prüffragen in einem zu hohen ($n = 39$) oder zu niedrigen ($n = 75$) Detaillierungsgrad beantwortet wurden. Eine zu hohe Differenzierung beinhaltet in der Regel unterschiedliche dokumentierte Aspekte, infolgedessen als Mängelbegründung zu viele Qualitätsdimensionen des Gutachtens angesprochen werden und kein direkter Rückschluss auf die Prüffrage gezogen werden kann. Andererseits beinhaltet ein zu niedriger Differenzierungsgrad eine zu pauschale Beantwortung der Prüffrage, sodass sich die Inhaltlichen Vorgaben der jeweiligen Mängelkategorie nicht in den Antworten der Peers widerspiegeln. Hierbei wird der Mangel nicht punktgenau und ohne Beispiel beschrieben, so dass z.T. lediglich die bereits durch das Manual vorgegeben Mängelkategorien ohne Bezug zum vorliegenden Gutachten wiederholt werden.

Kritik am Gutachten und am Verfahren. Es wurde in den Freitextbegründungen auch Kritik bzgl. des Vorgehens des Gutachters bzw. am Gutachten direkt geäußert, anstatt die Prüffrage zu beantworten. Neben Kritik am Gutachten finden sich auch vereinzelt kritische Anmerkungen zu einzelnen Prüffragen oder zum Verfahren des Peer Reviews selbst. So wurde mehrfach erwähnt, dass Informationen des Gutachtens geschwärzt worden sind, die eigentlich zur Beantwortung einer oder mehrerer Prüffragen benötigt werden würde (z. B. Facharztbezeichnung).

4.5 Handlungsempfehlungen

4.5.1 Empfehlungen für das Manual zum Peer Review

Die Ergebnisse weisen darauf hin, dass einzelne Prüffragen bzw. Mängeldefinitionen konnotativ unterschiedlich angewendet wurden. Es wird daher empfohlen, die Schulung für die Peertätigkeit zu intensivieren und das Manual zu überarbeiten und regelmäßig weiterzuentwickeln. Durch den Kalibrierungsworkshop wurde jedoch deutlich, dass die Konkretisierung von Fragen und Bewertungsschemata Grenzen hat, die u.a. im Grundverständnis der sozialmedizinisch tätigen Ärzte sowie in den etablierten sozialmedizinischen Arbeitsweisen begründet liegen, die nur durch einen anhaltenden Austauschprozess konvergieren können.

Erweiterung der Mängelkategorien. Die Mängelkategorien der Einzelkriterien erschienen in der praktischen Anwendung nicht immer hinreichend trennscharf. Die bislang konsequent angewendete vierstufige Rating-Skala bei den Einzelkriterien sollte nach Meinung der Peers bei bestimmten Prüffragen aufgehoben werden, indem zwei Mängelkategorien zusammengefasst werden. Ferner wurde auch der Wunsch geäußert, bei einigen Prüffragen eine zusätzliche Antwortkategorie „trifft nicht zu“ anzufügen. Um jedoch zu vermeiden, dass bei schwierigen Entscheidungsprozessen diese Kategorie lediglich aus ökonomischen Gesichtspunkten angekreuzt wird, sollte sie nur gezielt bei Bedarf vorgegeben und mit einer verpflichtenden freitextlichen Begründung verbunden sein.

Überarbeitung der Prüffragen. Hinsichtlich der Prüffragen, die sich ausschließlich auf Teilbereiche des Gutachtens beziehen, wird empfohlen, diese für die Peers deutlich im Manual zu kennzeichnen. Außerdem wird angeraten, häufiger Beispiele in den Prüffragen aufzuzeigen. Die konkreten Optimierungsvorschläge für die Zusammenfassung einzelner Mängelkategorien, die Umformulierung einzelner Prüffragen sowie die Implementierung der Antwortmodi wurden nach Projektende der DRV Bund übermittelt und können auf Anfrage beim Autor der vorliegenden Arbeit eingesehen werden.

Hinweise für die Anonymisierung. Die Freitextanmerkungen beinhalteten auch Kritik bezüglich der Anonymisierungspraxis der Gutachten. Es wird daher empfohlen, einen standardisierten Kriterienkatalog zu erstellen, der Informationen darüber enthält, welche Informationen geschwärzt werden sollen, damit in der Routine alle 16 Rentenversicherungsträger demselben modus operandi folgen.

4.5.2 Empfehlung für die Peer-Schulung

Ausarbeitung einer standardisierten Peer-Schulung. Um im Peer Review-Verfahren Fehlerquellen im Umgang mit den Prüffragen zu minimieren, ist es notwendig, die Peers mit dem Manual vertraut zu machen. Dazu gehört neben der Kenntnis von Aufbau und Anwendung der Prüffragen auch die Kenntnis über die der Prüffrage hinterlegten Bewertungscharakteristika. Die Ausarbeitung einer standardisierten Peer-Schulung könnte einen Beitrag dazu leisten, den systematischen Bias bei der Schulung der teilnehmenden Sozialmediziner zu reduzieren (vgl. Tuijn et al., 2012). Die niedrige Reliabilität und die erhöhte interindividuelle Varianz der Peers könnten voraussichtlich durch eine initiale Schulung und die regelmäßige Teilnahme an einer standardisierten Peer-Schulung optimiert werden. Damit jeder Peer im Rahmen der Qualitätssicherung gleiche Beurteilungsstandards anwendet, müssen alle Teilnehmer einheitlich geschult und fortgebildet werden. Zur Standardisierung muss ein ausformuliertes Curriculum mit konkreten Umsetzungshinweisen und festgelegten Methoden erarbeitet werden. Hierbei soll das Ziel der Schulung sowohl der **Wissenserwerb** als auch das **Training von Fertigkeiten** in der Methode des Peer Review-Verfahrens zur Qualitätssicherung der Sozialmedizinischen Begutachtung sein.

Der **Erwerb von Wissen** über den Hintergrund des Qualitätssicherungsverfahrens, den Aufbau des Manuals und die Aufgaben als Peer sind notwendig, um die Funktion als Peer im Peer Review ganzheitlich wahrzunehmen. Das Wissen fördert den Perspektivwechsel vom gutachterlich tätigen Sozialmediziner hin zum Peer im Review-Verfahren. Dies ist notwendig, da die Peerfunktion mit einer veränderten Herangehensweise an die Sichtung und das Lesen von Gutachten einher geht. Während Sozialmediziner in der Routine beispielsweise häufig die zusammenfassende Epikrise eines Gutachtens zuerst lesen, ist es im Review-Prozess vorgesehen, dass Gutachten von der ersten bis zur letzten Seite zu lesen. Zudem erlangen die Peers Wissen darüber, dass die Mängelkategorien unterschiedlichen Charakteristiken unterliegen und dementsprechend Begründungen eines Qualitätsdefizits individuell formuliert werden müssen. Zusätzlich sind der Erwerb und das **Training von praktischen Fertigkeiten** im Umgang mit den Prüffragen erforderliche Grundvoraussetzungen zur richtigen und einheitlichen Gutachtenbeurteilung. Dazu müssen im geschützten Rahmen der Peer-Schulung möglichst viele Gutachtenfälle bewertet und in einer Kleingruppe diskutiert werden. Da die Sozialmedizinische Begutachtung eine höchst komplexe und fallspezifische Leistungsbeurteilung darstellt, sollten die Peers viele unterschiedliche Fallkonstellationen aus den unterschiedlichen Indikationen kennen lernen. Infolgedessen ist es notwendig, die Schulung als regelmäßige Fortbildungsmaßnahme für

bereits aktive Peers auszugestalten. Dadurch wird einerseits der Erfahrungsschatz der Peers erweitert und andererseits deren Beurteilungsmaßstäbe regelmäßig kalibriert.

Entsprechend dieser inhaltlichen Anforderungen sollte die Schulung in ihrer Ausgestaltung mindestens zwei volle Tage umfassen und einen workshopartigen Charakter haben, der es erlaubt, in Kleingruppen mit interaktiven Methoden zu arbeiten. Die Schulung soll dem theoretischen Konzept für gute Unterrichtsqualität entsprechen, welches eine klare Schulungsstruktur, ein teilnehmerorientiertes Sozialklima sowie eine kognitive Aktivierung der Teilnehmer vorsieht (Feicke & Spörhase, 2012).

Schulungsinhalt: Aufbau der Prüffragen. Bei der Auswertung der qualitativen Inhaltsanalyse zeigten sich einige Verstöße gegen die im Manual festgelegten Einschränkungen. So bedarf es bspw. bei einigen Prüffragen lediglich der Beurteilung von sozialmedizinischen Sachverhalten in bestimmten Abschnitten des Gutachtens, wie Diagnosen und Epikrise. Auch formulieren einige Prüffragen in ihren Begleittexten, dass bestimmte Sachverhalte nicht beurteilt werden sollen (z. B. Prüffrage A1: In Abgrenzung zu den Prüffragen des Einzelkriteriums **"Vollständigkeit"** geht es hier um das rein formale Vorhandensein der Gliederungspunkte. Inhaltliche Aspekte, wie z. B. Art, Umfang, Angemessenheit oder Vollständigkeit der Gliederungspunkte sind hier nicht Gegenstand der Bewertung). Für die Ausarbeitung einer standardisierten Schulung wird daher empfohlen, im Rahmen eines eigenständigen Lernzieles diese Besonderheiten aufzuzeigen.

Schulungsinhalt: freitextliche Mängeldokumentation. Um die Probleme und Defizite bei der Bearbeitung der Gutachten zu beseitigen, sollte in einer standardisierten Schulung explizit darauf hingewiesen werden, keine Verweise auf andere Prüffragen (z. B. siehe D5) als Begründung zu verwenden und stets eine Mängelbegründung für ein dokumentiertes Qualitätsdefizit zu formulieren. Prüffragen aus dem Einzelkriterium **„Transparenz“**, deren Bewertungsmatrix für die Graduierung der Mängelstufen eine Strukturierungsdimension zugrunde legen (sozialmedizinisch relevant vs. nicht relevant), sollten speziell geschult werden, sodass bei der freitextlichen Begründung im Review Beispiele nicht als alleinige Begründung Verwendung finden, sondern mit einem begleitenden Text verfasst werden.

4.6 Kernbotschaft

Die Güte des Verfahrens (Interrater-Übereinstimmung und -Reliabilität) wurde unter der Perspektive des spezifischen Beurteilungskontextes bewertet. Die Koeffizienten erreichen den definierten Richtwert nicht, liegen jedoch im Rahmen existierender Forschungsergebnisse und sind mit der Güte des Qualitätssicherungsverfahrens der Reha-Entlassungsberichte vergleichbar. Das Qualitätssicherungsverfahren im Bereich der Sozialmedizinischen Begutachtung erwies sich in der Evaluation als sachgerecht und umsetzbar. Daher kann das Review nach Überarbeitung zur Einführung einer trägerübergreifenden Qualitätssicherung sozialmedizinischer Gutachten empfohlen werden. Gleichzeitig sollte das Potenzial des Verfahrens kontinuierlich ausgeschöpft werden. Neben einer regelmäßigen Überarbeitung des Prüffragenkatalogs scheinen insbesondere eine standardisierte initiale Peer-Schulung und ein zusätzlicher konsekutiver regelmäßiger Schulungsworkshop, bei dem mit konkreten Gutachtenfällen gearbeitet wird, geeignet, die Güte des Verfahrens langfristig zu verstetigen.

5 Teilstudie 3: Entwicklung und Evaluation einer Peer-Schulung

Die dritte Teilstudie umfasst die Entwicklung und Evaluation eines strukturierten Peer-Schulungsprogramms für das trägerübergreifende Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung der DRV. Ziel der Untersuchung war die Erstellung einer standardisierten Peer-Schulung, inkl. Schulungsunterlagen, Umsetzungsempfehlungen, organisatorischen Rahmenbedingungen sowie die Erprobung und Evaluation des Schulungsprogramms in einem kleinen Peer Review-Verfahren.

Die Ergebnisse der vorangegangenen Reliabilitätsanalyse beschreiben einen individuellen Peer-Bias bei der Bearbeitung der Reviews sowie einen systematischen Bias, der aus dem Verfahren bzw. der Schulung selbst generiert wurde. Die Übereinstimmung von Peers ist im Allgemeinen häufig limitiert und benötigt Verbesserung (Smith et al., 1997). Um das Qualitätssicherungsverfahren sozialmedizinischer Gutachten trägerübergreifend in die Routine zu implementieren, bedarf es neben einer hohen methodischen Güte des Verfahrens, vor allem einheitlicher Bewertungsmaßstäbe seitens der Peers. Die hohen Anforderungen, einheitliche Bewertungsmaßstäbe bei der Prüfung der Gutachten anzulegen, verlangen eine fundierte Einführung der Gutachter in die Methoden des Peer Review-Verfahrens. Aus der zweiten Teilstudie wurden Anregung zur Überarbeitung des Manuals sowie zur Erarbeitung einer Peer-Schulung aufgegriffen und im vorliegenden Projekt umgesetzt.

Um die Interrater-Variabilität zu reduzieren und infolgedessen die Reliabilität zu erhöhen, wird nach Tuijn et al. (2012) ein möglichst standardisiertes Peer Review-Instrument sowie ein standardisiertes Peer Training benötigt. Die Schulung soll den Peers die Methodik des Review-Verfahrens vermitteln sowie anhand von ausgewählten Beispielgutachten den Umgang mit dem Manual zum Peer Review-Verfahren lehren. Mit der Schulung sollen drei zentrale Ziele erreicht werden:

1. Die teilnehmenden Peers sollen das Bewertungssystem der Einzelkriterien und des übergeordneten Kriteriums im Manual zum Peer Review kennen und verstehen.
2. Die teilnehmenden Peers sollen die Anwendung des Bewertungssystem anhand von Beispielgutachten handhaben können
3. Die Peers sollen bei der Anwendung des Bewertungssystems einheitliche Bewertungsmaßstäbe kalibrieren und in individuellen Gutachtenkonstellationen die Prüffragen einheitlich und unabhängig voneinander bearbeiten.

Hierzu soll die Schulung vorrangig in Form von praktischen Übungen und Kleingruppenarbeiten realisiert und durchgeführt werden. Das heterogene Raterverhalten, welches Einfluss auf die Höhe der Konkordanz hat, gibt eindeutige Hinweise darauf, dass vor dem Einsatz in der Routine eine systematische Peer-Schulung erstellt und evaluiert werden muss, um die Qualitätsbeurteilungen der Peers einander anzugleichen.

Eine Meta-Analyse zur Analyse von Maßnahmen zur Erhöhung der Interrater-Reliabilität bei Angehörigen von Gesundheitsberufen belegt, dass Verbesserungen des eingesetzten Review-Instrumentes, Verbesserungen der Peer-Schulung sowie die Kombination aus beiden signifikante Auswirkung auf die Interrater-Reliabilität haben. Alle drei Interventionsarten werden als wirksam beschrieben, die Bearbeitung des Review-Instruments sei jedoch die effektivste Maßnahme zur Erhöhung der Reliabilität (Tuijn et al., 2012). Zudem konnte im Rahmen eines strukturierten impliziten Reviews gezeigt werden, dass eine Diskussion zwischen zwei Ärzten die Übereinstimmung in einem erheblichen Maße verbessern kann. Die allgemeine Reliabilität über mehrere an der Studie teilnehmenden Ärzte wurde jedoch nicht signifikant beeinflusst, da die Beurteilung ausschließlich in Peer-Paaren diskutiert wurde. Zwischen den Paaren war keine Steigerung der Reliabilität zu verzeichnen (Hofer, Bernstein, DeMonner & Hayward, 2000). Die Ergebnisse deuten darauf hin, dass die direkte Kommunikation über die Review-Objekte anhand eines Review-Instruments zu einem gemeinsamen Konsensurteil führt.

Erste Daten zur Evaluation der Peer-Schulung wurden in Teilen bereits in der Zeitschrift „BMC Medical Education“ publiziert (Strahl et al., 2018). Im Gegensatz zu den veröffentlichten Studiendaten wird die Konzeption der Peer-Schulung und des dazugehörigen Schulungscurriculums in der vorliegenden Arbeit detaillierter beschrieben. Darüber hinaus werden zusätzliche Daten zur Reliabilität und zur Schulungsevaluation vorgelegt, die neu berechnet und bisher nicht publiziert wurden.

5.1 Fragestellung

Ziel der Untersuchung ist die Erstellung eines strukturierten Peer-Schulungsprogramms zum Wissenserwerb als auch zum Training von Fähig- und Fertigkeiten für aktive Peers im Peer Review-Verfahrens zur Qualitätssicherung der Sozialmedizinischen Begutachtung von Gutachten zur EM-Rente. In einem ersten Schritt werden die Inhalte, das didaktische Vorgehen sowie die Rahmenbedingungen der Schulung festgelegt und ausgearbeitet. In einem zweiten Arbeitsschritt wird die Schulung durchgeführt und evaluiert. Der nach Tuijn und Kollegen (2012) zentrale Schritt der Überarbeitung des Manuals, ist dabei nicht Gegenstand der vorgestellten Studie. Die Überarbeitung wurde nach Abschluss des vorangegangenen Projektes von der DRV Bund in Abstimmung mit den Leitenden Ärzten der DRV-Träger vorgenommen.

Das Peer-Schulungsprogramm soll vollständig manualisiert und evaluiert werden. Die Entwicklung der Schulung wird daher durch eine formative und summative Evaluation flankiert, wie es im Bereich der Patientenschulung üblich ist (Faller, Reusch & Meng, 2011; Meng et al., 2009; Seekatz et al., 2013; Worringen, Meng, Bitzer, Brandes & Faller, 2017). Die formative Evaluation dient der systematischen Informationssammlung zur Messung, Bewertung und Optimierung der Schulungsqualität auf Basis empirischer Daten. Die anschließende summative Evaluation untersucht den Schulungserfolg anhand von drei zu beurteilenden sozialmedizinischen Gutachten, die von allen Schulungsteilnehmern beurteilt werden. Entsprechend dieser Prämissen werden in der Studie, die in Tabelle 54 dargestellten Hauptfragestellungen untersucht.

Tabelle 54: Fragestellungen der formativen und summativen Evaluation eines strukturierten Peer-Schulungsprogramms.

1. Hauptfragestellung
Ist die Übereinstimmung der Peers nach dem Besuch einer standardisierten Peer-Schulung höher als nach einer nicht-strukturierten Peer-Schulung? (vgl. Strahl et al., 2018)
Wie hoch ist die prozentuale Übereinstimmung mit dem Peer-Konsens ⁶
2. Nebenfragestellungen
Wie ist die Umsetzbarkeit der standardisierten Peer-Schulung zu beurteilen?
Wie ist die Akzeptanz der standardisierten Peer-Schulung?

⁶ Der Peer-Konsens wurde definiert als der Modus aller Peer-Urteile einer Prüffrage. Der Modus entspricht dabei der Mängelkategorie, die bei der Beurteilung eines Gutachtens bei einer spezifischen Prüffragen am häufigsten vergeben wurde.

5.2 Untersuchungsdesign

Die Studie wurde in enger Kooperation mit der DRV Bund, Bereich Sozialmedizin, und den teilnehmenden Regionalträgern durchgeführt. Das Projekt untergliederte sich in drei Phasen:

1. **Entwicklung des standardisierten Peer-Schulungsprogramms:** Zur Entwicklung der standardisierten Peer-Schulung wurde ein theoriegeleitetes Verfahren zur Definition von Lernzielen und entsprechenden Schulungsmethoden eingesetzt. Dabei richtete sich das didaktische Vorgehen der Schulung an übergeordneten Lernzielen aus. Diese wurden in einzelne konkretisierende Lernziele überführt und jeweils mit einer geeigneten didaktischen Vermittlungsmethode hinterlegt (Birkenbihl, 2014, S. 129-131). Die Phase beinhaltete die Erstellung und Ausarbeitung der Schulung und Schulungsmaterialien sowie die Durchführung eines Vorbereitungsworkshops mit Experten der DRV, die auch als Co-Moderatoren während der Schulung fungierten. Die Phase endet mit der Vorstellung und Konsentierung des Schulungsprogramms mit der DRV Bund, Bereich Sozialmedizin, und der Durchführung der Schulung.
2. **Formative Evaluation:** Nach Durchführung der Schulung wurde anhand selbstkonstruierter Fragebögen die Akzeptanz und Umsetzbarkeit der Peer-Schulung seitens der teilnehmenden Peers geprüft. Potenzielle Änderungsbedarfe wurden in dieser Projektphase analysiert und bei Bedarf umgesetzt.
3. **Summative Evaluation:** Nach Abschluss der Peer-Schulung wurde der Schulungserfolg anhand von drei sozialmedizinischen Gutachten untersucht. Für die Evaluation wurden Maße der Interrater-Übereinstimmung und -Reliabilität berechnet.

Technische Umsetzung des Reviews. Die Bearbeitung und der Versand der Reviews erfolgten in digitaler Form. Die im Vorgängerprojekt bewährte Evaluations- und Umfragesoftware EvaSys® fand auch in diesem Projekt Anwendung. Jedes Gutachten erhielt durch die DRV eine individuelle TAN-Nummer, die mit dem digitalisierten Prüffragenkatalog verknüpft war. Die Reviews wurden durch die DRV Bund, Bereich Sozialmedizin, gesammelt und von dort in pseudonymisierter Form den bearbeitenden Wissenschaftlern der Universität Würzburg zur Verfügung gestellt.

Datenschutz und Ethik. Das Datenschutzkonzept wurde vom Datenschutzbeauftragten der DRV Bund geprüft und bewilligt. Weiterhin wurden sämtliche Regionalträger der Rentenversicherung durch die DRV Bund angeschrieben und über das Projekt informiert sowie deren Einverständnis zur Teilnahme eingeholt. Für die Auswertung der Forschungsdaten wurde kein Ethikvotum beantragt, da die bereitgestellten Daten durch die DRV Bund vollständig pseudonymisiert werden. Entsprechend dieses Vorgehens ließen sich keine direkten Rückschlüsse auf den einzelnen Peer oder RV-Träger ziehen.

5.2.1 Entwicklung des standardisierten Peer-Schulungsprogramms

Vorüberlegungen. Die Ausarbeitung des Peer-Schulungsprogramms orientierte sich an den Erfahrungswerten aus der Entwicklung von Patientenschulungen für die medizinische Rehabilitation. Bei der Aufbereitung des Schulungscurriculums wurde daher auf die Qualitätskriterien der DRV Bund und des Zentrums für Patientenschulung (Ströbl, Friedl-Huber & Küffner, 2007) zurückgegriffen.

Als grundlegende **Qualitätsmerkmale für Schulungen** gelten eine Manualisierung, ein curricularer Aufbau, die Anwendung unterschiedlicher Vermittlungsmethoden, eine auf 15 Personen begrenzte Teilnehmerzahl, Multiprofessionalität und die (regelmäßige) Fortbildung der Trainer. Eine Manualisierung von Schulungscurricula gewährleistet in der Routine eine standardisierte Durchführung unabhängig vom Schulungsdozenten.

Vor dem Hintergrund, dass die Schulung als regelmäßiges Training für Peers angeboten werden soll und bei der Ausarbeitung und Evaluation des Peer Review-Verfahrens bereits Sozialmediziner aus den Häusern der RV-Träger an der Erprobung des Peer Review-Verfahrens teilgenommen haben, wurde das Peer-Schulungsprogramm in Form von zwei separaten Modulen erstellt:

- *Grundlegendes Basismodul* als Einführung für Peers, die bisher nicht mit dem Manual zum Peer Review-Verfahren und dem Prüffragenkatalog gearbeitet haben. Das Modul muss seitens der teilnehmenden Sozialmediziner mindestens einmal vollständig durchlaufen werden und befähigt allein noch nicht zur Ausübung der Peertätigkeit.
- *Konsekutives Anwendermodul* als spezielles Modul für Sozialmediziner, die mit dem Peer Review bereits vertraut sind, das Basismodul bereits absolviert haben oder bereits eine Peer-Funktion ausgeübt haben. Ziel des Anwendermoduls ist, das Beurteilerverhalten in optimaler Weise zu kalibrieren.

Aufgrund der im Gegensatz zur Patientenschulung verschiedenen Zielgruppe und Zielsetzung der Peer-Schulung, konnten a priori einige der aufgeführten Qualitätsmerkmale außer Acht gelassen werden. So wurde die Limitierung der Gesamtteilnehmerzahl aus Praktikabilitätsgründen sowie aufgrund der gewählten didaktischen Vermittlungsmethoden im ersten Teil der Schulung aufgehoben. Darüber hinaus wurde durch die sozialmedizinische Thematik der Schulung eine Multiprofessionalität im Dozententeam nicht zwangsläufig notwendig. Vielmehr sollte eine ausgeprägte sozialmedizinische Expertise und ein großes Sachverständnis über das Manual zum Qualitätssicherungsverfahren in der sozialmedizinischen Leistungsbeurteilung Voraussetzung für die Arbeit als Schulungsdozent sein.

Theoretischer Hintergrund der Lernzielformulierung. Die Erfahrungen aus dem Vorgängerprojekt (Strahl et al., 2016) sowie aus dem Peer-Review der Rehabilitation (Klosterhuis, 2010) deuteten darauf hin, dass ein interaktiver Trainingsansatz am besten geeignet erscheint, um die erforderlichen Peer-Fähigkeiten zur Teilnahme am Review-Verfahren zu erwerben. Die Literatur beschreibt, dass das Lernen in interaktiven Kleingruppen als Teil eines aktiven, kollaborativen, problemorientierten Lernprozesses eine intensive Informationsverarbeitung, ein tieferes Verständnis der Lerninhalte sowie die Möglichkeit, das Verständnis von Lerninhalten zu überprüfen und zu klären, fördert (Jones, 2007; Prince, 2004; Springer, Stanne & Donovan, 1999).

Zur Gewährleistung des interaktiven Trainingsansatzes erfolgte die Ausarbeitung des didaktischen Schulungskonzepts in Anlehnung an die Taxonomie kognitiver Lernziele von Bloom und Kollegen (vgl. Bloom, Engelhart, Furst, Hill & Krathwohl, 1956; Krathwohl, 2002). Die Bloom'sche Taxonomie definiert eine systematische Klassifizierung zur Gliederung von Lern- und Lehrzielen. Bloom definierte sechs hierarchische, kognitive Bildungsziele, die jeweils unterschiedlich detailliert gegliedert werden können (Baumgartner, 2014, S. 36). Jede Hierarchieebene beschreibt ein erhöhtes Abstrahierungsniveau und definiert dadurch die Schwierigkeit des Lernzieles und die Kompetenzen, die die Schulungsteilnehmer erlernen sollen. Die von Krathwohl (2002) revidierte Taxonomie beinhaltet die Dimensionen „Erinnern“, „Verstehen“, „Anwenden“, „Analysieren“, „Bewerten“ und „Erzeugen“ (Tab. 55).

Tabelle 55: Adaptierte Bloom'sche Taxonomie der kognitiven Lernziele (modifiziert nach Krathwohl, 2002, S. 215; Baumgartner, 2014, S. 43-46).

Lernziel-dimension	Beschreibung	Schlüsselverb für Lernziel
Erinnern	Lernende können relevantes Wissen aus dem Gedächtnis wiedergeben.	identifizieren, abrufen, benennen, kennen
Verstehen	Lernende können das Gelernte erklären, beschreiben oder aufzeichnen.	interpretieren, abbilden, erschließen, erklären,
Anwenden	Lernende können etwas Gelerntes in einer anderen Situation anwenden.	durchführen, umsetzen, verwenden, erledigen
Analysieren	Lernende können Materialien in seine Bestandteile zerlegen und erfassen, wie sich die Teile zueinander und zu einer Gesamtstruktur verhalten.	Unterscheiden, auswählen, gliedern, strukturieren, dekonstruieren
Bewerten	Lernende können Beurteilungen nach vorher gelernten Kriterien und Standards vornehmen.	prüfen, urteilen, bewerten, testen, ermitteln
Erzeugen	Lernende können Elementen zu einem neuartigen Produkt zusammensetzen oder ein Originalprodukt herstellen.	entwerfen, Hypothesen formulieren, gestalten, konstruieren, erfinden

Die Lernzieldimensionen wurden dazu genutzt, anhand der mit ihnen assoziierten Schlüsselverben spezifische Lernziele zu formulieren.

Ausarbeitung des inhaltlichen und didaktischen Vorgehens. Ausgehend von den Vorüberlegungen wurde die Schulung in ein Basis- und ein Anwendermodul aufgeteilt. Für jedes der beiden Module wurde ein eigenständiges übergeordnetes Lernziel definiert:

- *Übergeordnetes Lernziel des Basismoduls:* In dem Modul werden die Teilnehmer über den Hintergrund des Qualitätssicherungsprogramms für die Sozialmedizinische Begutachtung informiert. Die Peers erwerben eine fundierte Wissensbasis für das Verständnis der Ziele der Qualitätssicherung und der Methode des Peer Reviews anhand des Manuals zur Qualitätssicherung der Sozialmedizinischen Begutachtung.
- *Übergeordnetes Lernziel des Anwendermoduls:* Dieses Modul versetzt die Teilnehmer in die Lage, die im Manual dargelegte Bewertungsmethode selbstständig und konsistent auf konkrete sozialmedizinische Gutachten anzuwenden und die in den Prüffragen angesprochenen Sachverhalte eines Gutachtens zu analysieren, die Einzelfakten unter dem jeweiligen Qualitätsaspekt (Formale Gestaltung, Verständlichkeit, Transparenz, Vollständigkeit, medizinisch-wissenschaftliche Grundlagen, Wirtschaftlichkeit) zu werten und vorliegende Gutachten im Hinblick auf die definierten Mängel konsistent zu evaluieren.

Im Anschluss wurden die übergeordneten Lernziele in konkretere Lernziele und Lerninhalte überführt und mit entsprechenden didaktischen Methoden hinterlegt. Die festgelegte Reihenfolge der Module resultierte infolge der aufeinander aufbauenden Inhalte der Schulung. Für jedes Modul wurden etwa fünf konkrete Lernziele formuliert. Das Schulungsprogramm beinhaltet sowohl frontale als auch aktivierende bzw. interaktive didaktische Techniken. Jede Vermittlungsmethode wurde unter der jeweiligen Lernzielperspektive begründet und unter Einbezug sozialmedizinischer Expertise ausgewählt. Als didaktische Szenarien oder Bausteine wurden ausgearbeitet:

- Plenarvortrag/Impulsbeitrag (z. B. unter der Perspektive: Kennenlernen des historischen Hintergrunds der Qualitätssicherung),
- Gruppendiskussion (z. B. unter der Perspektive: Identifikation mit den Zielsetzungen des Qualitätssicherungsprogramms),
- Kleingruppenarbeit (z. B. unter der Perspektive: Verständnis der Bewertungssystematik),
- Einzelarbeit (z. B. unter der Perspektive: Verständnisevaluation),
- Übung (z. B. unter der Perspektive: Anwendung EvaSys®),
- Hausaufgabe (z. B. unter der Perspektive: Verstetigung und Konsistenz des Bewertungsmaßstabs)

Das Peer-Schulungsprogramm sollte durch erfahrene Sozialmediziner als Co-Moderatoren unterstützt werden. Durch den Einbezug der sozialmedizinischen Experten wurden besonders in den Kleingruppenphasen der Beurteilungsspielraum auf die in den Qualitätskriterien definierten Aspekte fokussiert, um eventuell individuell ausgeprägte, abweichende Bewertungsmaßstäbe zu harmonisieren.

Schulungsunterlagen. Zur Unterstützung der Lernzielumsetzung wurden folgende Materialien einbezogen:

- *Manual zum Peer Review-Verfahren für die Sozialmedizinische Begutachtung:* Grundlage des gesamten Peer Reviews war das Manual mit den definierten Prüffragen. Dieses wurde im Rahmen der Peer-Schulung sowohl in einem einführenden Vortrag vorgestellt und in Kleingruppen besprochen als auch in einer Einzelübung vom Peer bezüglich prüffragenspezifischer Problemlagen selbstständig bearbeitet und anschließend im Plenum nachbearbeitet.
- *Begutachtungsliteratur der DRV:* Auf die vorhandene Begutachtungsliteratur „DRV Schriften, Band 21, Das ärztliche Gutachten für die gesetzliche Rentenversicherung“, die Buch-Publikation „Leitlinien für die Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung“, die „Leitlinien für die Sozialmedizinische Begutachtung“ sowie auf die „DRV Schriften, Band 81, Sozialmedizinisches Glossar der Deutschen Rentenversicherung“ wurde als Grundlagenliteratur hingewiesen und der Schulung zugrunde gelegt bzw. bei Bedarf einbezogen.
- *Anonymisierte sozialmedizinische „Schulungsgutachten“:* Im Gegensatz zum Basismodul wurde im Anwendermodul der Umgang mit dem Manual zum Peer Review-Verfahren an praktischen Beispielen erörtert. Das Anwendermodul beinhaltet schwerpunktartig die richtige Anwendung des Vier-Kategorien-Bewertungssystem der Prüffragen der Einzelkriterien, die Beurteilung der Nachvollziehbarkeit eines Gutachtens gemäß der „Ampelbewertung“ und die Dokumentation der gefundenen Qualitätsdefizite. Für die praxisnahe Übung und Anwendung des Manuals wurden drei sozialmedizinische Beispielgutachten der Hauptindikationsbereiche (Orthopädie, Innere/Allgemeinmedizin, Neurologie/Psychiatrie) von der DRV Bund, Bereich Sozialmedizin, zur Verfügung gestellt.
- *Anonymisierte sozialmedizinische „Hausaufgabengutachten“:* Nach Beendigung der Schulung sollten die Peers, ähnlich wie in Teilstudie 2, autonom drei Gutachten anhand des Prüffragenkatalogs beurteilen. Diese Gutachten wurden ebenfalls in anonymisierter Form von der DRV Bund, Bereich Sozialmedizin, zur Verfügung gestellt.

Im Schulungscurriculum wurden die einzelnen Lernziele, die zu vermittelnden Inhalte und die entsprechend gewählten Vermittlungsmethoden detailliert dargestellt. Es wurden die Dozentenvoraussetzungen festgelegt und der organisatorische Rahmen beschrieben. Mit dem Schulungscurriculum sollte eine standardisierte Durchführung der beiden Module des Peer-Schulungsprogramms dozentenunabhängig erreicht werden. Abbildung 17 veranschaulicht beispielhaft den Aufbau eines Lernzieles im Schulungscurriculum.

Definition der Rahmenbedingungen. Das Peer-Schulungsprogramm sollte, vor allem im Anwendermodul, wesentlich auf Kleingruppenarbeit ausgelegt sein. Gleichzeitig enthielt es aber auch Einheiten, die in der Großgruppe (Plenum) durchgeführt werden. Da der Austausch zwischen Peers als wesentliches Element zur Verbesserung des trägerübergreifenden Qualitätssicherungsverfahrens angenommen wird, sollte eine Gesamtteilnehmerzahl von 24 Teilnehmern nicht überschritten werden. Dabei sollte die Kleingruppenarbeit in zwei parallelen Gruppen mit maximal zwölf Teilnehmern erfolgen, um die aktivierenden und teilnehmerorientierten Methoden erfolgreich anzuwenden. Jede Kleingruppe benötigte für die Durchführung einen eigenen Sozialmedizinischen Experten als Moderator. Im Sinne der einzuhaltenden Qualitätsdimensionen sollten bei der Schulung Rahmenbedingungen wie bspw. ein adäquater Schulungsraum oder die Verfügbarkeit von Medien und Material (Flipchart, Beamer) eingehalten werden.

Erwerb von Wissen	Erwerb von Handlungskompetenz
Lernziel	
Der Peer kennt den Aufbau der Prüffragen des Manuals zum Peer Review-Verfahren zum Qualitätssicherungsverfahren der sozialmedizinischen Begutachtung	
Begründung	
Eine sichere Ausführung der Peer-Tätigkeit setzt die Kenntnis über das Manual zum Qualitätssicherungsverfahren der sozialmedizinischen Begutachtung voraus. Die einzelnen Prüffragen bilden die Grundlage für die Bewertung der jeweiligen Qualitätskriterien.	
Inhalt	
• Erläuterung des hierarchischen Qualitätsmodells (Einzelkriterien, Übergeordnetes Kriterium) • Beschreibung des formalen Aufbaus einer Prüffrage – Einordnung in das Einzelkriterium – Prüffragennummer – Prüffrage – Erläuterung zur Prüffrage – Bewertungssystematik: Mängelbewertung – Hinweis zur Dokumentation der festgestellten Qualitätsdefizite, die der Mängelbewertung zu Grunde liegen	
Didaktische Vermittlungsmethode	
Vortrag, Diskussion.	
Zeit circa 25 Minuten	
Anmerkung	
Grundsatzdiskussionen über die Einzelkriterien und Prüffragen sollen vermieden werden. Ziel ist die Vermittlung der Inhalte des Manuals zum Peer Review-Verfahren zur Qualitätssicherung der sozialmedizinischen Begutachtung und der prinzipielle Aufbau einer Prüffrage.	

Abbildung 17: Veranschaulichungsbeispiel aus dem Schulungscurriculum (Entwurf).

Expertenworkshop. Um einen ersten Entwurf des Peer-Schulungsprogramms mit der DRV Bund zu konsentieren, wurde nach Erstellung einer ersten Entwurfsfassung ein Expertenworkshop mit sieben sozialmedizinischen Experten aus dem Lenkungsausschuss Qualitätssicherung und dem Expertengremium der DRV Bund durchgeführt.

Der Fokus des Expertenworkshops lag auf der Qualitätsbeurteilung der sechs sozialmedizinischen Gutachten, die innerhalb der Schulung und der Hausaufgabe zu bearbeiten waren. Ziel des Workshops war die Beurteilung dieser Gutachten anhand des Manuals zum Peer Review-Verfahren durch die Expertengruppe. Da vier der Experten gleichzeitig als Co-Moderatoren für die Kleingruppen im Anwendermodul der Schulung fungierten, wurden die Gutachten sowie die konsentierten Lösungen besprochen, um eine größtmögliche Übereinstimmung bei der Beantwortung der Prüffragen zu erzielen.

5.2.2 Durchführung der Peer-Schulung

Insgesamt meldeten sich 40 Sozialmediziner aus allen 16 Rentenversicherungsträgern zur Peer-Schulung an. Um eine gute Erreichbarkeit zum Schulungsort zu gewährleisten und die festgelegte Gruppengröße nicht zu übertreten, wurden in Würzburg und Berlin insgesamt zwei Peer-Schulung im Abstand von einer Woche durchgeführt. Dreizehn der 40 teilnehmenden Sozialmediziner waren bereits am vorangegangenen Projekt zur Analyse der Interrater-Reliabilität beteiligt, 27 Sozialmediziner wurden mit der strukturierten Schulung neu geschult.

Die teilnehmenden Peers hatten im Mittel 12 (SD 7,2) Jahre Erfahrung im sozialmedizinischen Dienst. Ferner übten 21 Peers sowohl gutachterliche als auch prüfärztliche Tätigkeiten aus. Acht Teilnehmer gaben an ausschließlich prüfärztliche Tätigkeiten durchzuführen, sieben weitere Peers übten nur gutachterliche Tätigkeiten im sozialmedizinischen Dienst der Rentenversicherung aus⁷.

Der Prüffragenkatalog des Manuals zum Peer Review-Verfahren wurde im Zuge der Ergebnisauswertung der Pilotstudie (vgl. Strahl et al., 2016) von 22 auf 17 Prüffragen um ein Fünftel komprimiert, in einzelnen Prüffragen sprachlich modifiziert, in seiner grundsätzlichen Struktur jedoch beibehalten (DRV, 2013). Diese Version des Manuals ist im Anhang 9 einzusehen und diente in der Schulung als Beurteilungsgrundlage der sozialmedizinischen Gutachten.

⁷ Daten basieren auf n = 36 Peers, 4 missings

5.2.3 Formative Evaluation

Die Peer-Schulung wurde von einer formativen Evaluation begleitet. Ziel war die Überprüfung der Umsetzbarkeit der Schulung nach Manual und die Akzeptanz der Schulung bei den Peers, um dadurch Anhaltspunkte für weitere Überarbeitungsschritte bzw. die Fertigstellung der Schulung zu erhalten.

Für die formative Evaluation des Peer-Schulungsprogramms wurde ein selbstkonstruiertes Bewertungsinstrument entwickelt und eingesetzt. Das Bewertungsinstrument orientierte sich an etablierten formativen Evaluationsinstrumenten der Patientenschulung (vgl. Meng, 2009), wurde aber auf das Schulungscurriculum zur Qualitätssicherung der Sozialmedizinischen Begutachtung adaptiert. Die Beurteilung fand nach Beendigung des letzten Lernziels statt. Als Komponenten der Schulungsbewertung wurden die Auswahl, Verständlichkeit und subjektiver Nutzen der Inhalte, Vortragsweise des Dozenten sowie die Materialgestaltung herangezogen. Darüber hinaus konnten über offene Fragen (z. B. „Was hat Ihnen besonders gut gefallen?“, „Was hat Ihnen nicht gefallen?“, „Welche Änderungen und Ergänzungen schlagen Sie vor?“) Feedback und Verbesserungsmöglichkeiten aus Teilnehmerperspektive benannt werden.

5.2.4 Summative Evaluation

Ziel der summativen Evaluation war die Prüfung der Effektivität der Peerschulung. Die Überprüfung des Schulungserfolgs fand im Anschluss an die Peer-Schulung statt. Dazu beurteilen alle Peers innerhalb von zwei Monaten drei Gutachten anhand des Manuals zum Peer Review-Verfahren selbstständig und unter Nutzung des Feedbacksystems EvaSys®.

Im Vorfeld der Schulung wurde im Rahmen des Expertenworkshops durch die sozialmedizinischen Experten eine konsentierte Empfehlung für die Beurteilungen der Prüffragen für alle drei Gutachten im Konsensverfahren erarbeitet („Musterlösung“). Dieses Vorgehen bot die Möglichkeit, den individuellen Schulungserfolg quantitativ zu bestimmen. Zum Einschluss in die Studie und Teilnahme am Projekt mussten die sozialmedizinischen Gutachten folgende Einschlusskriterien erfüllen:

- Erstgutachten auf EM-Rente,
- Gutachten mit körperlicher Untersuchung,
- letzter Rehabilitationsantrag des Antragstellers muss länger als ein Jahr zurück liegen.

Für die Analyse lagen von zwei der drei Gutachten vollständige Reviews von allen 40 teilnehmenden Peers vor, ein Gutachten wurde von 39 Sozialmedizinern beurteilt. Die datenanalytische Auswertung schließt 119 Reviews ein. Dies entspricht einem Rücklauf von 99,2 %.

Zielkriterium Reliabilität. Als wesentliches Gütekriterium ist die Reliabilität zu überprüfen (Goldman, 1994). Das Hauptbewertungskriterium dieser Studie war die Interrater-Reliabilität des übergeordneten Kriteriums, als wichtigstes Kriterium des Peer Review-Verfahrens. Neben der Untersuchung der Peer-Konkordanz erlaubte das vorliegende Studiendesign, durch die vorherige Konsentierung der Hausaufgabengutachten (Musterlösungen) die Übereinstimmung mit den Musterlösungen zu untersuchen bzw. die Übereinstimmung der Musterlösungen mit dem Mehrheitsurteil (Modus) der Peers in Beziehung zu setzen. Mit dieser Methode sollte die Erwartung überprüft werden, dass die Übereinstimmung mit dem Mehrheitsurteil („Peer-Konsens“) infolge der Schulung höher ausfällt als die Übereinstimmung mit der konsentierten Musterlösung („Experten-Konsens“).

Es ist davon ausgegangen worden, dass die Kalibrierung der Bewertungsmaßstäbe durch die Teilnahme am Anwendermodul gleichmäßige Beurteilungen erzeugt und die Übereinstimmung dadurch positiv beeinflusst wird. Für die Beurteilung der Effektivität wurden zunächst berechnet:

- die prozentuale Übereinstimmung, d.h. der prozentuale Anteil der Beurteilungen, in denen mehrere Peers das gleiche Urteil abgeben,
- die prozentuale Übereinstimmung mit dem Peer-Konsens,
- die prozentuale Übereinstimmung mit dem Experten-Konsens.

Schulungsevaluation. Ergänzend zur prozentualen Konkordanz griff die Evaluation der Peer-Schulung auf das Interrater-Reliabilitätsmaß Fleiss Kappa für kategorial und ordinal skalierte Daten nach Bortz und Lienert (2008 S. 454) zurück. Das Reliabilitätsmaß wurde gewählt, da die Reliabilitätsanalyse der Pilotstudie zeigte, dass die Höhe der Reliabilität bei Reduzierung der zu beurteilenden Fallzahl durch gebräuchliche Koeffizienten nicht mehr valide abgebildet werden kann (Strahl et al., 2012, S. 44). Im Übereinstimmungsmaß **κ_m -Koeffizient von Fleiss für multiple Rater** wird berücksichtigt, dass eine gewisse Urteilsübereinstimmung zufällig zustande kommen kann. Die resultierende Übereinstimmungswahrscheinlichkeit wird auf eine signifikante Abweichung von der Zufallsübereinstimmungsrate geprüft. Der Kennwert (p_i) relativiert die Anzahl der Urteilerpaare, die in ihren Urteilen pro Gutachten übereinstimmen, an der maximal möglichen Anzahl von übereinstimmenden Urteilerpaaren. Gemittelt über die Fälle (p) ergibt sich die Wahrscheinlichkeit, mit der die Gutachten eines beliebigen Urteilerpaares im Durchschnitt identisch beurteilt wurden. Die resultierende Übereinstimmungswahrscheinlichkeit wird auf eine signifikante Abweichung von der Zufallsübereinstimmungsrate geprüft. Ein Wert nahe 1 zeigt, dass die Peers vergleichbare Urteile vergeben, ein Wert von 0, dass dies nicht der Fall ist (Fleiss, 1971; Gisev et al., 2013).

Fleiss Kappa wurde zunächst als globales Reliabilitätsmaß berechnet. Um die Wirksamkeit der standardisierten Peer-Schulung zu untersuchen, wurde zusätzlich Fleiss Kappa der 27 neu geschulten Peers mit den Review-Ergebnissen aus dem früheren unstrukturierten, nicht-standardisierten Peer-Training ohne Schulungscurriculum aus der Teilstudie 2 verglichen. Die Kennwerte wurden aus den Reviews der drei sozialmedizinischen Probegutachten berechnet, die nach dem Schulungsworkshop im Rahmen der Evaluation des Manuals zum Peer Review-Verfahrens beurteilt wurden (vgl. Kap. 4.2.2). Die Reliabilität, die nach der ersten Schulung der Peers berechnet wurde, diente entsprechend als Kontrollgruppe (KG), während die Reliabilität der neu geschulten Peers die Interventionsgruppe (IG) darstellte. Die Reliabilität der KG wurde anhand von 56 Reviews (von 57 erwarteten Reviews; 98,2 % Rücklaufquote) ermittelt, die von 19 Peers erstellt wurden.

Abbildung 18 veranschaulicht zusammenfassend die dargestellten Berechnungsschritte der summativen Evaluation. Die Abbildung zeigt die drei in der Studie verwendeten Reliabilitätskonstellationen (Interrater-Reliabilität zwischen den Peers, Übereinstimmung der Peers mit dem Peer-Konsens, Übereinstimmung mit der Musterlösung), um die Effektivität der standardisierten Peer-Schulung abzubilden.

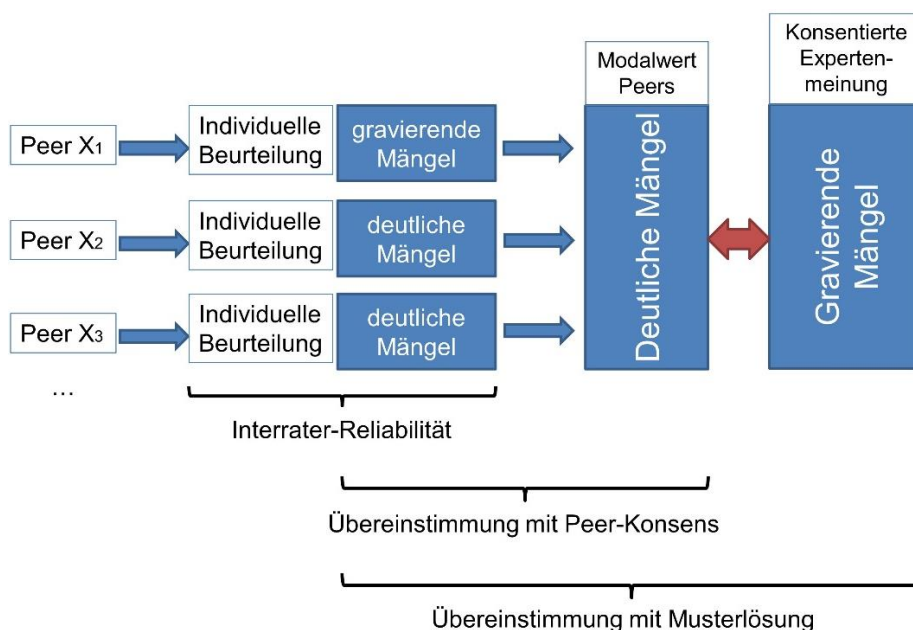


Abbildung 18: Ebenen der Reliabilitätsberechnungen im Rahmen der summativen Evaluation.

Alle Analysen wurden mit Excel und SPSS für Windows (Version 24.0, IBM) durchgeführt.

5.3 Ergebnisse

5.3.1 Aufbau des standardisierten Peer-Schulungsprogramms

Das ausgearbeitete konsentierende, zweieinhalbtägige Peer-Schulungsprogramm setzte sich aus zwei Modulen und insgesamt neun Lernzielen zusammen und wurde in einer geschlossenen Gruppe umgesetzt. Das Basismodul umfasste 270 min inklusive 30 min Pause. Das zeitlich länger konzipierte Anwendermodul, welches über zwei Tage verteilt stattfand, umfasste insgesamt 720 min inklusive 120 min Pause.

Die Schulung wurde durch erfahrene Sozialmediziner durchgeführt, die im Anwendermodul jeweils eine Kleingruppe von maximal 12 Personen anleiteten. Während der Durchführung im Rahmen der vorliegenden Teilstudie wurde das Basismodul durch zwei Psychologen und die Kleingruppen des Anwendermodul jeweils durch einen Psychologen und einen sozialmedizinisch erfahrenen Co-Moderator angeleitet.

Grundlage zur Durchführung bildete ein standardisiertes Schulungscurriculum. Bei der Umsetzung der Schulung wurden Präsentationsfolien, Flipchart und Arbeitsblätter eingesetzt. Bei der Vermittlung wurde ein interaktives Vorgehen mit hoher Peer-Beteiligung realisiert. Als Methoden wurden interaktive Kurzvorträge, Gruppendiskussionen, Gruppenarbeiten und Übungen angewandt. Jede Vermittlungsmethode wurde im Schulungscurriculum unter der jeweiligen Lernzielperspektive begründet (Tab. 56). Eine Übersicht über die Struktur, die Lernziele und die Methoden der Schulung zeigt Tabelle 57.

Tabelle 56: Lernzielinhalte und Vermittlungsmethoden.

Lernzielinhalt	Geeignete didaktische Methode
Erwerb von Wissen	• (Plenar-)Vortrag • Gruppendiskussion
Erwerb von Kompetenzen	• Übung • Gruppendiskussion
Erwerb von Fertigkeiten	• Kleingruppenarbeit • Übung

Das vollständige Schulungscurriculum im Rahmen des von der DRV Bund geförderten Projektes ist Eigentum der DRV Bund, Bereich Sozialmedizin, und kann daher nicht im Anhang dieser Arbeit zur Verfügung gestellt werden. Das Schulungscurriculum kann jedoch auf Anfrage und Genehmigung durch die DRV Bund, Bereich Sozialmedizin, vom Autor bezogen werden.

Tabelle 57: Aufbau und Lernziele des Schulungscurriculums (modifiziert nach Strahl et al., 2018, S. 4).

Modul	Zeit (min)	Lernziel	Lernziel-dimension nach Krathwohl	Methoden
Basis-modul	90	1.1 Der Peer kennt* die Grundlagen der Qualitätssicherung in der Sozialmedizinischen Begutachtung	Erinnern	Vortrag, Diskussion, Kleingruppenarbeit
	60	1.2 Der Peer kennt* Inhalt und Aufbau des Manuals zum Peer Review-Verfahren	Erinnern	Vortrag, Übung, Einzelarbeit
	30	1.3 Der Peer kennt* den formellen Ablauf des Peer Review-Verfahrens in der Routine	Erinnern	Vortrag
	45	1.4 Der Peer kennt* die Aufgaben in seiner Rolle als Peer	Erinnern	Vortrag, Diskussion
Anwender-modul	30	2.1 Der Peer wendet* eine einheitliche Herangehensweise bei der Bewertung eines sozialmedizinischen Gutachtens entlang der vorgegebenen Prüffragen an	Anwenden	Vortrag, Diskussion, Kleingruppenarbeit
	360	2.2 Der Peer kann die Prüffragen für individuelle Gutachten beantworten, indem das Vier-Kategorien-Bewertungssystem regelhaft angewendet* wird	Anwenden	Kleingruppenarbeit
	60	2.3 Der Peer kann die Nachvollziehbarkeit eines Gutachtens gemäß der Ampelbewertung bewerten*	Bewerten	Vortrag, Kleingruppenarbeit
	---(*)	2.4 Der Peer kann Qualitätsdefizite nachvollziehbar dokumentieren*	Analysieren	Kleingruppenarbeit
	220	2.5 Der Peer kann den Review-prozess eigenständig durchführen* und die Antworten der Prüffragen mit Hilfe des Informationsübermittlungssystem EvaSys® dokumentieren	Bewerten	Übung, Vortrag, Kleingruppenarbeit

(*) Lernziel 2.4 ist nicht mit einer bestimmten Bearbeitungszeit verbunden. Dieses Ziel ist in den Lernziele 2.2 und 2.3 integriert. Der Schulungsmoderator diskutierte während der Bearbeitung dieser Lernziele gemeinsam mit den Peers das adäquate Verfassen der Mängeldokumentation am konkreten Gutachtenbeispiel.

* Schlüsselverb im Zusammenhang mit der Wissensdimension der erweiterten Taxonomie nach Bloom zur Operationalisierung der Lernziele

5.3.2 Umsetzung des standardisierten Peer-Schulungsprogramms

Basismodul. Im Lernziel 1.1 wurden die Peers in die Grundlagen und Methoden der Qualitätssicherung in der Sozialmedizinischen Begutachtung eingeführt. Neben der Erläuterung des Anforderungsprofils und der Gliederung eines ärztlichen Gutachtens, war es Aufgabe der teilnehmenden Sozialmediziner, in einer Kleingruppenarbeit subjektive Qualitätskriterien von „guten“ und „schlechten“ Gutachten zu bestimmen.

Lernziel 1.2 vermittelte anschließend anhand eines Vortrages den Inhalt und Aufbau des Manuals zum Peer Review-Verfahren. Neben der Vorstellung des Manuals anhand einer Beispielprüffrage erarbeiteten die Peers zusätzlich in einer Übung die im Manual vorhandenen Einschränkungen, wie z. B. die Beschränkung von Prüffragen auf spezifische Gutachtenbereiche oder die im Ausschluss nicht zu beurteilenden Qualitätsaspekte. Den Peers wurde dazu ein Arbeitsblatt ausgeteilt, das für die weitere Nutzung in der Routine eingesetzt werden kann.

Im Lernziel 1.3 wurde den Peers der vorgesehene formelle Ablauf des Review-Verfahrens anhand eines Schaubildes erläutert. Die Peers erhielten Informationen über die vorgesehene Anzahl und Ziehungsmodalitäten der zu beurteilenden sozialmedizinischen Gutachten sowie das Online-Befragungssystem EvaSys®.

Im Lernziel 1.4 wurde in Form einer Gruppendiskussion die Notwendigkeit der Beurteilung von Qualität anhand festgelegter Qualitätskriterien erläutert. Ferner wurde abschließend das intuitive Qualitätsverständnis aus Lernziel 1.1 (subjektive Qualitätskriterien) mit den Qualitätsaspekten des Manuals abgeglichen.

Anwendermodul. Die Umsetzung dieses Moduls erfolgte annähernd vollständig in Kleingruppenarbeit. Lernziel 2.1 diente als Einstieg in die Kleingruppenarbeitsphase. Da am Anwenderworkshop auch Sozialmediziner teilnahmen, die das Basismodul bereits durchlaufen haben, wurden zunächst die nur am Anwendermodul teilnehmenden Peers (n = 13) begrüßt und die Notwendigkeit diskutiert, die im Manual vorgegebene Reihenfolge der Prüffragen bei der Erstellung eines Reviews abzuarbeiten. Anschließend wurden die bisherigen Erfahrungen der Peers im Umgang mit sozialmedizinischen Gutachten gesammelt. Nach Beendigung dieser Lerneinheit wurden die Teilnehmer der Pilot-Schulungen in zwei Gruppen zu je 10 Peers aufgeteilt, um kleingruppenorientierte Arbeit mit Beispielgutachten zu ermöglichen.

Im Lernziel 2.2 wurde anschließend in Kleingruppen die regelhafte Anwendung des Vier-Kategorien-Bewertungssystem am Schulungsgutachten 1 anhand des im Peer-Schulungscurriculum dargelegten Ablaufes geübt. Die 10 Peers jeder Kleingruppe fanden

sich in Murmelgruppen⁸ zu je drei bis vier Personen zusammen, um kontinuierlich an den Lernzielen 2.2 und 2.3 in eine festen Murmelgruppe zu arbeiten. Es wurde seitens der Moderatoren darauf geachtet, dass in jeder Murmelgruppe ein Peer eingebunden wurde, der mit dem Verfahren bereits vertraut war. Die Murmelgruppen sind fester Bestandteil der Kleingruppenarbeit und sollen zentral dazu beitragen, die Beurteilungsmaßstäbe der Peers zu vereinheitlichen. Jede Murmelgruppe bearbeitete sukzessiv alle Prüffragen der Einzelkriterien. Die Peers innerhalb einer jeden Kleingruppen sollten im Diskurs zu einem gemeinsamen Urteil kommen. Im Anschluss wurden die Bewertungen der einzelnen Murmelgruppen in der Kleingruppe diskutiert. Die Moderatoren hielten alle Ergebnisse stichpunktartig auf einer Flipchart fest. Die Beurteilung des übergeordneten Kriteriums wurde an zwei Gutachten eingeübt und diskutiert.

Das Lernziel 2.3 sollte die Peers in die Lage versetzen, das übergeordnete Kriterium des Qualitätssicherungsverfahrens, die Nachvollziehbarkeit des Gutachtens, im Sinne des Manuals zu interpretieren und die entsprechende Bewertungskategorien einheitlich zu vergeben. Dazu wurde das übergeordnete Kriterium an zwei sozialmedizinischen Gutachten in den Murmelgruppen bearbeitet. Zunächst wurde das bereits in den Einzelkriterien beurteilte Gutachten 1, im Anschluss das Schulungsgutachten 2 gereviewt.

Lernziel 2.4 wurde nicht als eigenständige Lerneinheit entwickelt. Es wurde vielmehr in die Lerneinheiten 2.2 und 2.3 integriert. Die Moderatoren diskutierten während der Bearbeitung dieser Lernziele gemeinsam mit den Peers über das adäquate Verfassen von Mängeldokumentation. Alle genannten Mängel wurden jeweils am Flipchart notiert und besprochen.

Lernziel 2.5 fand am dritten Tag der Schulung statt. In dieser Lerneinheit bearbeiteten die Peers eigenständig das Schulungsgutachten 3 anhand der Prüffragen des Manuals zum Peer Review-Verfahren. Da in der Pilotschulung kein EvaSys® zur Verfügung gestellt werden konnte, wurde für die Durchführung des Reviews auf eine Paper & Pencil-basierte Version des Prüffragenkatalogs zurückgegriffen. Im Gegensatz zum vorhergegangenen Lernziel 2.4 beurteilten die Peers innerhalb dieser Lerneinheit die Prüffragen und das übergeordnete Kriterium anhand des Schulungsgutachtens 3 eigenständig ohne vorangestellter Konsensbildung in der Murmelgruppe. Nach Abschluss des Reviews wurden die Ergebnisse in der Kleingruppe besprochen und ggf. abweichende Mängelbeurteilungen kalibriert.

⁸ Murmelgruppen sind kleine Diskussionsgruppen zur Aktivierung aller Teilnehmer, mit dem Ziel die Hemmschwelle herabsetzen, sich zu äußern. Dadurch können sich alle Peers am Beurteilungsprozess beteiligen (Sennekamp, Kollwe & Ochsendorf, 2018, S. 171).

Zum Abschluss der Schulung wurden mit allen teilnehmenden Sozialmediziner:innen im Plenum offene Fragen zu den Gutachten oder zum Peer Review-Verfahren diskutiert.

Die notierten Ergebnisse der Murmelgruppenbewertungen des Lernzieles 2.2 zu den Prüffragen der Einzelkriterien, des Lernzieles 2.3 zum übergeordneten Kriterium sowie die Ergebnisse der einzelnen Peerbeurteilungen aus dem Lernziel 2.5 sind in den Anhängen 10 und 11 aufgeführt.

5.3.3 Formative Evaluation

Die Ergebnisse der formativen Evaluation zeigten beim Fragenkomplex zur selbsteingeschätzten Sicherheit im Umgang mit dem Manual, dass sich die Peers ($n = 36$) im Umgang mit den Prüffragen des Manuals als sicher einschätzen (Tab. 58; Sicherheit im Umgang mit den Prüffragen (range): $70 \% < x < 84 \%$). Das Hauptzielkriterium des Manuals (Nachvollziehbarkeit des Gutachtens) konnte von den Peers mit einer Sicherheit von im Mittel 76% bewertet werden. Der Umgang mit den Prüffragen zur Transparenz wurde mit 70% am unsichersten beschrieben. Die durchschnittliche selbsteingeschätzte Gesamtsicherheit im Umgang mit den beurteilungsrelevanten Bestandteilen des Manuals betrug $76,7 \%$.

Tabelle 58: Sicherheit im Umgang mit den Bestandteilen des Manuals in Prozent ($n = 36$) (nach Strahl et al., 2018).

Sicherheit im Umgang...	Mittelwert	Median	min	max
...mit den Prüffragen zur formalen Gestaltung	83,3	90,0	30,0	100,0
...mit den Prüffragen zur Verständlichkeit	73,9	70,0	40,0	100,0
...mit den Prüffragen zur Transparenz	70,0	70,0	40,0	100,0
...mit den Prüffragen zur Vollständigkeit	71,7	70,0	40,0	100,0
...mit den Prüffragen zu den medizinisch-wissenschaftlichen Grundlagen	79,4	80,0	20,0	100,0
...mit den Prüffragen zur Wirtschaftlichkeit	83,9	90,0	50,0	100,0
...mit dem übergeordneten Kriterium	75,8	80,0	30,0	100,0
...bei der Dokumentation der festgestellten Mängel	75,3	80,0	30,0	100,0
Gesamtsicherheit	76,7	---	---	---

Anmerkung: 0% = sehr unsicher, 100% = sehr sicher.

Ferner wurde auf einer 4-stufigen Skala beurteilt, wie hilfreich für die Peers die didaktischen Methoden und Schulungsmaterialien waren. Es zeigte sich, dass die Möglichkeit den Prüffragenkatalog in der Kleingruppe zu bearbeiten und der kleingruppenorientierte

Erfahrungsaustausch von mehr als 90 % der Teilnehmer als ziemlich oder sehr hilfreich bewertet wurden (Tab. 59). Ferner wurden die konsentierten Lösungen und der Einbezug der Co-Moderation zwar als hilfreich beurteilt, jedoch wurden bei dieser Frage nicht so deutliche Ergebnisse erzielt, wie die Arbeit in den Kleingruppenarbeit.

Tabelle 59: Beurteilung der didaktischen Methoden und Schulungsmaterialien (n = 36).

Wie hilfreich fanden Sie...	gar nicht hilfreich n (%)	etwas hilfreich n (%)	ziemlich hilfreich n (%)	sehr hilfreich n (%)
...den Prüffragenkatalog in seinem formalen Aufbau kennen zu lernen*	0 (0 %)	3 (13,6 %)	8 (36,4 %)	11 (50,0 %)
...die Möglichkeit, den Prüffragenkatalog zu nächst in der Kleingruppe zu bearbeiten	2 (5,6 %)	0 (0 %)	5 (13,9 %)	29 (80,6 %)
...die Möglichkeit, den Prüffragenkatalog anschließend allein zu bearbeiten	0 (0 %)	3 (8,6 %)	6 (17,1 %)	26 (74,3 %)
...den kleingruppenorientierten Erfahrungsaustausch bei Beurteilung der Schulungsgutachten	0 (0 %)	1 (2,8 %)	9 (25,0 %)	26 (72,2 %)
...die konsentierten Lösungen	0 (0 %)	8 (22,2 %)	16 (44,4 %)	12 (33,3 %)
...den Einbezug von erfahrenen Sozialmedizinerinnen als Co-Moderatoren	0 (0 %)	6 (17,1 %)	10 (28,6 %)	19 (54,3 %)
...die Schulungsmaterialien	0 (0 %)	2 (5,9 %)	18 (52,9 %)	14 (41,2 %)

* Item 1 wurde nur von den Peers beantwortet, die am Basismodul teilgenommen haben (n = 22)

Die durchschnittliche Bewertung der Qualität der Durchführung der Schulungsmodule sowie der Schulung selbst lag im guten Bereich (Tab. 60, $1,4 < x < 2,5$). Auch hier wurde die Arbeit in den Kleingruppen im Mittel als sehr gut (1,4) bewertet. Hingegen lag die Beurteilung der Auswahl der Schulungsgutachten nur im gutem bis befriedigendem (2,5) Bereich.

Tabelle 60: Qualitative Beurteilung des Schulungscurriculums und dessen Durchführung (n = 36).

Wie beurteilen Sie...	$\bar{x}$	1 n (%)	2 n (%)	3 n (%)	4 n (%)	5 n (%)	6 n (%)
...die Auswahl der Schulungsgutachten	2,5	3 (8,3 %)	21 (58,3 %)	6 (16,7 %)	3 (8,3 %)	3 (8,3 %)	0 (0 %)
...die Schulungsmaterialien	2,0	8 (22,2 %)	23 (63,9 %)	4 (11,1 %)	0 (0 %)	1 (2,8 %)	0 (0 %)
...die Arbeit in den Kleingruppen	1,4	23 (63,9 %)	11 (30,6 %)	1 (2,8 %)	1 (2,8 %)	0 (0 %)	0 (0 %)
...die Moderation der Kleingruppen	1,5	21 (60,0 %)	12 (34,3 %)	2 (5,7 %)	0 (0 %)	0 (0 %)	0 (0 %)
...die Vorträge und Präsentation	1,8	15 (41,7 %)	17 (47,2 %)	3 (8,3 %)	0 (0 %)	1 (2,8 %)	0 (0 %)
...die Organisation der Veranstaltung	1,7	16 (44,4 %)	17 (47,2 %)	2 (5,6 %)	0 (0 %)	1 (2,8 %)	0 (0 %)

Anmerkung: 1=sehr gut, 2=gut, 3=befriedigend, 4=ausreichend, 5=mangelhaft, 6=ungenügend.

In der allgemeinen abschließenden Bewertung der Peer-Schulung wurde diese insgesamt als gelungen (M=1,6) und empfehlenswert (M = 1,6) eingeschätzt (Tab. 61). Insgesamt gaben 94 % der Teilnehmer an, dass die Schulung gelungen sei. Entsprechend würden auch 89 % der Teilnehmer die Schulung weiterempfehlen.

Tabelle 61: Allgemeine Beurteilung der Peer-Schulung (n = 36).

	trifft gar nicht zu n (%)	trifft eher nicht zu n (%)	trifft eher zu n (%)	trifft völlig zu n (%)
Meine Erwartungen an die Peer-Schulung wurden erfüllt.	0 (0 %)	2 (5,6 %)	18 (50,0 %)	16 (44,4 %)
Ich kann die Peer-Schulung weiterempfehlen.	1 (2,8 %)	3 (8,3 %)	13 (36,1 %)	19 (52,8 %)
Ich fühle mich gut auf die Tätigkeit als Peer im Qualitätssicherungsverfahren vorbereitet.	1 (2,8 %)	3 (8,3 %)	24 (66,7 %)	8 (22,2 %)
Insgesamt fand ich die Peer-Schulung gelungen.	0 (0 %)	2 (5,7 %)	17 (48,6 %)	16 (45,7 %)

Die Auswertung der offenen Fragen zeigte, dass den teilnehmenden Sozialmedizinerinnen die Möglichkeit zur Diskussion (n = 11), die Arbeit in den Kleingruppen (n = 10), die Art der Moderation (n = 5), der Einbezug von sozialmedizinischen Experten (n = 4) sowie die allgemeine Organisation (n = 6) besonders gut gefallen hat. Im Gegensatz dazu machten lediglich sechs Peers Angaben zu Dingen, die negativ aufgefallen seien. Bezogen auf Inhalt und Durchführung der Schulung wurde kritisiert, dass das erste Modul des Schulungscurriculums zu lang sei und die Einleitung komprimiert werden könne, die Experten bei der Auslegung des Manuals unsicher wirkten, es unmittelbar nach der Schulung noch Unklarheit über die Eingabe der Daten über EvaSys® gebe und dass die als Grundlage für einige Prüffragen des Manuals dienende DRV-Schrift 21 „Sozialmedizinisches Glossar“ nicht von Beginn an allen Peers zur Verfügung stand. Darüber hinaus wurde der „Formalismus des Konzeptes und die fehlende Relativierung der Einzelkriterien für die Brauchbarkeit des Gutachtens“ kritisiert. Infolgedessen werde die tatsächliche Güte relevanter Aspekte eines Gutachtens im Qualitätssicherungsverfahren nicht abgedeckt. Ferner wurde an dieser Stelle angegeben, dass die „Moderatoren häufig nicht moderieren“, die Schulung „langatmig und zäh“ sei und der Evaluationsbogen nicht ausführlich genug erklärt wurde.

Eine standardisierte Befragung der Schulungsmoderatoren wurde nicht durchgeführt. Das Peer-Training wurde mit den oben beschriebenen Inhalten und Zeitplanungen als durchführbar befunden. Bis auf das Lernziel 2.5 konnten alle Lernziele vollständig umgesetzt werden. Lernziel 2.5 wurde aus Zeitmangel in einer Arbeitsgruppe nur teilweise erfüllt, infolgedessen nicht alle Prüffragen abschließend besprochen werden konnten.

5.3.4 Beurteiler-Übereinstimmung

Die prozentuale Peer-Übereinstimmung wies trotz der standardisierten Peer-Schulung eine ausgeprägte Heterogenität in der Übereinstimmung auf. Die Range der Interrater-Übereinstimmung lag zwischen 25,4 % und 100,0 % für die Prüffragen der Einzelkriterien und zwischen 69,8 % und 44,5 % für das übergeordnete Kriterium (Tab. 62). Gemittelt über alle Prüffragen und Gutachten betrug die prozentuale Übereinstimmung der Peers 56,5 %.

Tabelle 62: Gutachtenspezifische prozentuale Übereinstimmung in Prozent (119 Reviews).

Prüffrage	PÜ Gutachten 1	PÜ Gutachten 2	PÜ Gutachten 3	PÜ Gesamt
Einzelkriterium A: Formale Gestaltung				
Prüffrage A.1	65,0 %	72,8 %	42,2 %	60, 3%
Prüffrage A.2	41,2 %	55,0 %	35,6 %	45,0 %
Einzelkriterium B: Verständlichkeit				
Prüffrage B.1	35,8 %	41,7 %	85,2 %	53,7 %
Prüffrage B.2	26,3 %	27,1 %	32,5 %	28,6 %
Prüffrage B.3	33,5 %	41,2 %	80,7 %	51,3 %
Einzelkriterium C: Transparenz				
Prüffrage C.1	41,9 %	48,8 %	90,0 %	59,8 %
Prüffrage C.2	66,3 %	34,2 %	67,6 %	55,8 %
Prüffrage C.3	39,6 %	26,0 %	40,2 %	35,2 %
Einzelkriterium D: Vollständigkeit				
Prüffrage D.1	25,4 %	37,8 %	72,2 %	44,7 %
Prüffrage D.2	72,8 %	55,5 %	72,6 %	66,9 %
Prüffrage D.3	80,9 %	27,1 %	56,7 %	54,8 %
Prüffrage D.4	90,3 %	100,0 %	64,8 %	85,4 %
Prüffrage D.5	28,8 %	26,5 %	56,7 %	37,0 %
Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen				
Prüffrage E.1	95,0 %	90,1 %	100,0 %	95,0 %
Prüffrage E.2	35,9 %	45,9 %	76,5 %	52,4 %
Einzelkriterium F: Wirtschaftlichkeit				
Prüffrage F.1	100,0 %	68,8 %	94,9 %	87,8 %
Prüffrage F.2	28,6 %	30,5 %	63,8 %	40,6 %
Übergeordnetes Kriterium				
Ampel- kriterium	67,2 %	44,5 %	69,8 %	60,5 %

Anmerkung: PÜ=Prozentuale Übereinstimmung.

Die heterogene Bewertung der Gutachten ließ sich auch exemplarisch an der Ampelbewertung des übergeordneten Kriteriums, dem Hauptzielkriterium des Verfahrens, veranschaulichen. In Abbildung 19 sind die prozentualen Anteile der drei Beurteilungsmöglichkeiten der Ampelbewertung für die drei Hausaufgabengutachten dargestellt, die von allen Peers beurteilt worden sind.

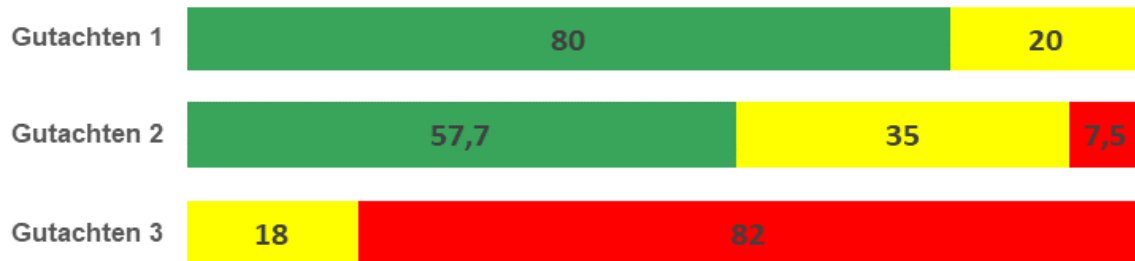


Abbildung 19: Varianz der Peerbeurteilungen des übergeordneten Kriteriums (in Prozent) der Hausaufgabengutachten.

Unterschied zwischen neu geschulten und erfahrenen Peers. Insgesamt 27 von 40 Peers wurden mit dem standardisierten Peer-Schulungsprogramm neu und erstmalig geschult und nahmen sowohl am Basis- als auch am Anwendermodul teil. Dreizehn Peers nahmen ausschließlich am Anwendermodul teil, da sie bereits im Vorgängerprojekt involviert waren. In Tabelle 63 wird die prozentuale Übereinstimmung der Peers separat nach „neuen“ und „erfahrenen“ Peers dargestellt.

Tabelle 63: Gutachtenspezifische prozentuale Übereinstimmung in Prozent (119 Reviews).

Prüf- frage	Gutachten 1			Gutachten 2			Gutachten 3		
	Ge- samt	„neue“ Peers	„erfah.“ Peers	Ge- samt	„neue“ Peers	„erfah.“ Peers	Ge- samt	„neue“ Peers	„erfah.“ Peers
A1	65,0	51,9	100,0	72,8	72,9	70,5	42,2	41,2	41,0
A2	41,2	46,4	32,1	55,0	61,5	39,7	38,6	37,5	35,9
B1	35,8	35,0	33,3	41,7	43,3	46,2	85,2	92,3	71,8
B2	26,3	25,9	21,8	27,1	25,9	24,4	32,5	31,4	30,8
B3	33,5	36,2	30,8	41,2	39,0	43,6	80,7	85,2	70,5
C1	41,9	39,6	43,6	48,9	47,0	57,7	90,0	92,3	84,6
C2	66,3	62,7	71,8	34,2	32,5	34,6	67,6	71,4	57,7
C3	39,6	41,9	33,3	26,0	25,4	21,8	40,2	41,2	47,4
D1	25,4	28,2	19,2	37,8	34,2	53,9	72,2	72,9	71,8
D2	72,8	67,5	84,6	55,5	58,4	48,7	72,6	78,8	59,0
D3	80,9	72,4	100,0	27,1	26,8	23,1	56,7	60,3	47,4
D4	90,3	85,8	100,0	100,0	100,0	100,0	64,8	60,3	71,8
D5	28,9	31,3	23,1	26,5	25,1	24,4	56,7	49,9	70,5
E1	95,0	100,0	84,6	90,1	100,0	70,5	100,0	100,0	100,0
E2	35,9	32,5	39,7	45,9	48,2	39,7	76,5	85,2	59,0
F1	100,0	100,0	100,0	68,9	61,5	84,6	94,9	92,3	100,0
F2	28,6	25,6	35,9	30,5	25,9	39,7	63,9	71,4	50,0
Ampel	67,2	60,1	84,6	44,5	41,9	50,0	69,8	78,8	53,9

Anmerkung: Die graue Farbmarkierung kennzeichnet die Prüffragen, bei denen die geschulten neuen Peers höhere Übereinstimmungswerte erreichten als die erfahrenen Peers.

Tabelle 63 beschreibt die prozentuale Übereinstimmung der Gutachter je Prüffrage. Der Vergleich der neuen und erfahrenen Peers zeigte, dass die neu geschulten Peers in den Prüffragen der Einzelkriterien in 28 von 51 Fällen (54,9 %) höhere prozentuale Übereinstimmungswerte aufwiesen als die erfahrenen Sozialmediziner. Bezüglich des übergeordneten Kriteriums, welches das Hauptzielparameter des Review-Verfahrens darstellte, wiesen die erfahrenen Peers in zwei von drei Fällen (66,7 %) eine höhere prozentuale Übereinstimmung auf. Über alle beurteilten Gutachten ist dieser Unterschied von im Mittel 2,6 % jedoch zu vernachlässigen.

Über alle Prüffragen der Einzelkriterien betrug die Abweichung der neuen von den erfahrenen Peers im Mittel nur 0,6 %. Die stärkste Divergenz von mehr als 48 % zeigte das Gutachten 1 bei Prüffrage A1. Die Range der Differenzen beider Gruppen lag für die Prüffragen der Einzelkriterien zwischen 0 % und 48,1 % und für das übergeordnete Kriterium zwischen 8,1 % und 24,9 %.

5.3.5 Beurteiler-Reliabilität

Für standardisierten Peer-Schulung ergab sich eine Prozentübereinstimmung aller möglichen Rater-Paare in Höhe von 61 %. Zwei zufällig ausgewählte und voneinander unabhängig urteilende Peers würden demnach bei Beurteilung der Gutachten zu 61 % in ihren Urteilen übereinstimmen. Der Werte für die Beurteilerübereinstimmung unterschieden sich zwar nominal, aber nicht signifikant von der erwarteten zufälligen Übereinstimmungsrate von 36 % ($\kappa_m = 0,39$, $p > 0,05$, siehe Tab. 64).

Tabelle 64: Beurteilerübereinstimmung mittels Kappa-Koeffizient von Fleiss (κ_m) für mehrere Beurteiler (117 Reviews)⁹.

Gutachten	Übergeordnetes Kriterium			Statistische Kennwerte		
	grün (n)	gelb (n)	rot (n)	p_i	p_e	κ_m
#1	32	7	0	0,70		
#2	3	13	23	0,45		
#3	0	8	31	0,67		
standardisierte Peer-Schulung (n = 39)				p = 0,61	0,36	0,39

Anmerkung. p_i = Anzahl der in ihren Urteilen übereinstimmenden Urteilerpaare in Bezug zur maximalen Anzahl übereinstimmender Urteilerpaare. p = Wahrscheinlichkeit, mit der die Gutachten eines beliebigen Urteilerpaares im Durchschnitt identisch beurteilt wurden. p_e = Wahrscheinlichkeit, mit der die Gutachten eines beliebigen Urteilsaares bei Zufall übereinstimmen würden.

⁹ Daten basieren auf n = 39 Peers, 1 missing. Ein Peer wurde aus der Analyse ausgeschlossen, da nicht alle drei Gutachten vollständig beurteilt wurden.

5.3.6 Übereinstimmung mit dem Peer und Experten-Konsens

Peer-Konsens. Im Gegensatz zur prozentualen Übereinstimmung sowie dem Kappa-Koeffizienten von Fleiss, die zwischen den Peers berechnet wurden, wurde bei der Übereinstimmung mit dem Peer-Konsens der Modus (der Wert, der am häufigsten zur Beurteilung einer Prüffrage genutzt wurde) bei der Beurteilung der Mängelkategorien als Referenzgröße herangezogen (Tab. 65).

Tabelle 65: Prozentuale Übereinstimmung mit dem Peer-Konsens (Modus) in Prozent (119 Reviews).

Prüf- frage	Gutachten 1			Gutachten 2			Gutachten 3		
	Konsens Ge- samt	Konsens „neue“ Peers	Konsens „erfah.“ Peers	Konsens Ge- samt	Konsens „neue“ Peers	Konsens „erfah.“ Peers	Konsens Ge- samt	Konsens „neue“ Peers	Konsens „erfah.“ Peers
A1	80,0	70,4	100,0	85,0	85,2	84,6	61,5	61,5	61,5
A2	60,0	48,1	84,6	72,5	77,8	61,5	53,8	53,8	53,8
B1	50,0	37,0	69,2	57,5	63,0	46,2	92,3	96,2	84,6
B2	40,0	37,0	38,5	40,0	40,7	38,5	38,5	34,6	46,2
B3	52,5	55,6	46,2	60,0	59,3	61,5	89,7	92,3	84,6
C1	60,0	59,3	61,5	65,0	59,3	76,9	94,9	96,2	92,3
C2	80,0	77,8	84,6	50,0	48,1	53,8	82,1	84,6	76,9
C3	47,5	51,9	38,5	37,5	37,0	38,5	51,3	42,3	69,2
D1	32,5	37,0	23,1	45,0	33,3	69,2	84,6	84,6	84,6
D2	85,0	81,5	92,3	70,0	74,1	61,5	84,6	88,5	76,9
D3	90,0	85,2	100,0	35,0	33,3	38,5	74,4	76,9	69,2
D4	95,0	92,6	100,0	100,0	100,0	100,0	79,5	76,9	84,6
D5	45,0	48,2	38,5	35,0	33,3	38,5	74,4	69,2	84,6
E1	97,5	100,0	92,3	95,0	100,0	84,6	100,0	100,0	100,0
E2	45,0	40,7	53,8	62,5	66,7	53,8	87,2	92,3	76,9
F1	100,0	100,0	100,0	82,5	77,8	92,3	97,4	96,2	100,0
F2	40,0	33,3	53,8	42,5	37,0	53,8	79,5	84,6	69,2
Ampel	80,0	74,1	92,3	57,5	51,9	69,2	82,1	88,5	69,2

Anmerkung: Die graue Farbmarkierung kennzeichnet die Prüffragen, bei denen die geschulten neuen Peers höhere Übereinstimmungswerte mit dem Peer-Konsens erreichten, als die erfahrenen Peers.

Insgesamt fielen die Übereinstimmungen mit dem Peer-Konsens regelhaft höher aus, als die berechnete Konkordanz zwischen den Peers. Die Tabelle differenziert erneut zwischen neu geschulten und erfahrenen Peers. Der Vergleich der Gruppen zeigte, dass die neu geschulten Peers in den Prüffragen der Einzelkriterien in 19 von 51 Fällen (37,3 %) höhere prozentuale Übereinstimmungswerte im Vergleich zum Modus aufwiesen als erfahrenere Sozialmediziner. Bezüglich des übergeordneten Kriteriums wiesen im Gegensatz dazu die erfahrenen Peers in zwei von drei Fällen (66,7 %) eine höhere prozentuale Übereinstimmung mit dem Peer-Konsens auf.

Experten-Konsens. Ähnlich zur Übereinstimmung mit dem Peer-Konsens wurde im Folgenden die prozentuale Übereinstimmung mit den konsentierten Lösungen aus den Musterantworten der Experten genutzt, um damit die Übereinstimmung der Peer-Urteile abzugleichen (Tab. 66).

Tabelle 66: Prozentuale Übereinstimmung mit dem Experten-Konsens in Prozent (119 Reviews).

Prüf- frage	Gutachten 1			Gutachten 2			Gutachten 3		
	Konsens Ge- samt	Konsens „neue“ Peers	Konsens „erfah.“ Peers	Konsens Ge- samt	Konsens „neue“ Peers	Konsens „erfah.“ Peers	Konsens Ge- samt	Konsens „neue“ Peers	Konsens „erfah.“ Peers
A1	80,0	70,4	100,0	85,0	85,2	84,6	61,5	61,5	61,5
A2	60,0	48,1	84,6	72,5	77,8	61,5	53,8	53,8	53,8
B1	50,0	37,0	69,2	2,5	3,7	0,0	92,3	96,2	84,6
B2	17,5	14,8	23,1	40,0	40,7	38,5	38,5	34,6	46,2
B3	52,5	55,6	46,2	10,0	14,8	0,0	89,7	92,3	84,6
C1	60,0	59,3	61,5	65,0	59,3	76,9	94,9	96,2	92,3
C2	17,5	18,5	15,4	50,0	48,1	53,8	82,1	84,6	76,9
C3	47,5	51,9	38,5	27,5	29,6	23,1	38,5	50,0	15,4
D1	17,5	11,1	30,8	2,5	3,7	0,0	84,6	84,6	84,6
D2	85,0	81,5	92,3	70,0	74,1	61,5	84,6	88,5	76,9
D3	90,0	85,2	100,0	32,5	33,3	30,8	74,4	76,9	69,2
D4	95,0	92,6	100,0	100,0	100,0	100,0	79,5	76,9	84,6
D5	45,0	48,2	38,5	10,0	11,1	7,7	74,4	69,2	84,6
E1	97,5	100,0	92,3	95,0	100,0	84,6	100,0	100,0	100,0
E2	40,0	40,7	38,5	7,5	7,4	7,7	87,2	92,3	76,9
F1	100,0	100,0	100,0	82,5	77,8	92,3	97,4	96,2	100,0
F2	40,0	33,3	53,8	32,5	29,6	38,5	79,5	84,6	69,2
Ampel	80,0	74,1	92,3	57,5	51,9	69,2	17,9	11,5	30,8

Anmerkung: Die grüne Farbmarkierung kennzeichnet die Prüffragen, an denen eine Abweichung zwischen Experten-Konsens und dem Peer-Konsens, d.h. der Kategorie, die von den meisten Peers gewählt wurde, vorliegt.

Bezüglich der Prüffragen der Einzelkriterien stimmte die prozentuale Übereinstimmung der Peer-Urteile zu 74,5 % (38 von 51 Fällen), und beim übergeordneten Kriterium zu 66,7 % (2 von 3 Fällen), mit den konsentierten Gutachtenreviews der DRV Bund überein. Die Nicht-Übereinstimmung erschien abhängig vom Gutachten zu sein. Während bei Gutachten 3 lediglich eine Prüffrage und das übergeordnete Kriterium andere Beurteilungen als der Peer-Konsens aufwiesen, bewerteten die Peers bei Gutachten 2 mehrheitlich acht von 17 Einzelkriterien abweichend zum Experten-Konsens.

5.3.7 Effektivität der standardisierten Schulung

Zur explorativen Überprüfung der Effektivität des standardisierten Peer-Schulungsprogramms wurden die Reviews der drei Gutachten, die im Anschluss der curricularen Schulung beurteilt wurden, mit den drei Reviews aus dem Vorgängerprojekt „EvaBegut“ verglichen, in dem keine standardisierte Schulung durchgeführt wurde. Um

Verzerrungseffekte zu minimieren, wurden nur die Peers in die Analyse der IG aufgenommen, die bisher nicht im Qualitätssicherungsverfahren oder dessen Evaluation involviert waren ($n = 26$).

Der Vergleich wurde mittels Fleiss Kappa am übergeordneten Kriterium dargestellt, da die Prüffragen der Einzelkriterien zwischen den Schulungszeitpunkten in Umfang und Wortlaut optimiert wurden. Eine direkte Vergleichbarkeit der Einzelkriterien war daher nicht möglich.

Die prozentuale Übereinstimmung für das übergeordnete Kriterium der IG, die nach der standardisierten Peer-Schulung bewertet wurden, betrug 78,8 % (Gutachten 1), 60,1 % (Gutachten 2) bzw. 41,8 % (Gutachten 3) mit durchschnittlich 60,2 %. Im Vergleich dazu wies die KG, die keine curriculare standardisierte Ausbildung erhielt, eine prozentuale Übereinstimmung von 52,9 % (Gutachten 4), 35,7 % (Gutachten 5) und 33,3 % (Gutachten 6) mit einem Durchschnitt von 40,2 % auf. Diese Differenz von 20 % zwischen IG und KG ist auf bei dem kritischen Alpha-Niveau von $p < 0.05$ nicht signifikant ($t(42) = 1.306$, $df = 42$, $p = 0,2$).

Die Interrater-Reliabilität des standardisierten Schulungscurriculums erreichte einen κ_m -Wert von 0,39, verglichen mit einem κ_m -Wert von 0,12 für die KG. Insgesamt belegen die Daten, dass unter beiden Untersuchungsbedingungen (IG/KG) die Übereinstimmung zwischen den Peers (p_i) höher war als durch den Zufall erwartet (p_e). Das beobachtete Wahrscheinlichkeitskonkordanzintervall war in der IG jedoch dreimal höher als in der KG (Tab. 67).

Tabelle 67: Beurteilerübereinstimmung mittels Kappa-Koeffizient von Fleiss (κ_m) im Vergleich von IG und KG (IG: 78 Reviews; KG: 54 Reviews).

Gutachten	Übergeordnetes Kriterium			Statistische Kennwerte		
	grün (n)	gelb (n)	rot (n)	p_i	p_e	κ_m
#1	23	3	0	0.79		
#2	2	10	14	0.59		
#3	0	7	19	0.42		
standardisierte Peer-Schulung (IG, n = 26)				$p = 0.60$	0.35	0.39
#4	12	6	0	0.53		
#5	3	5	10	0.38		
#6	3	9	6	0.35		
nicht- standardisierte- Peer-Schulung (KG, n = 18)				$p = 0.42$	0.34	0.13

Anmerkung. p_i = Anzahl der in ihren Urteilen übereinstimmenden Urteilerpaare in Bezug zur maximalen Anzahl übereinstimmender Urteilerpaare. p = Wahrscheinlichkeit, mit der die Gutachten eines beliebigen Urteilerpaares im Durchschnitt identisch beurteilt wurden. p_e = Wahrscheinlichkeit, mit der die Gutachten eines beliebigen Urteilsaares bei Zufall übereinstimmen würden.

5.4 Diskussion

5.4.1 Allgemeine Diskussion

Die Teilstudie beschreibt die Konstruktion und Evaluation eines standardisierten Peer-Schulungsprogramms für das trägerübergreifende Peer Review-Verfahren zur Qualitätssicherung der Sozialmedizinischen Begutachtung bei Anträgen zur EM-Rente. Die Schulung dient sowohl der Ausbildung von am Verfahren teilnehmenden sozialmedizinischen Sachverständigen als auch der Kalibrierung der Beurteilermaßstäbe aller teilnehmenden Peers. Das Ziel der dargelegten quantitativen Analysen bestand in der formativen und summativen Evaluation der Peer-Schulung. Da im Gegensatz zur vorangegangenen Teilstudie nur drei Gutachten von allen Peers beurteilt wurden, wurde als valider Evaluationskoeffizient die prozentuale Übereinstimmung und Fleiss Kappa als Maß der Interrater-Reliabilität berechnet.

Das in diesem Forschungsprojekt erarbeitete und evaluierte Schulungscurriculum beinhaltet im Sinne eines strukturierten Manuals differenzierte inhaltliche und methodische Durchführungshinweise für Moderatoren. Dadurch ist ein hohes Maß an Standardisierung, auch bei einem Wechsel der Schulungsdozenten, gewährleistet. Die Schulung setzt im Anwendermodul als didaktische Methoden ausschließlich Kleingruppenarbeit und Murmelgruppen ein und bildet zur Angleichung der Peer-Beurteilungen einen wesentlichen Bestandteil der Schulung. Die Konzeption der Schulung lässt offen, ob das Basis- und Anwendermodul zeitlich hintereinander oder getrennt durchgeführt werden. Dies bietet die Möglichkeit, das Anwendermodul als solitäre Fortbildungsveranstaltung vor jedem Peer Review-Verfahren anzubieten. Das Schulungscurriculum berücksichtigt weitgehend die Qualitätsmerkmale für (Patienten-)Schulungen nach Ströbl und Kollegen (2007). Es werden unterschiedliche didaktische Vermittlungsmethoden, geschlossene Gruppen und eine für das Anwendermodul begrenzte Teilnehmerzahl von zwölf Personen.

Formative Evaluation. Die formative Evaluation brachte keine nennenswerten inhaltlichen Verbesserungen für das Peer-Schulungscurriculum. Es wurde vereinzelt darauf hingewiesen, dass die in der Schulung beurteilten Gutachten nochmals durch eine sozialmedizinische Expertengruppe der DRV Bund auf ihre Eignung als Schulungsgutachten hin geprüft werden sollten. In der qualitativen Beurteilung des Schulungscurriculums und dessen Durchführung wurde daher die Auswahl der Schulungsgutachten auch am kritischsten bewertet. Ein wiederholter Einsatz dieser Gutachten ist jedoch nicht vorgesehen, da die Schulung, v.a. jedoch das Anwendermodul, offen für alle, d.h. neue und erfahrene Peers sein soll. Entsprechend müssen vor jeder Schulung sechs neue sozialmedizinische Gutachten von einer Expertengruppe der DRV

anhand der Prüffragen des Peer Review-Verfahrens beurteilt und deren Ergebnisse konsentiert werden.

Um die Auswahl der Schulungsgutachten zukünftig stärker zu standardisieren, bedarf es einer **Richtlinie zur Auswahl von Schulungsgutachten**. Die Gutachten müssen dahingehend speziellen Anforderungen genügen, um als Schulungsgutachten in der Schulung Verwendung zu finden. So sollten die Gutachten zur optimalen Anwendung des Prüffragenkatalogs nicht älter als drei Jahre sein, um zu gewährleisten, dass die aktuellsten Deck- und Schlussblätter genutzt werden. Durch die Beachtung dieses Kriteriums wird vermieden, dass Gutachten, die vor der Reform der DRV erstellt wurden, in die Peer-Schulung einbezogen werden. Ferner sollten in jeder Peer-Schulung jeweils ein im übergeordneten Kriterium als „grün“, „gelb“ und „rot“ beurteiltes Gutachten durch die Peers beurteilt und diskutiert werden. Da es sich beim Peer Review um ein indikationsübergreifendes Qualitätssicherungsverfahren handelt, sollten zudem die drei Gutachten, die im Anwendermodul beurteilt werden, aus den Hauptindikationen (Orthopädie/Innere/ Allgemeinmedizin/Neurologie/Psychiatrie) stammen. Insgesamt sollten zur Umsetzung dieses Vorgehens mindestens fünf bis sieben Gutachten im Vorfeld einer Peer-Schulung durch die Moderatoren und Co-Moderatoren konsentiert werden. Im Rahmen eines Vorbereitungswokshops sollten die drei Schulungsgutachten, sowie die drei Gutachten für die „Hausaufgabe“ im Anschluss der Schulung ausgewählt werden. Dieser vorbereitende Workshop sollte vor jeder Peer-Schulung obligat sein.

Ferner kann aus den Ergebnissen abgeleitet werden, dass bei der Durchführung der Schulung darauf geachtet werden sollte, das Lernziel 2.5 (Der Peer kann den Reviewprozess eigenständig durchführen und die Antworten der Prüffragen mit Hilfe des Informationsübermittlungssystem EvaSys® dokumentieren) zeitlich straff zu organisieren und länger andauernde Diskussionen zu begrenzen, da in einer Kleingruppe in der ersten Schulung in Würzburg aufgrund von Zeitmangel nicht alle Einzelkriterien in ihrer Ausführlichkeit besprochen werden konnten. Prinzipiell ist für das Lesen des Gutachtens sowie die Bearbeitung und Besprechung ausreichend Zeit im Schulungscurriculum eingeplant. Seitens der Teilnehmer wurde der gemeinsame Austausch zur Kalibrierung der Bewertungsmaßstäbe und zum Verständnis der Prüffragen sowie die gemeinsame Arbeit mit konkreten Beispielgutachten besonders positiv wahrgenommen. Ferner weist das Schulungscurriculum eine hohe Akzeptanz und eine gute Durchführbarkeit auf. Im Mittel wurde die Sicherheit im Umgang mit dem Manual mit 76,7 % als hoch eingeschätzt. Es ist davon auszugehen, dass die direkt nach der Schulung erhobenen Messwerte zur selbstberichteten Sicherheit im Umgang mit dem Manual nach Bearbeitung der „Hausaufgabe“ und regelmäßiger Peer-Tätigkeit weiter ansteigen werden.

Bei der Durchführung des Anwendermoduls konnte bei der Erstbearbeitung der Gutachten eine große Heterogenität in Bezug auf die Beurteilung der Prüffragen beobachtet werden (vgl. Anhang 10). Die Moderatoren waren meistens in der Lage, mittels der konsentierten Lösungen oder im Diskurs mit den Peers eine Homogenisierung der Bewertungsmaßstäbe herzustellen. Jedoch nicht in allen Fällen war es möglich einen Konsens zwischen den Peers zu erzielen. Zudem wurde bei einigen Prüffragen auch kein Konsens zu den durch die Expertengruppe im Vorfeld konsentierten Musterlösungen hergestellt. In diesen Fällen wurde in Absprache mit den Moderatoren der Peer-Konsens als gültige Beurteilung für die zu bearbeitende Prüffrage angenommen. In der Folge schlug sich dies in der formativen Evaluation in einer eher zurückhaltenden Beurteilung der Nützlichkeit der konsentierten Lösungen nieder.

Summative Evaluation. In der summativen Evaluation wurde ein eher heterogenes Bild der prozentualen Übereinstimmung mit einer großen Spannweite festgestellt. Die Interrater-Reliabilität $K_m = 0,39$ zeigt eine ausreichende (Landis & Koch, 1977) bzw. mittlere (Cohen, 1988, S. 82) Güte, erreicht jedoch nicht den allgemein anerkannten Standard von 0,7. Verglichen mit den publizierten Ergebnissen aus der Literatur und der Evaluation des Prüffragenkatalogs (Barth et al., 2017; Goldman, 1994; Harris & Bratzler, 2013; Kameoka et al., 2014; Neuderth, 2004; Strahl et al., 2016) liegt die Reliabilität der Peers jedoch in einer vergleichbaren Höhe.

Neben der allgemeinen Konkordanz wurde in der vorliegenden Teilstudie auch die Übereinstimmung mit dem Peer-Konsens und dem Experten-Konsens untersucht. Der Peer-Konsens beschreibt die Mangelkategorie, die die meisten Peers bei der Beantwortung einer Prüffrage genutzt haben. Dieser Konsens soll damit die über die Peers mehrheitlich determinierte „richtige“ Mangelkategorie zur adäquaten Beantwortung einer Prüffrage repräsentieren. Der Konsens könnte ein Indikator dafür darstellen, welche Mangelkategorie, die annähernd richtige ist, um die Prüffrage im Sinne des Manuals zu beantworten. Die prozentuale Übereinstimmung mit dem Peer-Konsens ist bei allen drei beurteilten Gutachten höher als mit der konsentierten Musterlösung der Experten. Es kann entsprechend davon ausgegangen werden, dass die Schulung einen positiven Effekt auf die interne Kongruenz der Beurteilungsmaßstäbe der Peers hat.

Um diesen Effekt statistisch darzustellen, wurde ein Studiendesign konstruiert, in dem die Reliabilität des standardisierten Peer-Schulungsprogramms mit der Reliabilität einer nicht strukturierten und manualisierten Peer-Schulung verglichen wurde. Da die teilnehmenden Sozialmediziner die Reviews neben ihrer Routinetätigkeit angefertigt haben und die Verteilung der Hausaufgabe lediglich zu einem Zeitpunkt an alle Peers möglich war, konnte kein echtes Parallelgruppen- oder Wartekontrollgruppendesign zur Schulungsevaluation

eingesetzt werden. Stattdessen wurden als KG die Reliabilitätskoeffizienten der Reviews definiert, die unmittelbar nach dem Schulungsworkshop der dargestellten Teilstudie 2 erstellt wurden. Die Übereinstimmung mit dem übergeordneten Kriterium war, unter standardisierten Trainingsbedingungen (61 %), fast 20 % höher als in der KG (42 %). Fleiss Kappa konnte zudem zeigen, dass die Peerübereinstimmung 25% höher ist als die zufällig zu erwartende Übereinstimmung von 36%. Obgleich die Beurteilungen nicht vollkommen übereinstimmten, kann dennoch davon ausgegangen werden, dass die Konkordanz höher ist, als es der Zufall zulassen würde, auch wenn dieser Unterschied aufgrund des geringen Stichprobenumfangs von $n = 3$ Gutachten nicht signifikant ist. In allen drei untersuchten Gutachten war die Übereinstimmung des übergeordneten Kriteriums der standardisierten Schulung höher als in der KG, was für die Effektivität der standardisierten Schulung spricht.

Bei der summativen Evaluation zeigte sich zudem bei Gutachten 3 die Besonderheit, dass die neu geschulten Peers sowohl im übergeordnetem Kriterium als auch bei zwölf der 17 Prüffragen der Einzelkriterien höhere Übereinstimmungswerte erreichten als die erfahrenen Peers (vgl. Tab. 63). Bei keinem der anderen Gutachten konnte dies in so ausgeprägter Form beobachtet werden. Gemittelt über alle Prüffragen der Einzelkriterien war die Übereinstimmung der neu geschulten Peers im Gegensatz zu den erfahrenen Peers um 11,7% höher. Die Range der erhöhten prozentualen Übereinstimmung lag zwischen 0,2% bei der Prüffrage A1 und 26,2% bei der Prüffrage E2. In Bezug auf die Nachvollziehbarkeit des Gutachtens (Ampelkriterium) erreichten die neu geschulten Peers eine 24,9% höhere Übereinstimmung als die erfahrenen Peers. Eine Subgruppenanalyse am Beispiel des übergeordneten Kriteriums zeigt, dass es sich bei diesem Gutachten insgesamt um ein „gutes“ Gutachtachten ohne wesentliche Brüche in der Argumentationskette handelt. Die beurteilenden Peers vergaben 32-mal die Mängelkategorie „grün“ (keine Mängel) und nur 7-mal die Kategorie „gelb“ (Mängel, die durch Sozialmediziner überbrückt werden können). Der Anteil der vergebenen Mängelkategorie „gelb“ war mit 31 % bei den erfahrenen Peers jedoch deutlich höher als bei den neu geschulten Peers mit 10 %, wodurch die Übereinstimmung in dieser Gruppe deutlich limitiert wurde.

Analog der Ergebnisse aus der qualitativen Inhaltsanalyse aus Teilstudie 2 der vorliegenden Arbeit, zeigt sich auch bei diesem Schulungsgutachten die Heterogenität der Peerbeurteilungen, obgleich dieselbe Mängelkategorie („gelb“) gewählt wurde. Die Peers bemängelten konkret:

- Fehlende Neutral-Null-Methode, sodass das Ausmaß der Einschränkung der Hüfte nicht eindeutig erkennbar wurde ($n = 1$),
- Fehlende Angaben darüber, dass keine relevante hirnorganische oder körperliche Beeinträchtigung durch den Alkohol vorliegt ($n = 2$),

- Fehlende Angaben darüber, warum trotz fehlender OP eine Leistungsfähigkeit von >6 Std. möglich ist (n = 1),
- Lückenhafte Argumentationskette durch fehlende Angaben zum Wirbelsäulenleiden in der Epikrise (n = 1),
- Lückenhafte Informationssammlung durch fehlende internistisch-apparative Untersuchungen (n = 1),
- Kleine Unklarheiten im Detail, z.B. warum bei Hüftgelenksdysplasie Tätigkeiten überwiegend im Gehen für leistungsgerecht gehalten werden oder warum beim negativen Leistungsbild Tätigkeiten mit Notwendigkeit des Besteigens von Leitern und Gerüsten nicht aufgenommen wurden (n = 1).

Die Auflistung der freitextlichen Mängeldokumentationen belegt an dieser Stelle erneut die Komplexität bei der Beurteilung des übergeordneten Kriteriums. Das Kriterium stellt sehr hohe Anforderungen an die Fähigkeiten eines Peers. Bereits im Pilotprojekt „EvaBegut“ wurde angemerkt, dass das übergeordnete Kriterium ein relevanter Teil des Qualitätssicherungsverfahrens darstellt, der jedoch schwer zu operationalisieren sei. Im Rahmen der Schulungsevaluation zeigte sich, dass die Peers in ihren Urteilen häufig übereinstimmten, Abweichungen jedoch nicht gänzlich auszuschließen sind und diese Abweichungen auch unterschiedlich begründet werden.

Limitationen. Diese Studie untersucht den Nachweis der Wirksamkeit einer standardisierten Peer-Schulung eher unter Alltagsbedingungen in einem wirklichkeitsnahen Umfeld (effectiveness) als unter strengen Laborbedingungen (efficacy). Entsprechend konnten Störvariablen nicht beeinflusst werden (Haynes, 1999; Kim, 2013). Dies beeinflusst auch das generelle Studiendesign. Die KG unterschied sich von der IG dadurch, dass sie ein nicht-standardisiertes Peer-Training auf Basis eines anderen Manuals zum Peer Review-Verfahren erhalten hat. Da sich das übergeordnete Kriterium zum Manual aus dem Jahr 2013 nicht verändert hat, wurde die Evaluation vorrangig anhand dieses Kriteriums durchgeführt. Zudem hat die KG und die IG verschiedene sozialmedizinische Gutachten beurteilt. Um eine Vergleichbarkeit zu gewährleisten, müssten beide Gruppen die gleichen Gutachten verwendet haben. Aus organisatorischen Gründen wurden jedoch alle Teilnehmer mit der standardisierten Schulung geschult, da unmittelbar nach der Evaluation mit der routinemäßigen Umsetzung des Peer-Review-Prozesses begonnen wurde. Eine Um- oder Nachschulung der Teilnehmer wäre nicht möglich gewesen. Daher verwendete die KG, die zu einem anderen Zeitpunkt anhand einer nicht-standardisierten Schulung geschult wurde, unterschiedliche Gutachten. Prinzipiell sind die Ergebnisse der summativen Evaluation entsprechend nicht direkt miteinander vergleichbar, da die Ergebnisse nicht in einem randomisierten Parallelgruppendesign erhoben wurden (siehe

Kap. 5.5). Die Unterschiede in den herangezogenen Gutachten könnten entsprechend zu einer unterschiedlichen Übereinstimmung zwischen den Peers führen. Da jedem Gutachten eine individuelle Krankengeschichte zugrunde liegt, können diese nicht direkt miteinander verglichen werden. Dementsprechend wird die Reliabilität der Schulungen nur indirekt verglichen und interpretiert, da die Variabilität auf die Berichte selbst zurückzuführen sein kann. Verzerrungen durch die Verwendung unterschiedlicher medizinischer Berichte in den untersuchten Gruppen können nicht endgültig ausgeschlossen werden.

Eine weitere grundlegende Einschränkung des Trainingserfolgs könnte in den zu prüfenden Gutachten selbst begründet sein, unabhängig davon, dass IG und KG unterschiedliche Gutachten beurteilt haben. In Deutschland wird die Arbeitsfähigkeit direkt durch Sozialmediziner im Auftrag der DRV beurteilt. Bei der Begutachtung kann es vereinzelt zu Abweichungen zwischen der formalen Definition der zu bewertenden Arbeitsfähigkeit und den tatsächlichen Kriterien der Gutachter kommen. Das Restleistungsvermögen wird determiniert durch die Anzahl der Stunden pro Tag, die ein Versicherter noch arbeiten kann (>6 Stunden, 3-6 Stunden oder <3 Stunden). In der Praxis kann diese strenge Klassifizierung nicht immer auf der Grundlage von funktionalen Einschränkungen quantifiziert werden. Dementsprechend ist es möglich, dass der Mediziner bei der persönlichen Untersuchung im Einzelfall und nach einer persönlichen Faustregel über die Arbeitsfähigkeit entscheidet (Aurich-Beerheide & Brüssig, 2017; Geiger, Garthwaite, Warren & Bamba, 2017). Wenn die Beurteilung der Arbeitsfähigkeit nicht objektiv nachvollziehbar ist, können Peers nur erschwerte einheitliche Bewertungen im Rahmen des Peer Reviews oder der Peer-Schulung vornehmen, was sich in niedrigen Reliabilitätskennwerten ausdrückt.

Darüber hinaus beinhaltet das standardisierte Peer-Schulungsprogramm keine Überprüfung der Lernziele. Die nach der Bloom'schen Taxonomie entwickelten, konkret zu überprüfenden Lernziele könnten direkt im Anschluss der Schulung mittels eines Wissenstests geprüft werden. Während und nach der Schulung erfolgte jedoch keine direkte Lernzielkontrolle mit den teilnehmenden Peers. Vielmehr basiert die Ausbildung auf der gemeinsamen Anwendung der Qualitätskriterien unter Aufsicht von zwei Moderatoren. Eine zusätzliche Überprüfung des Grundlagenwissens nach Abschluss des Basismoduls hätte dahingehend Vorteile, den Schulungserfolg direkt am Peer evaluieren zu können und den Peers dadurch die an sie gestellten Erwartungen zu vermitteln. Insgesamt könnte eine eingesetzte Lernzielkontrolle die Fähigkeit zur Anwendung und Interpretation der Prüffragen verbessern (Scriven, 1972, S. 77).

Für das Anwendermodul scheint eine Lernzielkontrolle kaum umsetzbar. Die Lernziele dieses Schulungsabschnittes basieren ausschließlich auf der praktischen Anwendung der

Review-Prüffragen in Kleingruppen und können nicht standardisiert getestet werden. Eine stringendere Operationalisierung der Lernziele und der untergeordneten Lehrinhalte erscheint für die zukünftige Weiterentwicklung der Schulung notwendig zu sein. Im Basismodul untergliedern sich die Lernziele in spezifische Lehrinhalte, die wiederum in Lernziele überführt werden können, um eine Erfolgskontrolle an mehreren Stellen im Training zu ermöglichen. Lernziel 1.2 (Der Peer kennt Inhalt und Aufbau des Manuals zum Peer Review-Verfahren) könnte somit in mehrere Teilziele untergliedert werden:

- **Lernziel 1.2.1:** Der Peer kann den hierarchischen Aufbau der Qualitätskriterien im Manual zum Peer Review-Verfahren erklären,
- **Lernziel 1.2.2:** Der Peer kann den Aufbau einer Prüffrage erklären,
- **Lernziel 1.2.3:** Der Peer kennt die Abstufungen der Mängelkategorien und kann die verschiedenen Möglichkeiten der Mängelabstufungen (quantitative Steigerung der Mängel, Sozialmedizinische Relevanz der Mängel, UND/ODER Verknüpfungen von Mängeln) beschreiben,
- **Lernziel 1.2.4:** Der Peer kann zu jeder Prüffrage erklären, ob sich diese auf einen spezifischen Gutachtenbereich bezieht (z. B. Frage B2 „Frage bezieht sich nur auf Diagnose und Epikrise“) oder ob bestimmte Aspekte explizit ausgeschlossen werden und nicht beurteilt werden müssen (z. B. Frage A1 „inhaltliche Aspekte [...] sind hier nicht Gegenstand der Bewertung“).

Die Aufgliederung würde eine Erfolgskontrolle an mehreren Stellen im Training ermöglichen, was mit einer Verbesserung der Übereinstimmung verbunden sein könnte.

Eine weitere Limitation ist der praktikablen Umsetzung der Schulung geschuldet. Insgesamt wurden vier Kleingruppen mittels unterschiedlicher Moderatoren geschult. Im Anhang 10 kann abgelesen werden, dass die unterschiedlichen Kleingruppen teilweise zu unterschiedlichen Konsensergebnissen bei der Beurteilung der Schulungsgutachten kommen. Vor dem Hintergrund, dass der Experten-Konsens in einigen Fällen durch den Peer-Konsens aufgehoben wurde, lag es an der individuellen Gruppendynamik und Einstellung des sozialmedizinischen Moderators, wie mit diesen Inkonsistenzen umgegangen wurde. Die Ergebnisse weisen darauf hin, dass die zwischen den Kleingruppen teilweise unterschiedlichen Beurteilungen der Prüffragen existieren. Obgleich es nach der Kleingruppenphasen noch eine abschließende Plenumsrunde gab, ist nicht auszuschließen, dass die Bewertungsmaßstäbe zwischen den Gruppen adäquat kalibriert wurden.

5.4.2 Beantwortung der Fragestellung

Die Teilstudie 3 zur Entwicklung und Evaluation einer standardisierten Peer-Schulung im Rahmen der trägerübergreifenden Qualitätssicherung sozialmedizinischer Gutachten untersuchte die Fragestellungen, die in Tabelle 54 zusammengetragen wurden:

1. *Ist die Übereinstimmung der Peers nach dem Besuch einer standardisierten Peer-Schulung höher als nach einer nicht-strukturierten Peer-Schulung?*

Die Übereinstimmung der Peers, die an der standardisierten Peer-Schulung teilgenommen haben und zuvor keine Erfahrung im Umgang mit dem Manual zum Peer Review-Verfahren hatten, zeigten bei der Beurteilung von drei sozialmedizinischen Gutachten eine 20 % höhere Übereinstimmung als die Peers der KG, die zuvor an einer unstrukturierten Schulung im Rahmen eines Projektworkshops teilgenommen hatten. Auch die Interrater-Reliabilität Fleiss Kappa war in der IG ($K_m = 0,39$) höher als in der KG ($K_m = 0,13$). Anhand der o.g. Limitationen sind bei dieser Berechnung Verzerrungseffekte nicht gänzlich auszuschließen, sodass anhand des vorliegenden Studiendesigns keine eindeutigen Aussagen zu dieser Fragestellung gemacht werden können.

2. *Wie hoch ist die prozentuale Übereinstimmung mit dem Peer-Konsens?*

Der Konsens beschreibt einen Indikator für die annähernd richtige Beantwortung einer Prüffrage im Sinne des Manuals. Er wird aus dem Modus aller Peers abgeleitet, die dieselbe Prüffrage zum selben sozialmedizinischen Gutachten bearbeitet haben. Über alle Gutachten und Prüffragen der Einzelkriterien gemittelt, betrug die prozentuale Übereinstimmung mit dem Peer-Konsens 67,9 % (Median: 70 %). Für das übergeordnete Kriterium betrug die Übereinstimmung mit dem Peer-Konsens 73,2 % (Median: 80 %). Damit weist das standardisierte Peer-Schulungscurriculum insgesamt eine gute interne Kongruenz der Beurteilungsmaßstäbe der Peers auf.

3. *Wie ist die Umsetzbarkeit der standardisierten Peer-Schulung zu beurteilen?*

Die formative Evaluation hat keine gravierenden Veränderungsbedarfe an der Peer-Schulung feststellen können. In der zeitlichen Umsetzbarkeit zeigten sich über die neun definierten Lernziele fast keine Defizite. Lediglich Lernziel 1.5 (Der Peer kann den Reviewprozess eigenständig durchführen und die Antworten der Prüffragen mit Hilfe des Informationsübermittlungssystems EvaSys® dokumentieren) konnte in einer von vier Kleingruppen nicht vollständig umgesetzt werden.

4. *Wie hoch ist die Akzeptanz der standardisierten Peer-Schulung?*

Die formative Evaluation spricht für eine sehr hohe Akzeptanz der standardisierten Peer Schulung. Die durchschnittliche Bewertung der Qualität der Durchführung der Schulungsmodule sowie der Schulung selbst lag im guten Bereich ($1,4 < x < 2,5$). Am besten und hilfreichsten wird die Arbeit in den Kleingruppen zur Beurteilung der Schulungsgutachten und zum Erfahrungsaustausch bewertet. Insgesamt 94,3 % der Peers beschrieben die Schulung als gelungen und 88,9 % würden diese für andere Peers als Einstieg in die Qualitätssicherung weiterempfehlen. Darüber hinaus zeigt die Evaluation, dass sich unmittelbar nach der Schulung 88,9 % der teilnehmenden Sozialmediziner gut auf die Tätigkeit als Peer im Qualitätssicherungsverfahren vorbereitet fühlten.

5.5 Handlungsempfehlungen

Kontinuierliche Peer-Schulungen. Mit Abschluss der Teilstudie liegt ein standardisiertes Peer-Schulungscurriculum für das Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung für EM-Renten mit differenzierten Durchführungshinweisen vor. Die Ergebnisse weisen darauf hin, dass die Kleingruppenarbeit als zentraler Bestandteil zur Angleichung der Peer-Beurteilungen wahrgenommen werden. Die intensive Diskussion und Auseinandersetzung mit den Prüffragen im Team ermöglichte es, Antworttendenzen zur Diskussion zu stellen und abweichende Peerbeurteilungen zu korrigieren. Es wird daher empfohlen, langfristig regelmäßige Peer-Schulungen, im Sinne des Schulungscurriculums, durchzuführen und zu evaluieren. Da aus der Evaluation des Prüffragenkatalogs (siehe Teilstudie 2) deutlich wurde, dass die Konkretisierung der Prüffragen und des Bewertungsschemas Grenzen haben, die u.a. im individuellen Begutachtungsgeschehen begründet sind, bietet ausschließlich die Peer-Schulung einen Rahmen, um im Austauschprozess mit anderen Peers die Antworttendenzen bei spezielle Gutachtenkonstellationen zu kalibrieren. Wenn seitens der DRV ein ausreichender Bestand an aktiven Peers aufgebaut wird, kann das Anwendermodul als regelmäßiger „Kalibrierungsworkshop“ für aktiv tätige Peers einmal im Jahr angeboten werden.

Re-evaluation der Schulungsgutachten. Die summative Evaluation zeigte, dass der Peer-Konsens sowohl bei den drei Schulungsgutachten als auch bei den drei Hausaufgabengutachten in einigen Fällen vom Experten-Konsens abwich. Auch wenn im Rahmen einer wiederholten Schulung die einmal eingesetzten Gutachten nicht wiederverwendet werden, sollte die Expertengruppe die sechs sozialmedizinischen Gutachten anhand der Beurteilungen und freitextlichen Mängeldokumentationen der Peers nochmals sichten und auf Abweichungen prüfen. Die erneute Beschäftigung mit den bereits konsentierten Gutachten bietet auch für die Expertengruppe die Möglichkeit, eigene Beurteilungsmaßstäbe neu zu kalibrieren und die damit gewonnenen Erfahrungen für die Konsentierung kommender Schulungsgutachten zweckdienlich einzubringen.

Lernzielkontrolle. Aus den Limitationen der Studie gehen weitere Handlungsempfehlungen hervor. Das Schulungscurriculum enthält keine Lernzielkontrolle, sondern untersuchte im Rahmen der formativen Evaluation die subjektive Angabe zur Sicherheit im Umgang mit den Einzelkriterien und dem übergeordneten Kriterium. Auf einzelne Prüffragen des Manuals wurde im Rahmen der Evaluation nicht näher eingegangen. Da eine Lernzielkontrolle im gutachtenabhängigen Anwendermodul im Sinne eines Wissenstestes nicht umsetzbar ist, könnten alle neuen Peers nach Abschluss des Basismodul theoretisch zu Inhalt und Anwendung der Prüffragen befragt werden.

Erweiterung der Lernziele. Der Schulungsumsetzung ist zudem zu entnehmen, dass der Konsens zur Kalibrierung der Bewertungsmaßstäbe innerhalb der zwei parallel stattfindenden Schulungsgruppen in einigen Prüffragen voneinander abweicht (vgl. Anhang 10). Obgleich in einem abschließenden Plenum mit allen an der Schulung teilnehmenden Peers offene Fragen und Anmerkungen besprochen wurden, erfolgte keine systematische Aufarbeitung dieser Divergenzen. Für eine notwendige kleingruppenübergreifenden Kalibrierung sollte das Anwendermodul um ein Lernziel 2.6 ergänzt werden, indem diese kleingruppenspezifischen Unterschiede im Plenum strukturiert dargestellt und diskutiert werden. Es ist zu vermuten, dass die Interrater-Konkordanz durch den bisher fehlenden gruppenübergreifenden Konsens weiter erhöht werden kann. Lernziel 2.6 würde den offiziellen Abschluss der Schulung darstellen, in dem die Schulungsmoderatoren die Ergebnisse der Kleingruppenarbeit systematisch im Plenum vorstellen und Abweichungen zur Diskussion stellen. Im bisherigen standardisierten Ablauf war dieses Lernziel nicht integriert. Didaktisch sollte das Lernziel aus einer Kombination aus Vortrag und (Plenums-)Diskussion aufgebaut sein. Ein vorläufiger Entwurf für das Lernziel 2.6 zur Integration in das Schulungsmanual wird in Abbildung 20 dargestellt. Das Lernziel ist nur umzusetzen, wenn mehr als eine Kleingruppe gleichzeitig geschult wurden.

Optimierung des Studiendesigns. Auch wenn die prozentuale Übereinstimmung mit dem Peer-Konsens auf eine gute interne Kongruenz der Peer-Beurteilungen schließen lässt, sollte die Evaluation der Schulung in einer weiteren Studie mit einem anderen Design nochmals durchgeführt werden. Der in dieser Studie verwendete Rückgriff auf die Schulungsgutachten der vorangegangenen nicht-standardisierten Peer-Schulung als KG erlaubt aufgrund der Diversität der Gutachten prinzipiell keinen Vergleich mit den aktuell erhobenen Reviews der Hausaufgabengutachten. In einem idealen RCT-Design würden die teilnehmenden Sozialmediziner in drei Gruppen im Verhältnis 1:1:1 aufgeteilt werden. Die IG würde mit der standardisierten Peer-Schulung die Handhabung mit dem Manual lernen, die KG₁ würde mit der unstrukturierten Schulung die Handhabung mit dem Manual lernen und KG₂ würde sich den Umgang mit dem Verfahren ausschließlich anhand des Manuals zum Peer Review-Verfahren beibringen. Im Anschluss könnten alle drei Gruppen dieselben sozialmedizinischen Gutachten beurteilen. Die mit einer geeigneten Fallzahl berechneten Reliabilitätskoeffizienten könnten endgültig Aufschluss über die Effektivität des standardisierten Peer-Schulungsprogramms geben.

Erwerb von Wissen	Erwerb von Handlungskompetenz
Lernziel	
1.6 Der Peer kalibriert seine Bewertungsmaßstäbe mit allen teilnehmenden Peers und wendet bei der Beurteilungen der Prüffragen der Einzelkriterien und des übergeordneten Kriteriums einheitliche Maßstäbe an	
Begründung	
Aufgrund der interaktiven didaktischen Umsetzung des Anwendermoduls, beschränkt sich die Teilnehmerzahl je Kleingruppe auf 12 Personen. Um bei mehreren parallel durchgeführten Kleingruppen für alle Peers gleiche Beurteilungsmaßstäbe herzustellen, bedarf es einer abschließende Diskussion kontrovers bewerteter Prüffragen im Plenum.	
Inhalt	
• Vergleichende Darstellung der Beurteilungen der Kleingruppenarbeit vom ersten Schulungstag des Anwendermoduls • Explizite Darstellung von Abweichungen zwischen den Kleingruppen – Divergenzen bei der Auswahl der Mängelkategorie – Divergenzen bei der Auswahl der Mängeldokumentation – Darstellung typischer Fallstricke bei der Beurteilung eines Gutachtens anhand von Beispielen aus den Kleingruppen, ggf. direkte Korrektur falsch angewendeten Prüffragen • Geleitete Plenumsdiskussion zur Kalibrierung der Bewertungsmaßstäbe kontroverser Peerbeurteilungen – Kalibrierung auf eine gemeinsame Beurteilung der Prüffragen – gesonderte Diskussion von Divergenzen bei der Auswahl der Mängelkategorie des übergeordneten Kriteriums als Hauptzielkriterium des Verfahren	
Didaktische Vermittlungsmethode	
Vortrag, Diskussion.	
Zeit circa 60 Minuten	
Anmerkung	
Da dieses Lernziel die Kalibrierung zwischen den Kleingruppen fokussiert, ist nur dann umzusetzen, wenn mehr als eine Kleingruppe im Rahmen des Anwendermoduls gleichzeitig geschult wurde.	

Abbildung 20: Entwurf des Lernziels 1.6 der standardisierten Peer-Schulung.

5.6 Kernbotschaft

Anhand der erweiterten Bloom'schen Taxonomie kognitiver Lernziele wurde ein standardisiertes Peer-Schulungsprogramm mit neun Lernzielen und interaktiver, didaktischer Methodik erarbeitet. Der Schulung liegt ein ausformuliertes Curriculum zugrunde, das es erlaubt, unabhängig vom Moderator das Programm anhand konkreter Umsetzungsanweisungen durchzuführen. Zentraler Bestandteil der Schulung ist das Anwendermodul zur Kalibrierung der Peer-Urteile. In diesem Modul beurteilen Peers im Rahmen von Klein-, Murrelgruppen und Einzelarbeit drei sozialmedizinische Gutachten anhand des Manuals zum Peer Review-Verfahren. In der formativen Evaluation erwies sich die Schulung als hilfreich und umsetzbar. Die Güte des Verfahrens (Interrater-Übereinstimmung und -Reliabilität) wurde mittels der prozentualen Übereinstimmung und Fleiss Kappa bewertet. Die Höhe des Fleiss Kappa-Koeffizienten liegt im Rahmen existierender Forschungsergebnisse und ist äquivalent mit der Höhe der Reliabilität der Teilstudie 2. Im Vergleich zur einer unstrukturierten Peer-Schulung zeigen sich Tendenzen, dass das strukturierte Schulungscurriculum die Reliabilität der teilnehmenden Peers erhöhen kann. Weitere Forschungsbemühungen in diesem Bereich sind jedoch notwendig, um die Effektivität der Schulung prospektiv nachzuweisen.

6 Zusammenfassende Diskussion

Gegenstand der Begutachtung. Die sozialmedizinische Beurteilung des Restleistungsvermögens ist eine wesentliche Aufgabe der Sozialmedizin. In der sozialmedizinischen Sachaufklärung sind vor allem die langfristigen oder dauerhaften Auswirkungen von Krankheit und Behinderung auf die Ausübung der Erwerbstätigkeit von besonderer Bedeutung (Derra, 2016, S. 71). Mit der steigenden Lebensarbeitszeit steigt voraussichtlich auch die Prävalenz chronischer Erkrankungen (Prütz et al., 2014, S. 123) und damit gleichzeitig die Zahl der Anträge zur Rehabilitation und EM-Rente (Bäcker et al., 2013, S. 54; Dannenberg et al., 2010; Deck & Pohontsch, 2011; Nowossadeck, 2018). Entsprechend wird der gesellschaftliche Stellenwert der Sozialmedizin als eigenständiges Fachgebiet der Medizin, das die multifaktoriellen Wechselwirkungen zwischen Gesundheit und Krankheit beschreibt und analysiert, in den nächsten Jahren stetig wachsen. Um bei der steigenden Zahl von benötigten Sozialmedizinischen Begutachtungen und Gutachtern die Gleichbehandlung aller Versicherten zu gewährleisten, bedarf es einer validen Messung des Leistungsvermögens auf Basis objektivierbarer Befunde.

Im Gegensatz zur Unfallversicherung wird bei der DRV nicht die Einschränkung der Funktionsfähigkeit im Sinne eines prozentualen Körperschadens, eines Grades der Schädigungsfolgen oder eines GdB geschätzt (vgl. Burggraf, 2009). Stattdessen wird das Restleistungsvermögen im Rahmen einer Sozialmedizinischen Sachaufklärung vor dem Hintergrund der aktuellen Leistungsfähigkeit in Bezug zur letzten Tätigkeit und auf dem allgemeinen Arbeitsmarkt quantitativ und qualitativ beurteilt. Untersuchungen beschreiben übereinstimmend, dass der komplexe und individuelle Begutachtungsprozess auf Basis der ICF durch verschiedene Störfaktoren einer gewissen Beurteilungsvarianz unterliegt und die Übereinstimmung bezüglich der sozialmedizinischen Leistungsbeurteilung häufig limitiert ist (Barth et al., 2017; Clark et al., 1988; Dickmann & Brooks, 2007; Elder et al., 1994; Ingravallo et al., 2008). Diese Ergebnisse werden gestützt durch die Aussagen von Linden (2007, S. 379), der beschreibt, dass die wissenschaftlichen Grundlagen der Entscheidungsfindung und -begründung in der Sozialmedizin gänzlich fehlen würden. Daraus resultiert, dass Mediziner bei der persönlichen Untersuchung teilweise nach einer persönlichen Faustregel über die Arbeitsfähigkeit entscheiden (Aurich-Beerheide & Brussig, 2017; Geiger et al., 2017).

Vor dem Hintergrund der komplexen Begutachtung und der Forderung zur Gleichbehandlung aller Versicherten durch alle RV-Träger ist die Etablierung einer einheitlichen, systematischen und trägerübergreifenden Qualitätssicherung ein zentrales Forderungsanliegen der DRV Bund, das bereits seit 2004 in Vorbereitung war (Cibis et al.,

2011, S. 108; VDR, 2004). Die Projektgruppe „Qualitätssicherung der Sozialmedizinischen Begutachtung“ legte den Grundstein für die konsekutiven Forschungsergebnisse der vorliegenden Arbeit.

Ziel der Dissertationsarbeit. Die vorliegenden Dissertationsarbeit hatte das Ziel, die dargelegten Defizite in der Versorgungslandschaft der Sozialmedizinischen Begutachtung zu beschreiben und anhand von drei Teilstudien valide Lösungsmöglichkeiten zu konzipieren. Dazu wurden konkret eine systematische Literaturrecherche zur Identifikation wissenschaftlich fundierter Literaturstellen zur Evidenzbasierung in der Sozialmedizinischen Beurteilung durchgeführt, die Evaluation eines trägerübergreifenden Qualitätssicherungsverfahrens der Sozialmedizinischen Begutachtung auf Basis eines Peer Review-Verfahrens abgeschlossen sowie eine Peer-Schulung für das trägerübergreifende Qualitätssicherungsverfahren entwickelt und evaluiert. Die Teilstudien basieren auf den durch die DRV Bund, Bereich Sozialmedizin, geförderten Drittmittelprojekten „Evaluation und Aufwandschätzung zum Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung“ sowie „Entwicklung eines Peer-Schulungsprogramms für das trägerübergreifende Qualitätssicherungsverfahren der sozialmedizinischen Begutachtung“. Da die Diskussion der Ergebnisteile aller Teilstudien bereits in den entsprechenden Kapiteln 3 bis 5 ausführlich diskutiert und methodenkritisch analysiert wurden, erfolgt im Folgenden eine zusammenfassende Diskussion der wichtigsten Erkenntnisse, um abschließend Implikationen für die Praxis und Forschung aufzuzeigen.

Systematische Literaturrecherche. Die Qualitätssicherung der Sozialmedizinischen Begutachtung basiert auf einem Peer Review-Verfahren von EM-Rentengutachten mit persönlicher Untersuchung. Die erste Teilstudie beschäftigte sich indirekt mit der Objektivität, Reliabilität und Validität sozialmedizinischer Gutachten. Vor der Prämisse, dass evidenzbasierte Informationen und Entscheidungsalgorithmen als festgelegter Begutachtungskorridor die unzufriedenstellende Übereinstimmung bezüglich der Beurteilung der beruflichen Leistungsfähigkeit erhöhen kann, wurde eine systematische Literaturrecherche zur Detektion evidenzbasierter Informationen in der Sozialmedizinischen Begutachtung durchgeführt. In einer Voruntersuchung wurde bereits beschrieben, dass die Evidenzbasierung in der Sozialmedizinischen Begutachtungsliteratur häufig auf Expertenkonsens „limitiert“ ist und keine höherwertigen Evidenzen gefunden wurden (Strahl, Rose, Brüggemann & Vogel, 2016a, 2016b; Strahl & Vogel, 2016). Da diese Recherche jedoch ohne jegliche Eingrenzung der Literatur und medizinischer Indikation durchgeführt wurde, hat die in dieser Arbeit beschriebene erste Teilstudie die zu untersuchende Population nach der Definition im PRISMA-Statement (Moher et al., 2011)

genauer eingegrenzt und die häufigsten Erkrankungen bei Beantragung einer EM-Rente in die Recherche eingeschlossen.

Nach Analyse der Datenbankergebnisse wurden 16 Publikationen in die systematische Literaturrecherche aufgenommen und deren Daten extrahiert. Die Studien lassen sich den drei Kategorien „Allgemeine und indikationsspezifische Hinweise zur Begutachtung“, „ICF und ICF Core Sets“ sowie „Begutachtung aus Perspektive der gutachterlich tätigen Ärzte“ zuordnen. Im Ergebnis wurden primär narrative Reviews und Beobachtungsstudien identifiziert (n = 10), die nach den Evidenzstufenmodell nach Kliche und Kollegen (2006) die beiden niedrigsten Evidenzklassen (Klasse IV und V) aufwiesen. In den identifizierten narrativen Reviews wurden zwar häufig zentrale zu begutachtende Bestandteile des sozialmedizinischen Gutachtens beschrieben, jedoch war die Evidenz und damit die Verallgemeinerbarkeit dieser Publikationen begrenzt.

Für die Qualitätssicherung könnten die dargestellten Analysen in Kombination mit den Ergebnissen von Strahl und Vogel (2016) bedeuten, dass die Objektivität von Gutachten eingeschränkt sein kann. Durch die Fülle an individuellen Fallkonstellationen in der Begutachtungssituation konnte keine Studie identifiziert werden, die die Ableitung der sozialmedizinischen Leistungsbeurteilung aus den Diagnosen, Befunden oder Epikrise verallgemeinerbar und evidenzbasiert darstellen kann. Zwar existieren standardisierte Vorlagen für den Aufbau eines sozialmedizinischen Gutachtens (VDR, 2001), jedoch werden Inhalt und Richtigkeit der Sozialmedizinischen Begutachtungsabfolge nicht geprüft. Es lassen sich auch keine Aussagen über die Durchführungsobjektivität der körperlichen Untersuchung treffen. Wenn die Durchführung einer Untersuchung anfällig für Störfaktoren ist, können keine fairen Untersuchungsbedingungen hergestellt werden (Rammstedt, 2010). Zudem wird die Höhe der Reliabilität bei niedriger Objektivität signifikant beeinflusst. Im Allgemeinen werden gute Reliabilitätswerte nur erreicht, wenn die Umstände der Testdurchführung bzw. Begutachtung standardisiert werden können (Schermelleh-Engel & Werner, 2008, S. 114).

Qualitätssicherung mittels Peer Review. Unter der Nutzung des von der DRV Bund entwickelten und von der Universität Würzburg manualisierten Prüffragenkatalogs zur Qualitätssicherung sozialmedizinischer Gutachten (vgl. DRV, 2009c; Löffler et al., 2010) wurde in der zweiten Teilstudie ein Peer Review-Verfahren zur Evaluation des Peer Review-Prüffragenkatalogs durchgeführt. Das Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung basiert auf einem Peer Review-Verfahren von Gutachten mit persönlicher Untersuchung, die infolge eines EM-Rentenanspruchs gestellt wurden. Ähnlich wie in den Studien zur Gutachter-Übereinstimmung bei der sozialmedizinischen Leistungsbeurteilung, zeigt auch die Studienlage zur Interrater-

Reliabilität der Peer-Übereinstimmungen niedrige Kennweite im Setting der Qualitätssicherung und scheint limitiert zu sein (Broder et al., 2002; Farin et al., 2003; Goldman, 1994; Hofer et al., 2004; Lilford et al., 2007; Neuderth, 2004; Smith et al., 1997; Strahl et al., 2016). Die Interrater-Übereinstimmung und -Reliabilität wurde unter der Perspektive des spezifischen Beurteilungskontextes von 0,4 bewertet. Mit einer Reliabilität von 0,39 überschritt Kendalls Konkordanzkoeffizient W für das übergeordnete Kriterium diese kritische Höhe nicht, lag jedoch im Rahmen existierender Forschungsergebnisse und war mit der Güte des Qualitätssicherungsverfahrens der Reha-Entlassungsberichte vergleichbar (vgl. Farin et al., 2003). Eine explorative qualitative Inhaltsanalyse der freitextlichen Mängeldokumentation zeigte, dass je nach Prüffrage gutachtenspezifische oder vorwiegend gutachtenübergreifende Mängelbegründungen detektiert wurden. Es konnte anhand deduktiver und induktiver Verfahrensweisen für jede Prüffrage charakteristische Mängelbegründungen identifiziert, codiert und quantifiziert werden. Aus der Evaluation des Prüffragenkatalogs gingen Empfehlungen für die Weiterentwicklung des Manuals zum Peer Review-Verfahren als auch für eine Peer-Schulung hervor.

Peer-Schulung. Die sich konsekutiv anschließende dritte Teilstudie beschreibt die Konstruktion und Evaluation eines standardisierten Peer-Schulungscurriculums. Tuijn und Kollegen (2012) konnten in einer Meta-Analyse zeigen, dass sowohl die Überarbeitung des Peer Review-Instruments als auch eine Intensivierung und Standardisierung der Peer-Schulung positive Auswirkungen auf die Interrater-Reliabilität haben kann. Ergänzende Studienergebnisse beschreiben zudem, dass die Interrater-Reliabilität vermutlich durch die Kommunikation zwischen den Peers erhöht werden kann (Hofer, Bernstein, DeMonner & Hayward, 2000; Strahl et al., 2016). Die Studie beinhaltete die Entwicklung sowie die formative und summative Evaluation einer Peer-Schulung. Das ausgearbeitete zweieinhalbtägige Peer-Schulungsprogramm besteht aus zwei Modulen und neun Lernzielen und wird als geschlossenes Gruppenangebot umgesetzt. Die Schulung basiert in großen Teilen auf interaktiven didaktischen Elementen wie Kleingruppenarbeit, Murmelgruppen, Diskussion und auch Einzelarbeiten. Die Ergebnisse der formativen Evaluation belegen, dass die Peers sich nach Besuch der standardisierten Peer-Schulung im Umgang mit den Prüffragen des Manuals als sicher einschätzen, die Möglichkeit den Prüffragenkatalog in der Kleingruppe zu bearbeiten und den kleingruppenorientierten Erfahrungsaustausch als sehr hilfreich empfinden sowie im Allgemeinen die Schulung als sehr gut organisiert und durchgeführt einschätzen. Bezogen auf die Hauptfragestellung im Rahmen der summativen Evaluation wurde gezeigt, dass die standardisierte Peer-Schulung ($K_m = 0,39$) im Vergleich zu einer zuvor unstrukturierten Schulung ($K_m = 0,13$) höhere Koeffizienten von Fleiss Kappa aufweist. Die Aussagekraft dieser

Studienergebnisse wird maßgeblich durch das zugrunde gelegte Studiendesign determiniert. Da die Peers der IG und KG unterschiedliche Gutachten zu unterschiedlichen Messzeitpunkten beurteilten, sind die beiden Schulungsbedingungen nur indirekt miteinander vergleichbar. Während der standardisierten Peer-Schulung zeigte sich zudem, dass ein Konsens innerhalb der teilnehmenden Sozialmediziner nicht immer herzustellen war. Auch wurde der Konsens zu den im Vorfeld durch eine Expertengruppe konsentierten Musterlösungen bei einigen Prüffragen nicht hergestellt. Stattdessen wurden die Musterlösungen innerhalb der Schulungen teilweise durch den Peer-Konsens revidiert.

Insgesamt veranschaulicht dies die gesamte Problematik der Qualitätssicherung der Sozialmedizinischen Begutachtung mittels des Peer Review-Verfahrens. Wenn, wie bereits dargelegt, die Beurteilung der Arbeitsfähigkeit nicht objektiv nachvollziehbar ist und das Review-Instrument bei einigen Prüffragen, v. a. dem übergeordneten Kriterium, viel Varianz in der Beurteilung zulässt, können Peers nur unter erschwerten Bedingungen im Rahmen des Peer Reviews valide und reliable Gutachtenbeurteilungen vornehmen.

Nichtsdestotrotz ist die Etablierung eines einheitlichen, trägerübergreifenden Qualitätssicherungssystems und die Erweiterung der Begutachtung mit der aktuell verfügbaren Evidenz ein erster Schritt, diese Defizite zu erfassen und systematisch zu beheben. Alle drei Teilstudien bearbeiten für sich genommen ein spezifisch abgrenzbares Themengebiet im Bereich der Sozialmedizinischen Begutachtung. Am Beispiel des Public Health Action Cycle von Kolip (2006) lässt sich das Ineinandergreifen der drei Teilstudien sehr anschaulich darstellen.

6.1 Einordnung in den Public Health Action Cycle

Der Public Health Action Cycle nach Kolip (2006) bestehend aus Problemdefinition, Strategieformulierung, Implementation/Umsetzung und Bewertung/Evaluation, wurde in den vorliegenden Teilstudien mehrfach vollständig durchlaufen (vgl. Abb. 8). Jede Teilstudie hat ihre eigene Problemdefinition, Strategie und Evaluation, aus der wiederum eine Problemdefinition abgeleitet werden kann.

Das Ausgangsproblem ist die in der Literatur beschriebene unzureichende Übereinstimmung von Gutachter- und Peerbeurteilungen, wenn zwei oder mehr Sozialmediziner denselben Patienten bzw. dasselbe Gutachten beurteilen. Linden (2007) begründet dies mit einer unzureichenden wissenschaftlichen Grundlage in der Entscheidungsfindung. Vor dem Hintergrund der in dieser Arbeit thematisierten Qualitätssicherung mittels des Peer Review-Verfahrens sollte die Konkordanz zwischen Gutachtern jedoch möglichst hoch sein, um objektive, reliable und valide Beurteilungen durchführen zu können. Die Strategieformulierung beschreibt die Planung einer systematischen Literaturrecherche zur Evidenzbasierung in der Begutachtung. Hintergrund dieser Recherche war die Annahme, dass evidenzbasierte Entscheidungsalgorithmen für die Begutachtung identifiziert werden können. Nach Umsetzung der systematischen Literaturrecherche in sechs wissenschaftlichen Datenbanken wurden die als relevant identifizierten Ergebnisse gesichtet und anhand der Fragestellung bewertet (Abb. 21). Da im Ergebnis keine höherwertigen Evidenzen identifiziert wurden, konnte kein allgemeingültiges Modell zur objektiven Ableitung des Restleistungsvermögens erarbeitet werden. Auch konnten anhand der Rechercheergebnisse keine Maßnahmen identifiziert werden, die Qualität der Gutachten selbst zu optimieren.

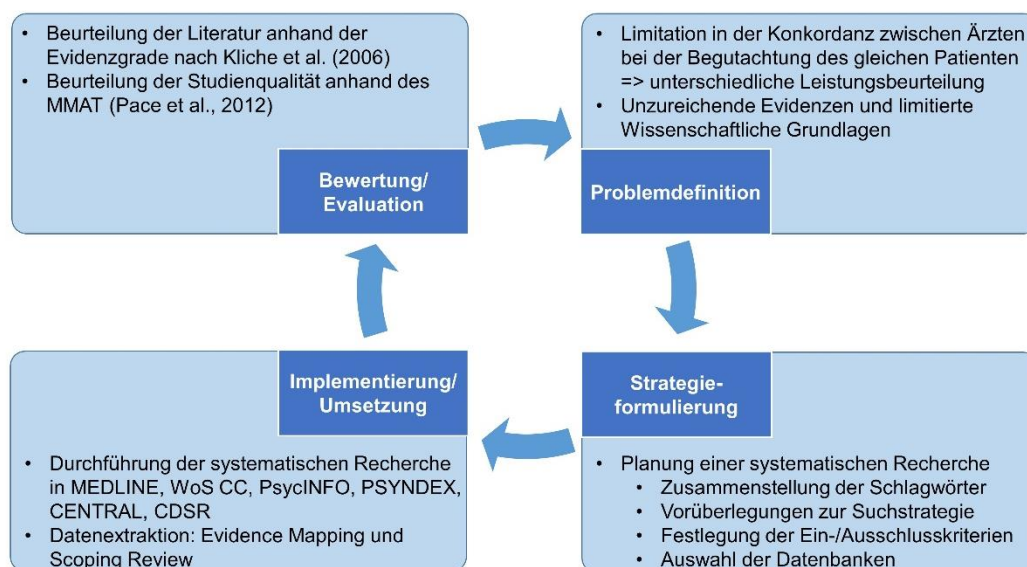


Abbildung 21: Einordnung der Teilstudie 1 in den Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).

In einem zweiten Durchlauf des Public Health Action Cycle wurde ein trägerübergreifendes Qualitätssicherungssystem für sozialmedizinische Gutachten evaluiert. Hier liegt die theoretische Annahme zugrunde, dass durch die externe Qualitätssicherung langfristig die Nachvollziehbarkeit, d.h. die Prozessqualität der Gutachten, verbessert werden kann. Die Problemdefinition der zweiten Teilstudie besteht daher wiederum aus der unzureichenden wissenschaftlichen Grundlage zur Gutachtenerstellung sowie aus der Fragestellung, wie reliabel das neue entwickelte Peer Review-Verfahren zur Qualitätssicherung der Gutachten sei. Im Rahmen der Strategieformulierung wurde ein Peer Review mit 260 Gutachten und 19 Peers geplant und in der Implementation/Umsetzung realisiert. Die Evaluation zeigte, dass die Interrater-Reliabilität der Peers limitiert ist, mit einer Höhe von $r_w = 0,39$ für das übergeordnete Kriterium jedoch mit anderen Studien vergleichbar sei (Abb. 22). Ob die Reliabilität durch die unzureichende Qualität der Review-Objekte oder aufgrund des Prüffragenkatalogs determiniert wurde, konnte nicht abschließend geklärt werden.

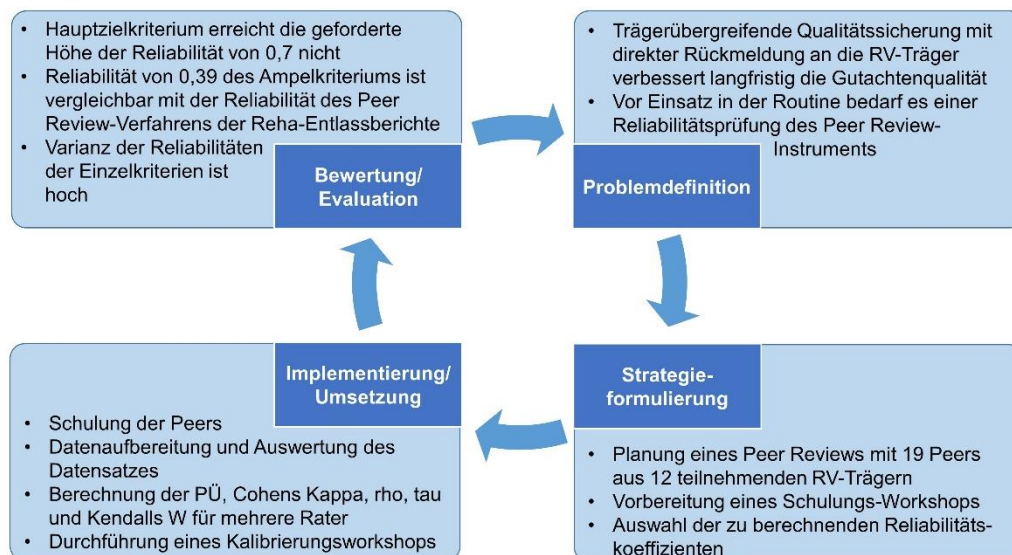


Abbildung 22: Einordnung der Teilstudie 2 in den Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).

Um die Interrater-Reliabilität der Peers im Rahmen der trägerübergreifenden Qualitätssicherungsprogramms der DRV zu optimieren, wurde im dritten Durchlauf des Public Health Action Cycle eine Peer-Schulung konzipiert und evaluiert, an der alle am Verfahren teilnehmenden Sozialmediziner verpflichtend teilnehmen müssen, wenn sie als Peer tätig sein möchten. Als Problem wurde die niedrige Reliabilität der Prüffragen aus dem Manual zum Peer Review-Verfahren festgelegt. Als Strategie zur Erhöhung der Reliabilität wurde die Schulung der Peers als Hauptzielparameter herangezogen. Im Rahmen der Studie wurde anhand der Taxonomie nach Bloom eine standardisierte, curriculare Peer-Schulung erstellt, mit 40 Peers durchgeführt und im Anschluss anhand von drei sozialmedizinischen Gutachten summativ evaluiert. Für die Evaluation wurde sowohl die

Übereinstimmung der Peers untereinander als auch mit dem Peer-Konsens und der Expertenbeurteilung berechnet. Darüber hinaus wurde im Rahmen einer „konstruierten“ Vergleichsbedingung die Reliabilität der neu geschulten Peers mit den Kennwerten derjenigen Peers verglichen, die im vorangegangenen Teilprojekt mittels einer unstrukturierten Peer-Schulung im Rahmen eines Schulungsworkshops mit dem Manual zum Peer Review-Verfahren vertraut gemacht wurden. Dabei erreichte die Interrater-Reliabilität des standardisierten Schulungscurriculums einen κ_m -Wert von 0,39, verglichen mit einem κ_m -Wert von 0,12 für die Vergleichsbedingung (Abb. 23).

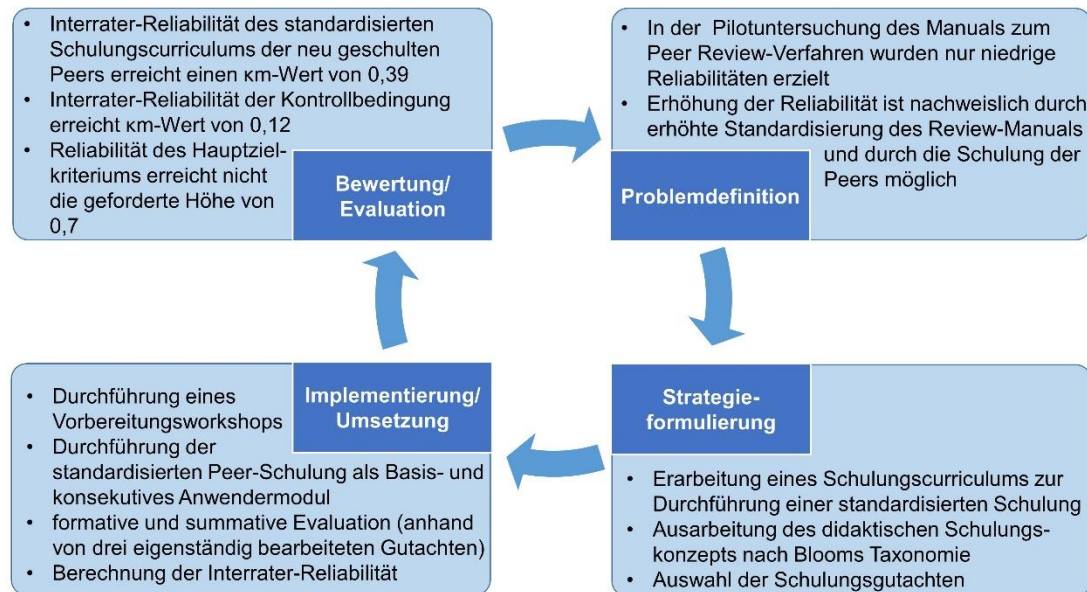


Abbildung 23: Einordnung der Teilstudie 3 in den Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).

Aus den Evaluationsergebnissen der drei Studien ergaben sich verschiedene Implikationen, für die (Schulungs-)Praxis, für das Manual zum Peer Review-Verfahren als auch für die Forschung die abschließend im Folgenden dargelegt werden.

6.2 Implikationen

Die Verbesserung der Qualität sozialmedizinischer Gutachten und die Entwicklung eines Qualitätssicherungssystems für die Sozialmedizinische Begutachtung war Teil der SOMEKO (VDR, 2004). Über einen Zeitraum von mehr als zehn Jahren hat die DRV, durch die Tätigkeit mehrerer Arbeitsgruppen und der Bewilligung verschiedener sozialmedizinischer Forschungsprojekte, die Qualität der Begutachtung optimiert. Die vorliegende Arbeit versuchte mittels einer multi-methodalen Herangehensweise, die sozialmedizinische Qualität sukzessiv zu steigern und damit sowohl die Gleichbehandlung aller Versicherten zu gewährleisten sowie die Arbeit der sozialmedizinischen Gutachter evidenzbasierter, standardisierter und damit transparenter zu gestalten. Alle vorgestellten Arbeitsschritte und Maßnahmen zielen darauf ab, sowohl direkt (über Evidenzbasierung der Begutachtung) als auch indirekt (über die Einflussnahme auf die gutachterliche Arbeitsweise durch ein trägerübergreifendes Qualitätssicherungssystem) die Qualität der Begutachtung zu verbessern.

Evidenz in der Sozialmedizin. Die Evidenz in der Sozialmedizinischen Begutachtung ist zum jetzigen Zeitpunkt stark begrenzt. Bereits zwei durchgeführte systematische Reviews kamen zu dem Ergebnis, dass für den Bereich der Begutachtung keine Studien mit höheren Evidenzleveln vorliegen [Stand: Dez. 2018]. Ebenfalls die von der DRV Bund zur Verfügung gestellten und zusätzlich publizierten Begutachtungsleitlinien besitzen ausschließlich ein S1-Evidenzniveau. Darüber hinaus empfiehlt die Rentenversicherung, selbst in Ausnahmefällen auf die subjektive, auf Erfahrungen beruhende, Meinung des Gutachters zu vertrauen, um ein Urteil über das Restleistungsvermögen zu fällen (Cibis et al., 2011, S. 100).

Auch gibt es nach Abschluss der Recherchen weiterhin keine evidenzbasierten Entscheidungspfade zur Begutachtung der Leistungsfähigkeit von Patienten mit gestelltem EM-Rentenantrag. Gleichzeitig ist an dieser Stelle die Überlegung anzustellen, welche Auswirkungen das auf die tägliche Arbeit der Sozialmediziner hat. Die Diskussion über die Verbindlichkeit von Leitlinien wird bereits seit mehreren Jahren geführt und findet auf beiden Seiten Befürworter und Gegner. Neben einer validen Entscheidung über die Gewährung einer EM-Rente würden evidenzbasierte Begutachtungsleitlinien das ärztliche und gutachterliche Handeln messbarer, transparenter und nachvollziehbarer gestalten und letztendlich die Qualität der Begutachtung im Wesentlichen verbessern (Leigemann, Lang, Kunz & Antes, 2005). Die Kehrseite der Leitliniennutzung besteht jedoch darin, dass das ärztliche Handeln u. U. restringiert wird und die ärztliche Erfahrung an Gewichtung verliert. Zudem müssen Leitlinien regelmäßig geprüft und aktualisiert werden. Ist eine Leitlinie veraltet, obliegt es dem behandelnden bzw. dem begutachtenden Sozialmediziner

abzuwägen, ob er dieser folgt oder sich aufgrund sachlicher Gründe oder persönlicher Erfahrungen für eine Nichtbefolgung entscheidet (Ulsenheimer, 2001).

Im Falle der Sozialmedizinischen Begutachtung bei EM-Rentenanträgen beinhalten die S1 Begutachtungsleitlinien der DRV Bund störungsspezifische Informationen über Auswirkungen einer Erkrankung, mögliche betroffene Bereiche im Sinne der ICF sowie Informationen zur Befundung. Der Unterschied zu evidenzbasierten Therapie- oder Diagnostikleitlinien besteht darin, dass der Gutachter die erhobenen Informationen zu einem sinnvollen Bild verknüpfen muss. Wie bereits im Theorieteil der vorliegenden Arbeit ausführlich dargestellt, ist die Variabilität und Individualität jedes Begutachtungsfalles so einzigartig, dass eine pauschale Ableitung des Restleistungsvermögen anhand einiger weniger Kriterien nicht möglich erscheint. Entsprechend dieser Prämisse wurden auch durch die Literaturrecherche keine evidenten, spezifischen Entscheidungsalgorithmen für die Begutachtung identifiziert.

Die narrativen Reviews von Kolenda (1998) und Kellner (2002) leiten aus dem Vorhandensein bestimmter Symptomkomplexe in Kombination mit spezifischen erhobenen Befunden einen GdB ab. Auf das vorhandene Restleistungsvermögen wird in diesen Publikationen jedoch nicht eingegangen. Auch die identifizierten Publikationen zu Prädiktoranalysen oder zur ICF bzw. ICF-Core Sets generieren keine Entscheidungshilfen bei der Leistungsbeurteilung, sondern ergänzen bestenfalls die bereits bestehenden S1 Leitlinien in Bezug auf die Identifikation oder Ergänzung relevanter und zwangsläufig zu erhebender Aspekte und Umstände einer chronischen Grunderkrankung.

Aus den vorliegenden Ergebnissen resultiert daher die Empfehlung, dass der globale Ansatz zur Identifikation evidenter Entscheidungshilfen langfristig fallen gelassen werden sollte. Stattdessen sollten vorrangig die solitären Sequenzen der Begutachtung (die Glieder der Argumentationskette) als eigenständige Forschungsobjekte untersucht werden. Wie im Kapitel 3.5 dargelegt, könnte mit diesem methodischen Ansatz die Evidenzbasierung in einzelne Bestandteile der Begutachtung integriert werden.

Qualitätssicherung und Qualität. Studien zeigen, dass medizinische Experten bezüglich der Beurteilung der beruflichen Leistungsfähigkeit häufig nicht übereinstimmen (Barth, 2017; Clark et al., 1988; Dickmann & Broocks, 2007; Elder, Symington & Symington, 1994; Ingravallo et al., 2008). Die Gründe für die reduzierte Konkordanz sind vielschichtig und liegen beim Gutachter, beim Patienten sowie beim Gegenstand der Begutachtung – das schwer zu erfassende Restleistungsvermögen – selbst. Um die Begutachtung zu harmonisieren, zu vereinheitlichen und transparenter zu gestalten, bedarf es einer grundlegenden anerkannten Gutachten- und Begutachtungsstruktur. Die

Rentenversicherung hat im Jahr 2001 verbindliche Hinweise für alle internen und externen Sozialmediziner publiziert, in denen der Ablauf der Begutachtung und der Aufbau eines Gutachtens dargestellt wurden (VDR, 2001). Dies waren die ersten verbindlichen Maßnahmen zur Qualitätssicherung in der Begutachtung, da Qualitätssicherung im Kern systematisch vordefinierte Qualitätsindikatoren erfasst, prüft und dadurch potenzielle Defizite beseitigt (Wirtz, 2018c). In der vorliegenden Arbeit wurde ein trägerübergreifendes Qualitätssicherungsverfahren für den Einsatz in der Routine evaluiert und optimiert. Zielsetzung des Verfahrens war es, eine systematische Bewertung der Qualität von Gutachten durch die Anwendung einheitlicher Qualitätskriterien und somit eine regional und überregional vergleichbare Gutachtenüberprüfung zu ermöglichen.

Am Beispiel der Einzelkriterien und des übergeordneten Kriteriums konnten die gegenseitige Wechselwirkungen zwischen Qualität und Qualitätssicherung aufgezeigt werden. So definieren die im Peer Review geprüften Qualitätskriterien gleichzeitig den zu prüfenden Untersuchungsgegenstand. Bei der Entwicklung des Verfahrens wurde darauf geachtet, die Review-Kategorien entlang der formalen und inhaltlichen Struktur der sozialmedizinischen Gutachten zu konzipieren, sodass nach Einführung der externen Qualitätssicherung keine neuen Anforderungen an die Gutachten gestellt werden müssen. Gleichzeitig werden die im Manual geprüften Kriterien sukzessiv in die Leitlinien der Sozialmedizinischen Begutachtung aufgenommen (z B. Leitlinie Sozialmedizinische Beurteilung bei psychischen und Verhaltensstörungen). Durch die Implementierung in die sozialmedizinischen Leitlinien wird sich zukünftig die einheitliche Vorgehensweise und Dokumentation der Begutachtung nachhaltig verfestigen. Mit den Ergebnissen der qualitativen Inhaltsanalyse konnte gezeigt werden, dass in vielen Fällen die im Manual vorgegebenen Mängeldokumentationen von den Peers genutzt wurden, um Qualitätsdefizite zu kodieren. Dies impliziert, dass das untersuchte Verfahren in der Lage ist, bestimmte Gutachteninhalte und -kriterien zu prüfen und im Rahmen der Qualitätssicherung keine Domänen im Peer Review überprüft werden, die nicht auch in den Gutachten verankert sind.

Im Rahmen eines Benchmarkings erhalten die Träger über das Review-Verfahren zukünftig Rückmeldung über die formale Gestaltung, Verständlichkeit, Transparenz, Vollständigkeit, medizinisch-wissenschaftliche Grundlagen, Wirtschaftlichkeit und Nachvollziehbarkeit ihrer Gutachten im Vergleich zu allen anderen Trägern. Die Ergebnisse des Peer Reviews können damit in das Qualitätsmanagement der Träger eingebunden werden. Dadurch kann mit Einführung der trägerübergreifenden Qualitätssicherung die Qualität sozialmedizinischer Gutachten sukzessiv optimiert werden.

Peer Review-Verfahren und Peer-Schulung. Die Evaluation des Peer Reviews und der Aufbau der Schulung stellte den letzten Schritt vor der deutschlandweiten Implementation des Verfahrens dar. Die erreichte Reliabilität ist je nach Prüffrage sehr heterogen. Wie im entsprechenden Kapitel 4.4.1.1 beschrieben wird die Höhe der Reliabilität jedoch durch zahlreiche Faktoren determiniert. Ein niedriger Konkordanzkoeffizient W bedeutet nicht automatisch, dass keine Übereinstimmung der Peers vorliegt. Es sei an dieser Stelle daher nochmals angemerkt, dass auch eine Übereinstimmung von 100 % mit einer niedrigen Reliabilität assoziiert sein kann, wenn nicht alle Skalen der Mängelkategorien ausgenutzt werden, da eine Reliabilität mit einer Konstanten nicht berechnet werden kann (Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990). Nach der Evaluation des Manuals in Teilstudie 2 wurden die Prüffragen seitens der DRV Bund nochmals grundlegend modifiziert, Prüffragen mit schlechter Reliabilität und Übereinstimmung wurden entfernt (vgl. Anhang 5 und Anhang 9). Eine Re-evaluation in einem ähnlichen Forschungsdesign fand mit der im Jahr 2013 publizierten Version des Manuals jedoch nicht statt. Obgleich sich die Interrater-Reliabilität auf einem ähnlichen Niveau wie nationale und internationale Qualitätssicherungsprogramme bewegte, wurde versucht, die Reliabilität des Verfahrens mittels einer standardisierten Peer-Schulung zu optimieren. Die Ergebnisse zeigen, dass die Schulung seitens der Peers gut angenommen wurde und einen wichtigen Bestandteil der Peer-Tätigkeit darstellt. Die Evaluation der Schulung als auch die Auswertung des Kalibrierungswrkshops legen nahe, dass die Kommunikation zwischen den Peers die Beurteilungsmaßstäbe auf ein nahezu ähnliches Niveau kalibrieren kann. Gestützt wird diese Annahme durch die Teilergebnisse von Hofer und Kollegen (2000), die vor allen bei bilateraler Kommunikation zwischen Peers Effekte für höhere Interrater-Reliabilität nachweisen konnten. Es ist anzunehmen, dass sich der Kalibrierungseffekt mit der Zahl der gemeinsam beurteilten Gutachten längerfristig noch verbessern wird. In der Routine sollte die Peer-Schulung einmal jährlich stattfinden. Ob dieser Zeitraum hinreichend ist, damit Schulungseffekte nicht verloren gehen, kann innerhalb dieser Arbeit nicht abschließend geklärt werden. Um die Schulungseffekte nachhaltig zu verstetigen, sollten Abseits des routinemäßig durchgeführten Reviews Qualitätszirkel eingerichtet werden, in denen Peers eines RV-Trägers oder überregional zusammenkommen und gemeinsam Gutachten anhand der Prüffragen beurteilen. Vor dem Hintergrund, dass die Begutachtung ein komplexes Geschehen darstellt, welches sich in seiner Komplexität innerhalb des Gutachtens wiederfindet, müssen die Peers ein möglichst breites Spektrum an Gutachten bearbeiten, um viele der typischen Gutachtenkonstellationen mindestens einmal beurteilen zu können. Eine einzige jährliche Schulung, in der drei Gutachten in einer Kleingruppe beurteilt werden, erscheint dabei nicht zielführend effektiv. Stattdessen wird eine Vernetzung der aktiven Peers in einem überregionalen Netzwerk mit regelmäßigen

Internetkonferenzen als zielführend erachtet, um die Bewertungsmaßstäbe aller Peers kontinuierlich über das Jahr verteilt zu kalibrieren.

Ausblick. Die vorliegende Forschungsarbeit untersuchte drei konsekutive Maßnahmen zur Verbesserung der Qualität der Sozialmedizinischen Begutachtung, die auf den Handlungsempfehlungen der SOMEKO Kommission für mehr Evidenzbasierung und Qualitätssicherung innerhalb der Begutachtung basieren (vgl. VDR, 2004). Zusammenfassend lässt sich diesbezüglich konstatieren, dass die vorliegenden Ergebnisse die Umsetzung dieser Handlungsempfehlungen darstellen und damit den vorläufigen Abschluss der SOMEKO Kommission bilden. Die vorliegende systematische Literaturrecherche griff die Vorschläge von Strahl und Vogel (2016) sowie Strahl et al. (2016a, 2016b) auf und ergänzte deren systematisches Literaturreview methodisch und inhaltlich. Im Rahmen der Recherche wurde die Forschungsfrage zur aktuellen Evidenz in der Begutachtung beantwortet. Es kristallisierte sich heraus, dass keine evidenzbasierten Entscheidungspfade zur Begutachtung vorliegen und Studien mit höheren Evidenzgraden rar sind. Die Ergebnisse zeigen auch, dass übergeordnete, themenspezifische Recherchen offenbar nicht geeignet sind, um Studien mit höherer Evidenz zu identifizieren. Somit sollten zukünftig weitere, tiefergehende Recherchen durchgeführt werden, welche die einzelnen Begutachtungsschritte separat prüfen. Zudem besteht ein genereller Forschungsbedarf mittels höherwertiger Studiendesigns, den Forschungsgegenstand der „Sozialmedizinischen Begutachtung“ angemessen zu untersuchen.

Abschließend ist festzuhalten, dass durch die Konstruktion und Evaluation des trägerübergreifenden Qualitätssicherungsverfahrens und dessen zugehöriger Schulung eine jahrzehntelange Versorgungslücke im Bereich der Sozialmedizin geschlossen werden konnte. Obgleich die Reliabilität des Verfahrens auch nach Abschluss der Studien optimiert werden sollte, wurde die erreichte Höhe der Reliabilitätskoeffizientenausprägung durch bereits publizierte Studienergebnisse bestätigt. Das gekürzte, aktuelle Manual zum Peer Review-Verfahren unterliegt einer ständigen Weiterentwicklung und sollte daher zukünftig nochmals re-evaluiert werden. Mit der hier vorliegenden Forschungsarbeit wurde erstmals eine vollständig standardisierte, curriculare Peer-Schulung erstellt, die durch das Vorhalten von konkreten Umsetzungsinstruktionen unabhängig vom Schulungspersonal durchgeführt werden kann. Die Ergebnisse leisten insgesamt einen wichtigen Beitrag zur Gleichbehandlung aller Versicherten mit aktiven Anträgen zur Gewährung einer EM-Rente. Alle vorgestellten Maßnahmen führen langfristig zu einer validen Sozialmedizinischen Begutachtung, einer einheitlichen Nutzung von Beurteilungshilfen zu sozialmedizinischen Fragestellungen sowie zu einer Gutachter-unabhängigen Gleichbehandlung aller versicherten Antragsteller.

7 Literaturverzeichnis

- Alex, C. (2005). Sozialmedizin - Was ist das? *Bayrisches Ärzteblatt*, 10, 719.
- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.
- Antes, G. (1998). Evidence-Based Medicine. *Der Internist*, 39 (9), 899–908.
- Antonovsky, A. (1997). *Salutogenese: Zur Entmystifizierung der Gesundheit*. Tübingen: dgvt-Verlag.
- American Psychological Association (APA). (2018). *PsycINFO*. Zugriff am 30.10.2018, verfügbar unter: <http://www.apa.org/pubs/databases/psycinfo/psycinfo-printable-fact-sheet.pdf>
- Aurich-Berheide, P. & Brussig, M. (2017). Assessment of work ability in competing strands of social insurance: The German case. *Journal of Poverty and Social Justice*, 25 (2), 163–176.
- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V (AWMF). (2018). *Leitlinien: Klassifikation S1 Leitlinien*. Zugriff am 24.11.2018, verfügbar unter: <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung/awmf-regelwerk-01-planung-und-organisation/po-stufenklassifikation/%0Aklassifikation-s1.html%0A>
- Baase, C. M. (2007). Auswirkungen chronischer Krankheiten auf Arbeitsproduktivität und Absentismus und daraus resultierende Kosten für die Betriebe. In B. Badura, H. Schellschmidt & C. Vetter (Hrsg.), *Fehlzeiten-Report 2006: Chronische Krankheiten* (S. 45–59). Berlin: Springer.
- Bäcker, G. (2013). Erwerbsminderungsrenten = Armutsrenten. Ein vergessenes soziales Problem?. *WSI-Mitteilungen*, 8, 572–579.
- Bäcker, G., Bispinck, R. & Hofemann, K. (2013). Zur Aktualität „alter“ sozialer Risiken: Erwerbsminderung. In G. Bäcker & R.G. Heinze (Hrsg.), *Soziale Gerontologie in gesellschaftlicher Verantwortung* (S. 51–68). Wiesbaden: Springer.
- Backman, C. L. (2004). Employment and work disability in rheumatoid arthritis. *Current Opinion in Rheumatology*, 16 (2), 148–152.
- Bundesarbeitsgemeinschaft für Rehabilitation (BAR). (2016). *Gemeinsame Empfehlungen - Begutachtung*. Frankfurt/Main: BAR.
- Barth, J., de Boer, W. E. L., Busse, J. W., Hoving, J. L., Kedzia, S., Couban, R. et al. (2017). Inter-rater agreement in evaluation of disability: Systematic review of reproducibility studies. *BMJ*, 356, j14.
- Bassler, M., Nosper, M., Follert, P., Böwering, L. & Polak, U. (2007). Datenquellen für eine kontinuierliche Qualitätsverbesserung in der medizinischen Rehabilitation: Das QS-Reha®-Verfahren der gesetzlichen Krankenkassen und das Dokumentationssystem Eva-Reha® des MDK Rheinland-Pfalz. *Die Rehabilitation*, 46 (3), 155–163.
- Baumgarten, E. & Klosterhuis, H. (2007). Aktuelles aus der Reha-Qualitätssicherung: Peer Review-Verfahren ausgewertet – bessere Reha-Qualität, aber deutliche Unterschiede zwischen Reha-Einrichtungen. *RVaktuell*, 54 (5), 152–154.
- Baumgartner, P. (2014). *Taxonomie von Unterrichtsmethoden: Ein Plädoyer für didaktische Vielfalt* (2. Aufl.). Münster: Waxmann Verlag GmbH.
- Birkenbihl, M. (2014). *Train the Trainer: Arbeitsbuch für Ausbilder und Dozenten* (21. Aufl.). München: mi-Wirtschaftsbuch.

- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.
- Blüher, S. & Kuhlmeier, A. (2016). Demographischer Wandel, Altern und Gesundheit. In M. Richter & K. Hurrelmann (Hrsg.), *Soziologie von Gesundheit und Krankheit* (S. 313–324). Wiesbaden: Springer.
- Blumenstock, G. (2011). Zur Qualität von Qualitätsindikatoren. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 54 (2), 154–159.
- Blümle, A., Lagrèze, W. A. & Motschall, E. (2018). Systematische Literaturrecherche in PubMed. *Der Ophthalmologe*, 115 (3), 243–260.
- Bomsdorf, E. (2015). Eine visionäre vorausschauende Kurzanalyse Lebenserwartung über 90 Jahre heute, Rente ab 70 morgen?. *ifo Schnelldienst*, 68 (23), 15–23.
- Bortz, J. & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (3. Aufl.). Berlin: Springer.
- Bortz, J., Lienert, G. A. & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3. Aufl.). Berlin: Springer.
- Brage, S., Donceel, P. & Falez, F. (2008). Development of ICF Core Set for disability evaluation in social security. *Disability and Rehabilitation*, 30 (18), 1392–1396.
- Bramesfeld, A. & Schwartz, F. (2007). Volkskrankheit Depression: Bestandsaufnahme und Perspektiven. *Psychiatrische Praxis*, 34 (3), 247–251.
- Breyer, F. (2015). Demographischer Wandel und Gesundheitsausgaben: Theorie, Empirie und Politikimplikationen. *Perspektiven der Wirtschaftspolitik*, 16 (3), 215–230.
- Broder, M. S., Oken, C., Parker, M., Giammona, M., Newman, J., Harrington Charlene et al. (2002). *Outpatient care: A conceptual framework and a form for structured implicit review*. Santa Monica: RAND Health.
- Bülau, N., Kessemeier, F., Petermann, F., Bassler, M. & Kobelt, A. (2016). Evaluation von Kontextfaktoren in der psychosomatischen Rehabilitation. *Die Rehabilitation*, 55 (06), 381–387.
- Bürger, W. & Deck, R. (2009). SIBAR – ein kurzes Screening-Instrument zur Messung des Bedarfs an berufsbezogenen Behandlungsangeboten in der medizinischen Rehabilitation. *Die Rehabilitation*, 48 (04), 211–221.
- Burggraf, H. (2009). Eigenheiten der Begutachtung in der Privaten und der Gesetzlichen Unfallversicherung. *Klinische Monatsblätter für Augenheilkunde*, 226 (08), 605–615.
- Burisch, J. & Munkholm, P. (2015). The epidemiology of inflammatory bowel disease. *Scandinavian Journal of Gastroenterology*, 50 (8), 942–951.
- Buslei, H. (2017). Erhöhung der Regelaltersgrenze über 67 Jahre hinaus trägt spürbar zur Konsolidierung der Rentenfinanzen und Sicherung der Alterseinkommen bei. *DIW Wochenbericht*, 48, 1090–1097.
- Canela, C., Schleifer, R., Dube, A., Hengartner, M. P., Ebner, G., Seifritz, E. et al. (2016). Funktionsbeschreibung in der Begutachtung der Arbeitsfähigkeit – Was ist „evidence-based“? *Psychiatrische Praxis*, 43 (2), 74–81.
- Canela, C., Schleifer, R., Jeger, J., Ebner, G., Seifritz, E. & Liebrecht, M. (2015). Die invalidenversicherungsrechtliche Begutachtung in der Schweiz vor dem Hintergrund der letzten Gesetzesrevision und neueren Rechtsprechung. *Forensische Psychiatrie, Psychologie, Kriminologie*, 9 (2), 106–116.

- Cerniauskaite, M., Quintas, R., Boldt, C., Raggi, A., Cieza, A., Bickenbach, J. E. et al. (2011). Systematic literature review on ICF from 2001 to 2009: Its use, implementation and operationalisation. *Disability and Rehabilitation* 33 (4), 281-309.
- Chatterji, S., Byles, J., Cutler, D., Seeman, T. & Verdes, E. (2015). Health, functioning, and disability in older adults - present status and future implications. *The Lancet*, 385 (9967), 563–575.
- Cibis, W., Paulus, E.-M., Mai, H. & Gehrke, J. (2011). Die sozialmedizinische Begutachtung. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 77–114). Berlin: Springer.
- Cicchetti, D. V. & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43 (6), 551–558.
- Cieza, A., Ewert, T., Üstün, T. B., Chatterji, S., Kostanjsek, N. & Stucki, G. (2004). Development of ICF Core Sets for patients with chronic conditions. *Journal of Rehabilitation Medicine*, 36 (44), 9–11.
- Clark, W. L., Haldeman, S., Johnson, P., Morris, J., Schulenberg, C., Trauner, D. et al. (1988). Back impairment and disability determination. Another attempt at objective, reliable rating. *Spine*, 13 (3), 332–341.
- Clemens, W. (2014). Zu früh oder wieder später in die „Späte Freiheit“? – Ältere Arbeitnehmer im gesellschaftlichen und demografischen Wandel. In A. Amann & F. Kolland (Hrsg.), *Das erzwungene Paradies des Alters?* (S. 109–127). Wiesbaden: Springer.
- Cochrane Deutschland. (2018). *Systematische Übersichtsarbeiten der Cochrane Library*. Zugriff am 30.10.2018, verfügbar unter: <http://www.cochrane.de/de/systematische-uebersichtsarbeiten>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). New Jersey: Lawrence Erlbaum Associates.
- Croft, P. R., Macfarlane, G. J., Papageorgiou, A. C., Thomas, E. & Silman, A. J. (1998). Outcome of low back pain in general practice: A prospective study. *BMJ*, 316 (7141), 1356–1359.
- Dannenberg, A., Hofmann, J., Kaldybajewa, K. & Kruse, E. (2010). Rentenzugang 2009: Weiterer Anstieg der Zugänge in Erwerbsminderungsrenten wegen psychischer Erkrankungen. *RVaktuell*, 9, 283–293.
- Deck, R. & Pohontsch, N. (2011). Älter, kränker und trotzdem länger arbeiten – neue Anforderungen an die medizinische Rehabilitation? *Public Health Forum*, 19 (4), 17.e1-17.e4.
- Dell-Kuster, S., Lauper, S., Koehler, J., Zwimpfer, J., Altermatt, B., Zwimpfer, T. et al. (2014). Assessing work ability: A cross-sectional study of interrater agreement between disability claimants, treating physicians, and medical experts. *Scandinavian Journal of Work Environment & Health*, 40 (5), 493–501.
- Derra, C. (2016). Sozialmedizinische Begutachtung. In J. Bengel & O. Mittag (Hrsg.), *Psychologie in der medizinischen Rehabilitation: Ein Lehr- und Praxisbuch* (S. 71–81). Berlin: Springer.
- DESTATIS (Statistisches Bundesamt). (2015). *Bevölkerung Deutschlands bis 2060 - 13. koordinierte Bevölkerungsvorausberechnung*. Wiesbaden: Statistisches Bundesamt. Zugriff am 24.7.2018, verfügbar unter: https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060Presse5124204159004.pdf?__blob=publicationFile

- Deutsche Vereinigung für Rehabilitation (DVfR). (2015). Nutzung der ICF im deutschen Rehabilitationssystem: Dringende Handlungsfelder für Politik und Rehabilitationsträger. *Die Rehabilitation*, 54 (01), 60–61.
- Dewa, C. S., Hees, H., Trojanowski, L. & Schene, A. H. (2015). Clinician experiences assessing work disability related to mental disorders. *PloS one*, 10 (3), e0119009.
- Dickmann, J. & Broocks, A. (2007). Das psychiatrische Gutachten im Rentenverfahren - wie reliabel? *Fortschritte der Neurologie/Psychiatrie*, 75 (7), 397–401.
- Diehl, R., Knisel, W. K. & Haupt, E. (2011). Metabolische und endokrine Krankheiten. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 269–287). Berlin: Springer.
- Deutsches Institut für Medizinische Dokumentation und Information/ Weltgesundheitsorganisation (DIMDI/WHO). (2005). *Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) der Weltgesundheitsorganisation (WHO)-deutschsprachige Fassung*. Genf: World Health Organization.
- Dodgshun, C. & Malios, J. (2011). Workers' compensation forms: A guide for GPs. *Australien Family Physician*, 40 (9), 730–733.
- Donabedian, A. (2005). Evaluating the quality of medical care [unveränderter Nachdruck von 1966]. *The Milbank Quarterly*, 83 (4), 691–729.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
- Deutsche Rentenversicherung (DRV). (2009a). *Leitlinien für die sozialmedizinische Begutachtung: Leistungsfähigkeit bei Bandscheiben- und bandscheibenassoziierten Erkrankungen*. Berlin: Deutsche Rentenversicherung Bund. Zugriff am 1.10.2018, verfügbar unter: https://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/3_Infos_fuer_Experten/01_sozialmedizin_forschung/downloads/sozmed/begutachtung/leitlinie_leistungsaehigkeit_bandscheibe_pdf.pdf?__blob=publicationFile&v=4
- Deutsche Rentenversicherung (DRV). (2009b). *DRV Schriften, Band 81: Sozialmedizinisches Glossar der Deutschen Rentenversicherung*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2009c). *Qualitätssicherung der sozialmedizinischen Begutachtung – Bericht zur Umsetzung des „Qualitätssicherungsverfahrens der sozialmedizinischen Begutachtung“ - aktueller Sachstand*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2011). *Leitlinien für die sozialmedizinische Begutachtung: Rehabilitation bei onkologischen Krankheiten*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2013). *Qualitätssicherung der sozialmedizinischen Begutachtung: Manual zum Peer Review-Verfahren*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2014). *Positionspapier der Deutschen Rentenversicherung zur Bedeutung psychischer Erkrankungen in der Rehabilitation und bei Erwerbsminderung*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2015). *Leitlinien für die sozialmedizinische Begutachtung: Sozialmedizinische Beurteilung von Menschen mit koronarer Herzkrankheit*. Berlin: Deutsche Rentenversicherung Bund.

- Deutsche Rentenversicherung (DRV). (2017a). *Statistik der Deutschen Rentenversicherung: Rentenversicherung in Zahlen 2017*. Zugriff am 24.7.2018, verfügbar unter: https://www.deutsche-rentenversicherung.de/cae/servlet/contentblob/238692/publicationFile/61815/01_rv_in_zahlen_2013.pdf
- Deutsche Rentenversicherung (DRV). (2017b). *Statistik der Deutschen Rentenversicherung: Erwerbsminderungsrenten im Zeitablauf 2017*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2017c). *Leitlinien für die sozialmedizinische Begutachtung: Beurteilung der Rehabilitationsbedürftigkeit von Menschen mit muskuloskeletalen Erkrankungen*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2018a). *Unsere Sozialversicherung - Wissenswertes speziell für junge Leute*. Berlin. Zugriff am 25.9.2018, verfügbar unter: https://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/5_Services/03_broschueren_und_mehr/01_broschueren/01_national/unsere_sozialversicherung.pdf?__blob=publicationFile&v=36
- Deutsche Rentenversicherung (DRV). (2018b). *Reha-Bericht 2018*. Zugriff am 24.9.2018, verfügbar unter: https://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/6_Wir_ueber_uns/03_fakten_und_zahlen/03_statistiken/02_statistikpublikationen/02_rehabericht_2018.html
- Deutsche Rentenversicherung (DRV). (2018c). *Erwerbsminderungsrente: Das Netz für alle Fälle*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung (DRV). (2018d). *Deutsche Rentenversicherung - Leitlinien*. Zugriff am 30.9.2018, verfügbar unter: https://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/3_Infos_fuer_Experten/01_sozialmedizin_forschung/01_sozialmedizin/03_begutachtung/leitlinien_index.html
- Edwards, M. T. (2011). The objective impact of clinical peer review on hospital quality and safety. *American Journal of Medical Quality*, 26 (2), 110–119.
- Egner, U., Gerwin, H. & Buschmann-Steinhage, R. (2006). Stand der Qualitätssicherung in der Rehabilitation der gesetzlichen Rentenversicherung. *Die Rehabilitation*, 45 (4), 221–231.
- Egner, U., Gerwin, H. & Schliehe, F. (2002). Das bundesweite Reha Qualitätssicherungsprogramm der gesetzlichen Rentenversicherung - Erfahrungen aus einem mehrjährigen Umsetzungsprozess. *Zeitschrift für Ärztliche Fortbildung und Qualitätssicherung*, 96, 4–9.
- Eikermann, B. & Zacharias, B. (2003). Sozialpsychiatrische Aspekte psychischer Krankheiten. In H.-J. Möller, G. Laux & H.-P. Kapfhammer (Hrsg.), *Psychiatrie und Psychotherapie* (S. 208–224). Berlin: Springer.
- Elder, A. G., Symington, I. S. & Symington, E. H. (1994). Do occupational physicians agree about ill-health retiral? A study of simulated retirement assessments. *Occupational Medicine*, 44 (5), 231–235.
- Ewert, T., Cieza, A. & Stucki, G. (2002). Die ICF in der Rehabilitation. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin*, 12 (3), 157–162.
- Ewert, T., Freudenstein, R. & Stucki, G. (2008). Die ICF in der Sozialmedizin. *Das Gesundheitswesen*, 70 (10), 600–616.
- Fach, E. M., Bethge, M., Spanier, K. & Richter, M. (2017). Soziale Ungleichheiten im Zugang zu Rehabilitations- und Rentenleistungen. Ergebnisse aus dem Dritten Sozialmedizinischen Panel. *Das Gesundheitswesen*, 79 (08/09), 565–804.

- Faller, H., Reusch, A. & Meng, K. (2011). DGRW-Update: Patientenschulung. *Die Rehabilitation*, 50 (05), 284–291.
- Farin, E. (2008). Patientenorientierung und ICF-Bezug als Herausforderungen für die Ergebnismessung in der Rehabilitation. *Die Rehabilitation*, 47 (2), 67–76.
- Farin, E., Carl, C., Lichtenberg, S., Jäckel, W. H., Maier-Riehle, B. & Rütten-Köppel, E. (2003). Die Bewertung des Rehabilitationsprozesses mittels des Peer-Review-Verfahrens: Methodische Prüfung und Ergebnisse der Erhebungsrunde 2000/2001 in den somatischen Indikationsbereichen. *Die Rehabilitation*, 42 (6), 323–334.
- Farin, E., Engel, E.-M., Dimou, A. & Jäckel, W. H. (2004). Die Erwartungen von Rehabilitations- und Vorsorgeeinrichtungen an externe Qualitätssicherungsmaßnahmen. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin*, 14 (2), 73–81.
- Farin, E. & Jäckel, W. H. (2011). Qualitätssicherung und Qualitätsmanagement in der medizinischen Rehabilitation. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 54 (2), 176–184.
- Farin, E., Jäckel, W. H., Schalaster, V. & AQMS. (2009). Das Qualitätssicherungsverfahren der GKV in der medizinischen Rehabilitation: Ergebnisse und Weiterentwicklung. *Das Gesundheitswesen*, 71 (03), 163–174.
- Feicke, J. & Spörhase, U. (2012). Impulse aus der Didaktik zur Verbesserung von Patientenschulungen. *Die Rehabilitation*, 51 (05), 300–307.
- Feinstein, A. R. & Cicchetti, D. V. (1990). High agreement but low Kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43 (6), 543–549.
- Fischer, S., Rath, T. & Neurath, M. F. (2018). Chronisch-entzündliche Darmerkrankungen. *Der Internist*, 59 (7), 681–693.
- Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76 (5), 378–282.
- Frerichs, F. (2015). Demografischer Wandel in der Erwerbsarbeit – Risiken und Potentiale alternder Belegschaften. *Journal for Labour Market Research*, 48 (3), 203–216.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303, 130–135.
- Fries, J. F. (2003). Measuring and monitoring success in compressing morbidity. *Annals of Internal Medicine*, 139, 455.
- Fries, J. F. (2005). The compression of morbidity. *The Milbank Quarterly*, 83 (4), 801–23.
- Fritze, E. & Fritze, J. (2008). Allgemeine Grundlagen. In E. Fritze & F. Mehrhoff (Hrsg.), *Die ärztliche Begutachtung: Rechtsfragen, Funktionsprüfungen, Beurteilungen* (7. vollständig überarbeitete Aufl., S. 1–11). Darmstadt: Steinkopff Verlag.
- Froggatt, K. A. (2001). The analysis of qualitative data: Processes and pitfalls. *Palliative Medicine*, 15 (5), 433–438.
- Gesundheit, Statistik, Gesundheitsberichterstattung des Bundes (GBE-Bund). (2018a). *Bei den Ärztekammern registrierte Ärztinnen und Ärzte mit Zusatz-Weiterbildungen (Zusatzbezeichnung). Gliederungsmerkmale: Jahre, Region, Geschlecht, Zusatz-Weiterbildung, Tätigkeitsbereich: Sozialmedizin*. Zugriff am 10.10.2018, verfügbar unter: http://www.gbe-bund.de/oowa921-install/servlet/oowa/aw92/dboowasys921.xwdevkit/xwd_init?gbe.isgbetol/xs_start_neu/&p_aid=3&p_aid=34485605&nummer=612&p_sprache=D&p_indsp=999999999&p_aid=58984385#SOURCES

- Gesundheit, Statistik, Gesundheitsberichterstattung des Bundes (GBE-Bund). (2018b). *Rentenzugänge wegen verminderter Erwerbsfähigkeit in der Gesetzlichen Rentenversicherung im Laufe des Berichtsjahres (Anzahl und je 100.000 aktiv Versicherte). Gliederungsmerkmale: Jahre, Region, Zugangsalter, Geschlecht, 1. Diagnose (ICD-10)*. Zugriff am 31.10.2018, verfügbar unter: http://www.gbe-bund.de/oowa921-install/servlet/oowa/aw92/WS0100/_XWD_FORMPROC
- Gechter, D., Nothacker, M., Khan, C., Rütters, D., Mavergames, C., Motschall, E. et al. (2013). *Manual: Systematische Literaturrecherche für die Erstellung von Leitlinien (Version 1 vom 10.05.2013)*. Zugriff am 02.11.2018, verfügbar unter: <https://doi.org/10.6094/UNIFR/2013/220/>
- Gehrke, J. (2008). Das Qualitätssicherungskonzept der sozialmedizinischen Begutachtung der Deutschen Rentenversicherung: Einführung und Überblick. *Das Gesundheitswesen*, 70 (07), A90.
- Gehrke, J. (2010). Sozialmedizinische Begutachtung der Deutschen Rentenversicherung: Qualitätssicherung von Gutachten bei Anträgen auf Erwerbsminderung. *Das Gesundheitswesen*, 72 (08/09), V203.
- Geiger, B. B., Garthwaite, K., Warren, J. & Bamba, C. (2017). Assessing work disability for social security benefits: International models for the direct assessment of work capacity. *Disability and Rehabilitation*, 1–9.
- Gellert, F. J., Kesselmann, M. & Wilke, C. B. (2018). Arbeitswelt im Wandel. *Prävention und Gesundheitsförderung*, 13 (1), 12–17.
- Gisev, N., Bell, J. S. & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9 (3), 330–338.
- Glattacker, M. & Jäckel, W. (2007). Evaluation der Qualitätssicherung - aktuelle Datenlage und Konsequenzen für die Forschung. *Das Gesundheitswesen*, 69 (5), 277–283.
- Göbel, D. (2015). Zehn Jahre nach der Organisationsreform – Rückblick und Umsetzungsergebnisse. *RVaktuell*, 4, 82–87.
- Goldman, H. H., Gattozzi, A. A. & Taube, C. A. (1981). Defining and counting the chronically mentally ill. *Psychiatric Services*, 32 (1), 21–27.
- Goldman, R. L. (1994). The reliability of peer assessments: A meta-analysis. *Evaluation & the Health Professions*, 17 (1), 3–21.
- Götz, M., Roth, S. & Chojetzki, U. (2011). Rechtliche Grundlagen für Leistungen der gesetzlichen Rentenversicherung. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 3–25). Berlin: Springer
- Goulet, F., Jacques, A., Gagnon, R., Racette, P. & Sieber, W. (2007). Assessment of family physicians' performance using patient charts: Interrater reliability and concordance with chart-stimulated recall interview. *Evaluation & the Health Professions*, 30 (4), 376–392.
- Gräßler, F. (2012). Die Institution der gesetzlichen Rentenversicherung zwischen Reform und Krise. In F. Gräßler (Hrsg.), *Die Restabilisierung der gesetzlichen Rentenversicherung* (S. 11–21). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Groner, A. (2012). Einführung in die Sozialmedizin, Grundlagen der Gesundheitsversorgung. In R. Diehl, E. Gebauer & A. Groner (Hrsg.), *Kursbuch Sozialmedizin: Lehrbuch zum Curriculum der Bundesärztekammer* (S. 3–140). Köln: Deutscher Ärzte-Verlag.

- Guba, B. (2008). Systematische Literatursuche. *Wiener Medizinische Wochenschrift*, 158 (1–2), 62–69.
- Guyatt, G. H., Oxman, A. D., Kunz, R., Atkins, D., Brozek, J., Vist, G. et al. (2011). GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology*, 64 (4), 395–400.
- Hagen, C., Himmelreicher, R. K., Kempfner, D. & Lampert, T. (2011). Soziale Ungleichheit und Risiken der Erwerbsminderung. *WSI-Mitteilungen*, 7, 336–344.
- Hammer, M. (2002). Begutachtung von Patienten mit rheumatoider Arthritis. *Zeitschrift für Rheumatologie*, 61 (6), 626–634.
- Harding, T. P. (2004). Psychiatric disability and clinical decision making: The impact of judgment error and bias. *Clinical Psychology Review*, 24 (6), 707–729.
- Harris, C. D. & Bratzler, D. W. (2013). Evaluating quality of care: The role of peer review. *The Journal of the Oklahoma State Medical Association*, 106 (7), 279, 281–284.
- Harris, R. E. (2010). *Epidemiology of chronic disease: Global perspectives*. Burlington: Michael Brown.
- Haynes, B. (1999). Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. *BMJ*, 319 (7211), 652–653.
- Hayward, R. A., McMahon, L. F. & Bernard, A. M. (1993). Evaluating the care of general medicine inpatients: How good is implicit review? *Annals of Internal Medicine*, 118 (7), 550.
- Heidemann, S., Kuhnert, R., Born, S. & Scheidt-Nave, C. (2017). 12-Monats-Prävalenz des bekannten Diabetes mellitus in Deutschland. *Journal of Health Monitoring*, 2 (1), 48–56.
- Heigl, A. (2004). Älter gleich kränker? In R. Scholz & J. Flöthmann (Hrsg.), *Materialien zur Bevölkerungswissenschaft: Jahrestagung 2002 der Deutschen Gesellschaft für Demographie in Rostock* (S. 41–49). Wiesbaden: Bundesinstitut für Bevölkerungsforschung beim Statistischen Bundesamt.
- Herget, S., Roski, C. & Grande, G. (2017). Inanspruchnahme psychosomatischer Rehabilitationsmaßnahmen und Beratungen wegen Erwerbsminderung aufgrund psychischer Erkrankungen – Eine qualitative Analyse. In Deutsche Rentenversicherung Bund (Hrsg.), *DRV-Schriften: Band 111: Tagungsband zum 26. Rehabilitationswissenschaftlichen Kolloquium: Prävention und Rehabilitation in Zeiten der Globalisierung vom 20. bis 22. März 2017 in Frankfurt am Main* (S. 419–421). Berlin: Deutsche Rentenversicherung Bund.
- Hesse, B. & Gebauer, E. (2011). Sozialmedizinische Begutachtung im Rentenverfahren: Stellenwert, Forschungsbedarf und Chancen. *Die Rehabilitation*, 50 (01), 17–24.
- Hesse, B. & Walden, J. (2014). Rehabilitation und Erwerbsminderungsrente bei psychischen Störungen. *DNP - Der Neurologe und Psychiater*, 15 (1), 57–63.
- Heubrock, D. & Petermann, F. (1998). Neuropsychological assessment of suspected malingering: Research results, evaluation techniques, and further directions of research and application. *European Journal of Psychological Assessment*, 14 (3), 211–225.
- Hien, P., Böhm, B., Claudi-Böhm, S., Krämer, C. & Kohlhas, K. (2013). Definition, Klassifikation, Inzidenz und Prävalenz des Diabetes mellitus. In P. Hien, B. Böhm, B. Claudi-Böhm, C. Krämer & K. Kohlhas (Hrsg.), *Diabetes-Handbuch* (S. 17–19). Berlin: Springer.

- Hofer, T. P., Asch, S. M., Hayward, R. A., Rubenstein, L. V., Hogan, M. M., Adams, J. et al. (2004). Profiling quality of care: Is there a role for peer review? *BMC Health Services Research*, 4 (1), 9.
- Hofer, T. P., Bernstein, S. J., DeMonner, S. & Hayward, R. A. (2000). Discussion between reviewers does not improve reliability of peer review of hospital quality. *Medical Care*, 38 (2), 152–161.
- Hoffmann, W., Vahlensieck, W. & Zermann, D.-H. (2016). Sozialmedizinische Begutachtung bei Patienten mit Prostatakarzinom. *Der Urologe*, 55 (11), 1481–1486.
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M. et al. (2018). *Mixed Methods Appraisal Tool (MMAT) VERSION 2018: Registration of Copyright (#1148552)*. Zugriff am 14.11.2018, verfügbar unter: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/attach/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf
- Horn, S. & Schuchardt, D. R. (2015). *Deutsche Rentenversicherung - Basis der Altersvorsorge*. Wiesbaden: Springer.
- Hoy, D., Brooks, P., Blyth, F. & Buchbinder, R. (2010). The Epidemiology of low back pain. *Best Practice & Research Clinical Rheumatology*, 24 (6), 769–781.
- Ingravallo, F., Vignatelli, L., Brini, M., Brugaletta, C., Franceschini, C., Lugaresi, F. et al. (2008). Medico-legal assessment of disability in narcolepsy: An interobserver reliability study. *Journal of Sleep Research*, 17 (1), 111–119.
- Jäckel, W. H. & Farin, E. (2004). Qualitätssicherung in der Rehabilitation: Wo stehen wir heute? *Die Rehabilitation*, 43 (5), 271–283.
- Jacobi, F., Höfler, M., Strehle, J., Mack, S., Gerschler, A., Scholl, L. et al. (2014). Psychische Störungen in der Allgemeinbevölkerung. *Der Nervenarzt*, 85 (1), 77–87.
- Jacobi, F., Höfler, M., Strehle, J., Mack, S., Gerschler, A., Scholl, L. et al. (2016). Erratum zu: Psychische Störungen in der Allgemeinbevölkerung. Studie zur Gesundheit Erwachsener in Deutschland und ihr Zusatzmodul „Psychische Gesundheit“ (DEGS1-MH). *Der Nervenarzt*, 87 (1), 88–90.
- Jacobs, E. & Rathmann, W. (2018). Epidemiologie des Diabetes in Deutschland. In Deutsche Diabetes Gesellschaft, diabetesDE & Deutsche Diabetes Hilfe (Hrsg.), *Deutscher Gesundheitsbericht: Diabetes 2018 - Die Bestandsaufnahme* (S. 9–22). Mainz: Kirchheim + Co GmbH.
- Jansen, R. & Müller, R. (2000). Arbeitsbelastungen und Gesundheit älterer Arbeitnehmer im Dienstleistungsbereich. *Zeitschrift für Gerontologie und Geriatrie*, 33 (4), 256–261.
- Jones, R. W. (2007). Learning and teaching in small groups: Characteristics, benefits, problems and approaches. *Anaesthesia and Intensive Care*, 35 (4), 587–592.
- Jüni, P., Altman, D. G. & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*, 323 (7303), 42–46.
- Kaldybajewa, K. & Kruse, E. (2012). Erwerbsminderungsrenten im Spiegel der Statistik der gesetzlichen Rentenversicherung: Unterschiede und Gemeinsamkeiten zwischen Männern und Frauen. *RVaktuell*, 8, 206–216.
- Kalousova, L. & Burgard, S. A. (2014). Unemployment, measured and perceived decline of economic resources: Contrasting three measures of recessionary hardships and their implications for adopting negative health behaviors. *Social Science & Medicine*, 106, 28–34.

- Kameoka, J., Okubo, T., Koguma, E., Takahashi, F., Ishii, S. & Kanatsuka, H. (2014). Development of a peer review system using patient records for outcome evaluation of medical education: Reliability analysis. *The Tohoku Journal of Experimental Medicine*, 233 (3), 189–195.
- Kawski, S. & Koch, U. (2004). Qualitätssicherung in der medizinischen Rehabilitation in Deutschland. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 47 (2), 111–117.
- Kellner, H. (2002). Begutachtung von Patienten mit seronegativen Spondylarthropathien. *Zeitschrift für Rheumatologie*, 61 (6), 643–651.
- Kemptoner, D. (2014). *Erwerbsminderung als Armutsrisiko. DIW Roundup: Politik im Fokus* (Band 8). Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R. & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62 (6), 593.
- Khan, C. & Ollenschläger, G. (2014). Wirksamkeit von Qualitätsprogrammen in der stationären Versorgung in Deutschland – eine Literaturanalyse. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 108 (10), 576–586.
- Kim, S. Y. (2013). Efficacy versus effectiveness. *Korean Journal of Family Medicine*, 34 (4), 227.
- Kirschneck, M., Legner, R., Armbrust, W., Nowak, D. & Cieza, A. (2015). Können ICF Core Sets für die Erstellung von sozialmedizinischen Gutachten bei Anträgen auf Erwerbsminderungsrente hilfreich sein? – Ein erster Vorschlag. *Die Rehabilitation*, 54 (02), 92–101.
- Kirschneck, M., Winkelmann, A., Kirchberger, I., Gläsel, A., Ewert, T., Stucki, G. et al. (2008). Anwendung der ICF Core Sets in der Begutachtung von Patienten mit lumbalen Rückenschmerzen und generalisiertem Schmerzsyndrom. *Das Gesundheitswesen*, 70 (11), 674–678.
- Kleibel, V. & Mayer, H. (2011). *Literaturrecherche für Gesundheitsberufe* (2., überar. Aufl.). Wien: Facultas Verlag.
- Klenk, T. (2005). Die Organisationsreform in der gesetzlichen Rentenversicherung: Eine Betrachtung aus politikwissenschaftlicher Perspektive. *Zeitschrift für Sozialreform*, 51 (1), 94–127.
- Kliche, T., Koch, U., Lehmann, H. & Töppich, J. (2006). Evidenzbasierte Prävention und Gesundheitsförderung. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 49 (2), 141–150.
- Klosterhuis, H., Baumgarten, E., Beckmann, U., Erbstößer, S., Lindow, B., Naumann, B. et al. (2010). Ein aktueller Überblick zur Reha-Qualitätssicherung der Rentenversicherung. *Die Rehabilitation*, 49 (06), 356–367.
- Kolenda, K. D. (1998). Sozialmedizinische Beurteilung und Beratung von Patienten mit koronarer Herzkrankheit. *Deutsche Medizinische Wochenschrift*, 123 (23), 741–747.
- Kolip, P. (2006). Evaluation, Evidenzbasierung und Qualitätsentwicklung. *Prävention und Gesundheitsförderung*, 1 (4), 234–239.
- König, V. (2011). Onkologische Erkrankungen. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 247–258). Berlin: Springer.
- Kopp, I. B. (2010). Perspektiven der Leitlinienentwicklung und -implementation aus der Sicht der AWMF. *Zeitschrift für Rheumatologie*, 69, 298–304.

- Körner, M. (2005). ICF und sozialmedizinische Beurteilung der Leistungsfähigkeit im Erwerbsleben: Alles klar? - Ein Diskussionsbeitrag. *Die Rehabilitation*, 44 (04), 229–236.
- Korsukéwitz, C. & Irle, H. (2010). Medizinische Rehabilitation. *Der Internist*, 51 (10), 1219–1230.
- Korsukéwitz, C. & Rehfeld, U. (2008). Rehabilitation und Erwerbsminderungsrenten - Aktueller Stand und Entwicklung. *RVaktuell*, 9, 274–284.
- Krahwinkel, W., Rink, O., Liebetrau, M., Günther, M., Schuler, E. & Kuhlen, R. (2011). 10 Jahre Peer Review – Verbesserung der medizinischen Behandlung durch Qualitätsindikatoren aus Routinedaten. *Deutsche Medizinische Wochenschrift*, 136 (41), 2083–2088.
- Krämer, L. & Bengel, J. (2016). Chronische körperliche Krankheit und Krankheitsbewältigung. In J. Bengel & O. Mittag (Hrsg.), *Psychologie in der medizinischen Rehabilitation: Ein Lehr- und Praxishandbuch* (S. 25–36). Berlin: Springer.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41 (4), 212–218.
- Krause, B. J., Khan, C., Antoch, G., Kotzerke, J. & Hosten, N. (2015). Wie schreibe ich eine gute Übersichtsarbeit? *Nuklearmedizin*, 54 (02), 69–74.
- Kroll, L. E. & Ziese, T. (2009). Kompression oder Expansion der Morbidität? In K. Böhm, Statistisches Bundesamt, C. Tesch-Römer, Deutsches Zentrum für Altersfragen, Z. Ziese & Robert Koch-Institut (Hrsg.), *Beiträge zur Gesundheitsberichterstattung des Bundes: Gesundheit und Krankheit im Alter* (S. 105–112). Berlin: Robert Koch-Institut.
- Lampert, T., Richter, M., Schneider, S., Spallek, J. & Dragano, N. (2016). Soziale Ungleichheit und Gesundheit. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 59 (2), 153–165.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159.
- Legner, R. & Cibis, W. (2007). Qualitätssicherung in der sozialmedizinischen Begutachtung. *Die Rehabilitation*, 46 (1), 57–61.
- Leigemann, M., Lang, B., Kunz, R. & Antes, G. (2005). Leitlinien - Was haben die Ärzte und Patienten davon? *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 48 (2), 215–220.
- Lilford, R., Edwards, A., Girling, A., Hofer, T., Di Tanna, G. L., Petty, J. et al. (2007). Inter-rater reliability of case-note audit: A systematic review. *Journal of Health Services Research & Policy*, 12 (3), 173–180.
- Linden, M. (2007). Das psychiatrische Gutachten im Rentenverfahren - Wie reliabel? *Fortschritte der Neurologie Psychiatrie*, 75 (7), 379–381.
- Linden, M. (2017). Definition and assessment of disability in mental disorders under the perspective of the International Classification of Functioning, Disability and Health (ICF). *Behavioral Sciences & the Law*, 35 (2), 124–134.
- Linden, M. & Baron, S. (2005). Das „Mini-ICF-Rating für psychische Störungen (Mini-ICF-P)“. Ein Kurzinstrument zur Beurteilung von Fähigkeitsstörungen bei psychischen Erkrankungen. *Die Rehabilitation*, 44 (3), 144–151.
- Loddo, I. & Romano, C. (2015). Inflammatory bowel disease: Genetics, epigenetics, and pathogenesis. *Frontiers in Immunology*, 6, 551.

- Löffler, S., Gerlich, C., Meng, K. & Vogel, H. (2010). *Manualisierung eines Prüffragenkatalogs zur Qualitätssicherung sozialmedizinischer Gutachten (QSGUT): unveröffentlichter Abschlussbericht*. Würzburg: Universität Würzburg, Abteilung für Medizinische Psychologie und Psychotherapie, Medizinische Soziologie und Rehabilitationswissenschaften.
- Maaz, A., Winter, M. H.-J. & Kuhlmei, A. (2007). Der Wandel des Krankheitspanoramas und die Bedeutung chronischer Erkrankungen. In B. Badura, H. Schellschmidt & C. Vetter (Hrsg.), *Fehlzeiten-Report 2006: Chronische Krankheiten* (S. 5–23). Berlin: Springer.
- Maher, C., Underwood, M. & Buchbinder, R. (2017). Non-specific low back pain. *The Lancet*, 389 (10070), 736–747.
- Martin, S. & Zollmann, P. (2013). Keine Reha vor Rente? Ergebnisse des Projekts „Sozioökonomische Situation von Personen mit Erwerbsminderung“. In Deutsche Rentenversicherung Bund (Hrsg.), *DRV-Schriften Band 101: Tagungsband zum 22. Rehabilitationswissenschaftliches Kolloquium: Teilhabe 2.0 - Reha neu denken? vom 4. bis 6. März 2013 in Mainz* (S. 109–111). Berlin: Deutsche Rentenversicherung Bund.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz Verlag.
- Mayring, P. & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 543–556). Wiesbaden: Springer.
- Meierjürgen, R., Becker, S. & Warnke, A. (2016). Die Entwicklung der Präventionsgesetzgebung in Deutschland. *Prävention und Gesundheitsförderung*, 11 (4), 206–213.
- Meng, K., Seekatz, B., Rossband, H., Worringer, U., Faller, H. & Vogel, H. (2009). Entwicklung eines standardisierten Rückenschulungsprogramms für die orthopädische Rehabilitation. *Die Rehabilitation*, 48 (6), 335–344.
- Meyer, C., Rumpf, H.-J., Hapke, U., Dilling, H. & John, U. (2000). Lebenszeitprävalenz psychischer Störungen in der erwachsenen Allgemeinbevölkerung. *Der Nervenarzt*, 71 (7), 535–542.
- Meyer, T., Quittan, M. & Rauch, A. (2017). Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF). In R. Crevenna (Hrsg.), *Kompodium Physikalische Medizin und Rehabilitation* (S. 71–88). Berlin: Springer.
- Mielck, A. (2012). Soziale Ungleichheit und Gesundheit. Empirische Belege für die zentrale Rolle der schulischen und beruflichen Bildung. In E. Brähler, J. Kiess, C. Schubert & W. Kiess (Hrsg.), *Gesund und gebildet. Voraussetzungen für eine moderne Gesellschaft* (S. 129–145). Göttingen: Vandenhoeck & Ruprecht.
- Mika, T. (2013). Risiken für eine Erwerbsminderung bei unterschiedlichen Berufsgruppen. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 56 (3), 391–398.
- Mittag, O., Reese, C. & Meffert, C. (2014). (Keine) Reha vor Rente: Analyse der Zugänge zur Erwerbsminderungsrente 2005 – 2009. *WSI-Mitteilungen*, 2, 149–155.
- Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. (2011). Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *Deutsche Medizinische Wochenschrift*, 136 (08), e9–e15.
- Moser, N.-T., Fischer, K. & Korsukéwitz, C. (2010). Prävention als Aufgabe der Rentenversicherung: Innovative Modelle ergänzen bewährte Konzepte. *Die Rehabilitation*, 49 (02), 80–86.

- Muche-Borowski, C. & Kopp, I. (2011). Wie eine Leitlinie entsteht. *Zeitschrift für Herz-, Thorax- und Gefäßchirurgie*, 25 (4), 217–223.
- van Muijen, P., Duijts, S. F. A., Kornet-van der Aa, D. A., van der Beek, A. J. & Anema, J. R. (2015). Work disability assessment of cancer survivors: Insurance physicians' perspectives. *Occupational Medicine*, 65 (7), 558–563.
- Mulrow, C. D. (1987). The medical review article: State of the science. *Annals of Internal Medicine*, 106 (3), 485–488.
- Neuderth, S. (2004). *Externe Qualitätssicherung durch Peer-Review*. Landau: Verlag Empirische Pädagogik.
- Neuhauser, H., Ellert, U. & Ziese, T. (2005). Chronische Rückenschmerzen in der Allgemeinbevölkerung in Deutschland 2002/2003: Prävalenz und besonders betroffene Bevölkerungsgruppen. *Das Gesundheitswesen*, 67 (10), 685–693.
- Niemeier, E. (2018). Lebensstandardsicherndes Niveau der gesetzlichen Rentenversicherung erforderlich. *Wirtschaftsdienst*, 98 (8), 600–604.
- Nowossadeck, E. (2018). Einfluss der demografischen Alterung auf die Inanspruchnahme der medizinischen Rehabilitation in Deutschland bis 2040. *Die Rehabilitation*. DOI: <https://doi.org/10.1055/a-0603-6864>.
- Nüchtern, E., Bahemann, A., Egdmann, W., van Essen, J., Gostomzyk, J., Hemmrich, K. et al. (2015). Soziale Sicherheit braucht Sozialmedizin Selbstverständnis von Ärztinnen und Ärzten in der sozialmedizinischen Begutachtung und Beratung. *Das Gesundheitswesen*, 77 (08/09), 580–585.
- Ostholt-Corsten, M. & Schuntermann, M. (2011). Ausgewählte Klassifikationssysteme. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 65–75). Berlin: Springer.
- Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J. et al. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49 (1), 47–53.
- Panic, N., Leoncini, E., de Belvis, G., Ricciardi, W. & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS ONE*, 8 (12), e83138.
- Pawils, S., Boettcher, A., Metzner, F., Plaumann, M. & Walter, U. (2015). Stellenwert und Umsetzung der Prävention in Deutschland – Eine bundesweite Befragung von Entscheidungsträgern in GKV-Spitzenverband, Bund, Ländern und Kommunen. *Das Gesundheitswesen*, 77 (S 01), S12–S13.
- Pengel, L. H. M., Herbert, R. D., Maher, C. G. & Refshauge, K. M. (2003). Acute low back pain: Systematic review of its prognosis. *BMJ*, 327 (7410), 323.
- Peters, E., Pritzkeleit, R., Beske, F. & Katalinic, A. (2010). Demografischer Wandel und Krankheitshäufigkeiten. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 53 (5), 417–426.
- Pfingsten, M. (2016). Risikofaktoren und psychobiologische Mechanismen der Chronifizierung. In H.-R. Casser, M. Hasenbring, A. Becker & Baron. R. (Hrsg.), *Rückenschmerzen und Nackenschmerzen* (S. 57–66). Berlin: Springer.
- Pfingsten, M. & Hildebrandt, J. (2007). Rückenschmerzen. In B. Kröner-Herwig, J. Frettlöh, R. Klinger & P. Nilges (Hrsg.), *Schmerzpsychotherapie* (S. 405–425). Berlin: Springer.

- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93 (3), 223–231.
- Prütz, F., Seeling, S., Ryl, L., Scheidt-Nave, C. E., Ziese, T. & Lampert, T. (2014). Welche Krankheiten bestimmen die Zukunft? In B. Badura, A. Ducki, H. Schröder, J. Klose & M. Meyer (Hrsg.), *Fehlzeiten-Report 2014: Erfolgreiche Unternehmen von morgen - gesunde Zukunft heute gestalten* (S. 113–126). Berlin: Springer.
- Rammstedt, B. (2010). Reliabilität, Validität, Objektivität. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 239–258). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Robert Koch-Institut (RKI). (2006). *Gesundheitsberichterstattung des Bundes Heft 33: Koronare Herzkrankheit und akuter Myokardinfarkt*. Berlin: Robert Koch-Institut.
- Robert Koch-Institut (RKI). (2011). *Beiträge zur Gesundheitsberichterstattung des Bundes - Daten und Fakten: Ergebnisse der Studie „Gesundheit in Deutschland aktuell 2009“*. Berlin: Robert Koch-Institut.
- Robert Koch-Institut (RKI). (2016). *Bericht zum Krebsgeschehen in Deutschland 2016*. Berlin: Robert Koch Institut.
- Rogers, R. & Payne, J. W. (2006). Damages and rewards: Assessment of malingered disorders in compensation cases. *Behavioral Sciences & the Law*, 24 (5), 645–658.
- Rosenbrock, R. (1995). Public Health als soziale Innovation. *Das Gesundheitswesen*, 57 (3), 140–144.
- Rosenbrock, R. (2004). Evidenzbasierung und Qualitätssicherung in der gesundheitsbezogenen Primärprävention. *Zeitschrift für Evaluation*, 1, 71–80.
- Sackett, D. L. (1997). Evidence-based medicine. *Seminars in Perinatology*, 21 (1), 3–5.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B. & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ*, 312 (7023), 71–72.
- Scheidt, S., Vavken, P., Jacobs, C., Koob, S., Cucchi, D., Kaup, E. et al. (2018). Systematische Reviews und Metaanalysen. *Zeitschrift für Orthopädie und Unfallchirurgie*. DOI: <https://doi.org/10.1055/a-0751-3156>.
- Schermelleh-Engel, K. & Werner, C. (2008). Methoden der Reliabilitätsbestimmung. In Moosbrugger & Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 113–133). Berlin: Springer.
- Schewe, P. (2017). *Ratgeber Erwerbsminderungsrente: Zuverlässiger Wegweiser und Praxishelfer für Versicherte*. Wiesbaden: Springer.
- Schilling, G. & Arnold, D. (2012). Therapiefolgen bei Krebs-Langzeitüberlebenden. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 55 (4), 493–500.
- Schiltenswolf, M. (2007). Begutachtung muskuloskelettaler Schmerzen. *Zeitschrift für Rheumatologie*, 66 (6), 473–483.
- Schliehe, F. & Ewert, T. (2013). Die Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) - Aktualisierung der problemorientierten Bestandsaufnahme. *Die Rehabilitation*, 52 (01), 40–50.
- Schmidt, C., Bernert, S. & Spyra, K. (2014). Zur Relevanz psychischer Komorbiditäten bei chronischem Rückenschmerz: Häufigkeitsbild, Erwerbsminderungsrenten und Reha-Nachsorge im Zeitverlauf der Reha-Kohorten 2002–2009. *Die Rehabilitation*, 53 (06), 384–389.

- Schmidt, C., Watzke, A.-B., Schulz, A., Baumeister, S., Freyberger, H. & Grabe, H.-J. (2013). Die Lebenszeitprävalenz psychischer Störungen in Vorpommern. *Psychiatrische Praxis*, 40 (04), 192–199.
- Schmidt, T., Lanquillon, S. & Ullmann, U. (2011). Kontroverse zu Beschwerdvalidierungsverfahren bei der Begutachtung psychischer Störungen. *Forensische Psychiatrie, Psychologie, Kriminologie*, 5 (3), 177–183.
- Schmucker, C., Motschall, E., Antes, G. & Meerpohl, J. J. (2013). Methoden des Evidence Mappings. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 56 (10), 1390–1397.
- Schneider, M. (2011). Rechtliche Rahmenbedingungen der Begutachtung im Verwaltungs- und Sozialgerichtsverfahren. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 27–44). Berlin: Springer.
- Schwaab, B. & Franz, I.-W. (2011). Herz-Kreislauf-Erkrankungen. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 289–324). Berlin: Springer.
- Schwegler, U., Anner, J., Boldt, C., Glässel, A., Lay, V., De Boer, W. E. L. et al. (2012). Aspects of functioning and environmental factors in medical work capacity evaluations of persons with chronic widespread pain and low back pain can be represented by a combination of applicable ICF Core Sets. *BMC Public Health*, 12, 1088.
- Schwegler, U., Selb, M., Escorpizo, R. & Stucki, G. (2015). Selection of ICF Core Sets for functioning assessment in disability evaluation toward the assignment to return to work programs and/or disability benefits. In R. Escorpizo, S. Brage, D. Homa & G. Stucki (Hrsg.), *Handbook of vocational rehabilitation and disability evaluation: Application and implementation of the ICF* (S. 413–435). Cham: Springer.
- Scriven, M. (1972). Die Methodologien der Evaluation. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 60–91). München: Piper & Co. Verlag.
- Seekatz, B., Haug, G., Mosler, G., Schwaab, B., Altstidl, R., Worringer, U. et al. (2013). Entwicklung und kurzfristige Effektivität eines standardisierten Schulungsprogramms für die Rehabilitation bei koronarer Herzkrankheit. *Die Rehabilitation*, 52 (5), 344–351.
- Selbmann, H. K. & Kopp, I. (2006). Leitlinien im Gesundheitswesen: Kompetenzen und Zuständigkeiten der AWMF. *Forum DKG*, 5, 5–8.
- Sennekamp, M., Kollwe, T. & Ochsendorf, F. (2018). Methodensammlung. In T. Kollwe, M. Sennekamp & F. Ochsendorf (Hrsg.), *Medizinindidaktik: Erfolgreich lehren und Wissen vermitteln* (S. 157–180). Berlin: Springer.
- Shaw, C. (2001). External assessment of health care. *BMJ*, 322 (7290), 851–854.
- Shaw, C. D. (2000). External quality mechanisms for health care: Summary of the ExPeRT project on visitatie, accreditation, EFQM and ISO assessment in European Union countries. External peer review techniques. European Foundation for quality management. International Organization for Standardization. *International Journal for Quality in Health Care*, 12 (3), 169–75.
- Shekelle, P. G., Woolf, S. H., Eccles, M. & Grimshaw, J. (1999). Clinical guidelines: Developing guidelines. *BMJ*, 318 (7183), 593–596.
- Simon, M. (2004). Die Begutachtung im Rahmen der sozialen Pflegeversicherung. *Journal of Public Health*, 12 (3), 218–228.

- Smith, M. A., Atherly, A. J., Kane, R. L. & Pacala, J. T. (1997). Peer review of the quality of care: Reliability and sources of variability for outcome and process assessments. *Journal of the American Medical Association*, 278 (19), 1573–1578.
- Solimene, M. C. (2010). Coronary heart disease in women: A challenge for the 21st century. *Clinics*, 65 (1), 99–106.
- Spies, M., Brütt, A., Freitag, M. & Buchholz, A. (2015). Anwendung der Internationalen Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) in der psychosomatischen Rehabilitation und Suchtrehabilitation in Deutschland – eine Bestandsaufnahme. *Die Rehabilitation*, 54 (05), 297–303.
- Springer, L., Stanne, M. E. & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69 (1), 21–51.
- Stengel, D., Bauwens, K. & Ekkernkamp, A. (2007). Qualitätssicherung und Qualitätsmanagement. *Orthopädie und Unfallchirurgie up2date*, 2 (6), 417–432.
- Stock, M. & Stock, W. G. (2003). Web of Knowledge: Wissenschaftliche Artikel, Patente und deren Zitationen: Der Wissenschaftsmarkt im Fokus. *PASSWORD*, 10, 30–37.
- Strahl, A., Gerlich, C., Alpers, G. W., Ehrmann, K., Gehrke, J., Müller-Garnn, A. et al. (2018). Development and evaluation of a standardized peer-training in the context of peer review for quality assurance in work capacity evaluation. *BMC Medical Education*, 18 (1), 135.
- Strahl, A., Gerlich, C., Wolf, H.-D. & Vogel, H. (2012). *Evaluation und Aufwandschätzung zum Qualitätssicherungsverfahren der sozialmedizinischen Begutachtung (EvaBegut): unveröffentlichter Abschlussbericht*. Würzburg: Universität Würzburg, Abteilung für Medizinische Psychologie und Psychotherapie, Medizinische Soziologie und Rehabilitationswissenschaften.
- Strahl, A., Rose, A., Brüggemann, S. & Vogel, H. (2016a). Systematische Literaturrecherche zur sozialmedizinischen Begutachtung bei Erwerbsminderungsrenten und Reha-Zugang. In Deutsch Rentenversicherung Bund (Hrsg.), *DRV-Schriften: Band 109: Tagungsband zum 25. Rehabilitationswissenschaftlichen Kolloquium: Gesundheitssystem im Wandel - Perspektiven der Rehabilitation vom 29. Februar bis 02. März 2016 in Aachen* (S. 106–108). Berlin: Deutsch Rentenversicherung Bund.
- Strahl, A., Rose, A., Brüggemann, S. & Vogel, H. (2016b). Evidenzbasierung in der sozialmedizinischen Begutachtung bei Erwerbsminderungsrenten und Reha-Zugang – Eine systematische Literaturrecherche. *Das Gesundheitswesen*, 78 (08/09), A5.
- Strahl, A. & Vogel, H. (2016). *Systematische Literaturrecherche zur sozialmedizinischen Begutachtung bei Erwerbsminderungsrenten und Reha-Zugang (LitBegut): unveröffentlichter Abschlussbericht*. Würzburg: Universität Würzburg, Abteilung für Medizinische Psychologie und Psychotherapie, Medizinische Soziologie und Rehabilitationswissenschaften.
- Strahl, A., Gerlich, C., Wolf, H.-D., Gehrke, J., Müller-Garnn, A. & Vogel, H. (2016). Qualitätssicherung in der Sozialmedizinischen Begutachtung durch Peer Review – ein Pilotprojekt der Deutschen Rentenversicherung. *Das Gesundheitswesen*, 78 (03), 156–160.
- Ströbl, V., Friedl-Huber, A. & Küffner, R. (2007). Beschreibungs- und Bewertungskriterien für Patientenschulungen. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 75, 11–14.

- Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14 (3), 207–222.
- Tuijn, S., Janssens, F., Robben, P. & van den Bergh, H. (2012). Reducing interrater variability and improving health care: A meta-analytical review. *Journal of Evaluation in Clinical Practice*, 18 (4), 887–895.
- Ueberschär, I. (2008). Qualitätssicherung bei der sozialmedizinischen Sachaufklärung in der Deutschen Rentenversicherung. *Das Gesundheitswesen*, 70 (11), 690–695.
- Ulsenheimer, K. (2001). Das Leid mit den Leitlinien. *Zentralblatt für Chirurgie*, 126 (9), 730–734.
- Vaccarino, V., Parsons, L., Every, N. R., Barron, H. V. & Krumholz, H. M. (1999). Sex-based differences in early mortality after myocardial infarction. *New England Journal of Medicine*, 341 (4), 217–225.
- Vaccarino, V., Parsons, L., Peterson, E. D., Rogers, W. J., Kiefe, C. I. & Canto, J. (2009). Sex differences in mortality after acute myocardial infarction: Changes from 1994 to 2006. *Archives of Internal Medicine*, 169 (19), 1767–1774.
- Verband Deutscher Rentenversicherungsträger (VDR). (2001). *DRV-Schriften, Band 21: Das ärztliche Gutachten für die gesetzliche Rentenversicherung*. Frankfurt/Main: Verband Deutscher Rentenversicherungsträger.
- Verband Deutscher Rentenversicherungsträger (VDR). (2004). *DRV Schriften Band 53: Abschlussbericht der Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung (SOMEKO)*. Frankfurt/Main: Verband Deutscher Rentenversicherungsträger.
- Verbrugge, L. M. (1984). Longer life but worsening health? Trends in health and mortality of middle-aged and older persons. *The Milbank Memorial Fund Quarterly. Health and Society*, 62 (3), 475–519.
- Vogel, H., Faller, H. & Reusch, A. (2016). Förderung und Erhaltung von Gesundheit: Prävention. In H. Faller & H. Lang (Hrsg.), *Medizinische Psychologie und Soziologie* (S. 329–358). Berlin: Springer.
- Vogt, V. (2017). Die Projektion der langfristigen Entwicklung der gesetzlichen Rentenversicherung. In K.-D. Altmeppen, F. Zschaler, H.-M. Zademach, C. Böttigheimer & M. Müller (Hrsg.), *Nachhaltigkeit in Umwelt, Wirtschaft und Gesellschaft* (S. 149–170). Wiesbaden: Springer.
- Walter, F., Geissler, N., Petermann, F. & Kobelt, A. (2015). Beeinflussen Instruktionen und die Motivation für eine Rente die Beschwerdedarstellung in Fragebogen?. *Die Rehabilitation*, 54 (03), 166–171.
- Walter, F., Petermann, F. & Kobelt, A. (2012). Beschwerdenuvalidierung: Ein aktueller Überblick. *Die Rehabilitation*, 51 (05), 342–348.
- Warm, K., Metzler, M. & Glaser, H.-J. (2018). Stoffwechsel und Verdauungstrakt. In Bundesarbeitsgemeinschaft für Rehabilitation e.V. (Hrsg.), *Rehabilitation: Vom Antrag bis zur Nachsorge – für Ärzte, Psychotherapeuten und andere Gesundheitsberufe* (S. 69–79). Berlin: Springer.
- Welti, F. (2001). Das neue Recht der Teilhabe und Rehabilitation (SGB IX): Chance und Verpflichtung. *Soziale Sicherheit*, 49 (6), 146–150.
- Welti, F. (2002). Das SGB IX in der Entwicklung des Sozialrechts. *Die Rehabilitation*, 41 (4), 268–273.

- Wenzel, T.-R. & Morfeld, M. (2016). Das biopsychosoziale Modell und die Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit: Beispiele für die Nutzung des Modells, der Teile und der Items. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 59 (9), 1125–1132.
- Wenzel, T.-R. & Morfeld, M. (2017). Nutzung der ICF in der medizinischen Rehabilitation in Deutschland: Anspruch und Wirklichkeit. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 60 (4), 386–393.
- World Health Organisation (WHO). (2001). *International Classification of Functioning, Disability and Health (ICF)*. Genf, Schweiz.
- World Health Organisation (WHO). (2006). *Zugewinn an Gesundheit: Die Europäische Strategie zu Prävention und Bekämpfung nichtübertragbarer Krankheiten; abschließender Entwurf*. Kopenhagen: WHO. Zugriff am 14.11.2018, verfügbar unter: <https://link.springer.com/content/pdf/10.1007%2F978-3-662-43531-1.pdf>
- Willms, G., Bramesfeld, A., Pottkämper, K., Broge, B. & Szecsenyi, J. (2013). Aktuelle Herausforderungen der externen Qualitätssicherung im deutschen Gesundheitswesen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 107 (8), 523–527.
- Wirtz, M. A. (2018a). Qualität. *Dorsch - Lexikon der Psychologie* (18. Aufl., S. 1276). Bern: Hogrefe.
- Wirtz, M. A. (2018b). Qualitätsmanagement. *Dorsch - Lexikon der Psychologie* (18. Aufl., S. 1280). Bern: Hogrefe.
- Wirtz, M. A. (2018c). Qualitätssicherung. *Dorsch - Lexikon der Psychologie* (18. Aufl., S. 1280). Bern: Hogrefe.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wolf-Kühn, N. & Morfeld, M. (2016). *Rehabilitationspsychologie*. Wiesbaden: Springer
- Worringen, U., Meng, K., Bitzer, E.-M., Brandes, I. & Faller, H. (2017). Entwicklung und aktueller Stand des Gesundheitstrainingsprogramms der Deutschen Rentenversicherung Bund. *Die Rehabilitation*, 56 (04), 232–239.
- Zillessen, E., Reichel, C. & Oehler, G. (2011). Krankheiten des Verdauungssystems. In Deutsche Rentenversicherung Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 363–390). Berlin: Springer.
- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID). (2018). *Datenbank PSYINDEX*. Zugriff am 25.11.2018, verfügbar unter: <https://www.psyindex.de/index.php?wahl=products&uwahl=fee&uuwahl=psyndexinfo>

Anhangsverzeichnis

Anhang 1:	einheitliches Deck- und Schlussblatt.....	S. 249
Anhang 2:	Suchstrategie und -begriffe der systematischen Literaturrecherche.....	S. 252
Anhang 3:	PRISMA-Statement: Qualitätskriterien für systematische Reviews.....	S. 262
Anhang 4:	Anhand der Volltexte von der Literaturrecherche ausgeschlossene Studien.....	S. 265
Anhang 5:	Manual zum Peer Review-Verfahren: Arbeitsversion 2010.....	S. 273
Anhang 6:	Prüffragenspezifische Codes der qualitativen Inhaltsanalyse der Einzelkriterien.....	S. 308
Anhang 7:	Prüffragenübergreifende Codes der qualitativen Inhaltsanalyse der Einzelkriterien.....	S. 316
Anhang 8:	Codes der qualitativen Inhaltsanalyse des übergeordneten Kriteriums.....	S. 318
Anhang 9:	Manual zum Peer Review-Verfahren.....	S. 320
Anhang 10:	Dokumentation der freitextlichen Mängelbegründungen während der standardisierten Peer-Schulung (Lernziele 2.2, 2.3, 2.5).....	S. 352

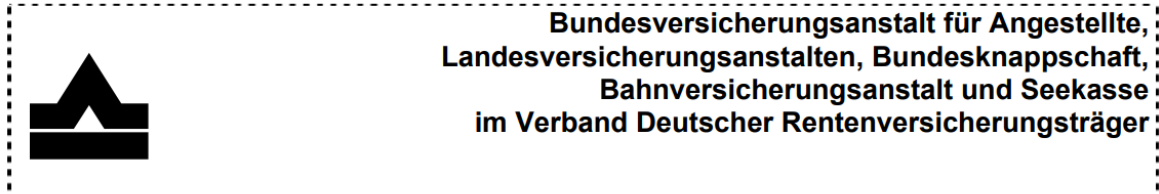
Anhang 1:

Einheitliches Deck- und Schlussblatt für sozialmedizinische Gutachten

(Stand: 2001)

Versicherungsnummer:

Blatt 1



**Bundesversicherungsanstalt für Angestellte,
Landesversicherungsanstalten, Bundesknappschaft,
Bahnversicherungsanstalt und Seekasse
im Verband Deutscher Rentenversicherungsträger**

ÄRZTLICHES GUTACHTEN FÜR DIE GESETZLICHE RENTENVERSICHERUNG

Angaben zum Begutachtungsanlass (Zutreffendes ankreuzen):

☐ Antrag auf Leistungen zur medizinischen Rehabilitation	☐ Antrag auf Rente wegen verminderter Erwerbsfähigkeit
☐ Antrag auf Leistungen zur beruflichen Rehabilitation	☐ Nachuntersuchung
☐ Antrag nach § 51 SGB V	☐ _____
☐ Antrag nach § 125 SGB III	

Name, Vorname, Geburtsname, Geburtsdatum:

Anschrift:

Die Angaben zur Person wurden überprüft

durch: ☐ Personalausweis/Reisepass ☐ Sonstiges _____

Besteht Arbeitsunfähigkeit?

- ☐ ja Seit wann? _____ Aufgrund welcher Erkrankungen?
- ☐ nein

Behandelnde Ärzte/Fachrichtung:

Gutachterin/Gutachter:

(Stempel)

Tag der Untersuchung:

Uhrzeit:

Das Gutachten ist entsprechend „Gliederung und Anforderungsprofil des ärztlichen Gutachtens für die gesetzliche Rentenversicherung“ abzufassen, d.h.

1. Anamnese
2. Untersuchungsbefunde
3. Diagnosen

4. Epikrise
5. Sozialmedizinische Leistungsbeurteilung

Versicherungsnummer:

Sozialmedizinische Leistungsbeurteilung

A. Letzte berufliche Tätigkeit

Bezeichnung der Tätigkeit	
Beurteilung des zeitlichen Umfanges, in dem die letzte berufliche Tätigkeit ausgeübt werden kann:	☐ 6 Stunden und mehr ☐ 3 bis unter 6 Stunden ☐ unter 3 Stunden

B. Positives und negatives Leistungsbild (allgemeiner Arbeitsmarkt) Zutreffendes ankreuzen (X), Mehrfachnennungen möglich

1. Positives Leistungsbild Folgende Arbeiten können verrichtet werden:	
Körperliche Arbeitsschwere	☐ schwere Arbeiten ☐ mittelschwere ☐ leichte bis mittelschwere ☐ leichte
Arbeitshaltung	im Stehen ☐ ständig ☐ über-wiegend ☐ zeit-weise im Gehen ☐ ständig ☐ über-wiegend ☐ zeit-weise im Sitzen ☐ ständig ☐ über-wiegend ☐ zeit-weise
Arbeitsorganisation	
☐ Tagesschicht ☐ Früh-/Spätschicht ☐ Nachtschicht	
☐ Keine wesentlichen Einschränkungen	

2. Negatives Leistungsbild Einschränkungen beziehen sich auf (Art/Ausmaß müssen differenziert unter Ziff. 3 beschrieben werden):
☐ geistig/psychische Belastbarkeit (Zu beachten sind insbesondere Konzentrations-/Reaktionsvermögen, Umstellungs-, Anpassungsvermögen, Verantwortung für Personen und Maschinen, Publikumsverkehr, Überwachung, Steuerung komplexer Arbeitsvorgänge).
☐ Sinnesorgane (Zu beachten sind insbesondere Seh-, Hör-, Sprach-, Sprech-, Tast- und Riechvermögen).
☐ Bewegungs-/Haltungsapparat (Zu beachten sind insbesondere Gebrauchsfähigkeit der Hände, häufiges Bücken, Ersteigen von Treppen, Leitern und Gerüsten, Heben, Tragen und Bewegen von Lasten, Gang- und Standsicherheit, Zwangshaltungen).
☐ Gefährdungs- und Belastungsfaktoren (Zu beachten sind insbesondere Nässe, Zugluft, extrem schwankende Temperaturen, inhalative Belastungen, Allergene, Lärm, Erschütterungen, Vibrationen, Tätigkeiten mit erhöhter Unfallgefahr, häufig wechselnde Arbeitszeiten).

3. Beschreibung des Leistungsbildes (insbesondere der unter Ziffer 2 genannten Einschränkungen).	
4. Beurteilung des zeitlichen Umfanges, in dem eine Tätigkeit entsprechend dem positiven und negativen Leistungsbild ausgeübt werden kann:	☐ 6 Stunden und mehr ☐ 3 bis unter 6 Stunden ☐ unter 3 Stunden

Bitte alle Blätter ausfüllen und jeweils mit Seitenzahl und Versicherungsnummer kennzeichnen!

Ärztliches Gutachten Schlussblatt Teil 2

Diagnosen	Diagnosenschlüssel ICD-10	-zusatz	-sicherheit
1.			
2.			
3.			

5. Die getroffenen Feststellungen gelten **seit** _____ (Tag, Monat, Jahr)

☐ Besserung wahrscheinlich bis _____

☐ Besserung unwahrscheinlich

(Begründung zu den Angaben in der Epikrise)

6. **Gesundheitsschäden/Leistungsminderung**
(vermutlich) verursacht **durch**

☐ Arbeitsunfall	☐ Wehrdienstbeschädigung
☐ Berufskrankheit	☐ Fremdverschulden (z.B. Unfall, gem. Opferentschädigungsgesetz)

7. Werden **Rehabilitationsleistungen** zur Besserung einer erheblich gefährdeten oder geminderten Leistungsfähigkeit vorgeschlagen?

Medizinische Rehabilitation	☐ Nein	☐ Ja
Berufliche Rehabilitation	☐ Nein	☐ Ja

(Begründung in der Epikrise)

8. Kann bei Vorliegen einer psychischen Krankheit oder einer geistigen oder seelischen Behinderung die Versicherte/der Versicherte ihre/seine Angelegenheiten selbst besorgen?

☐ Nein ☐ Ja

9. Wurden wesentliche, bisher unbekannte Erkrankungen oder Krankheitskomplikationen festgestellt?

☐ Nein ☐ Ja

Am _____ wurde _____ davon unterrichtet.
(Datum) (Versicherte(r))

Am _____ wurde _____ davon unterrichtet.
(Datum) (behandelnde Ärztin/behandelnder Arzt)

10. Für die Fahrt zur Untersuchung war(en) erforderlich:

a)	öffentliche Verkehrsmittel	☐	Nein	☐	Ja
b)	PKW	☐	Nein	☐	Ja
c)	eine Begleitperson	☐	Nein	☐	Ja

(Datum des Gutachtenabschlusses)

Anhang 2

Suchstrategie und -begriffe der systematischen Literaturrecherche

Datenbank	Suchterminus	Treffer
Medline via OVID	((((krebs.mp. or exp CARCINOMA, KREBS 2/) OR (Cancer.mp. or Neoplasms/) OR Neubildungen.mp. OR Neubildung*.mp. OR tumor.mp. OR tumour.mp. OR exp Carcinoma/ OR (exp Stroke/ or exp Brain Ischemia/ or Schlaganfall.mp. or exp Cerebral Infarction/) OR apoplexy.mp. OR (COPD.mp. or exp Pulmonary Disease, Chronic Obstructive/) OR CED.mp. OR chronisch entzündliche Darmerkrankung.mp. OR (exp Colitis, Ulcerative/ or exp Crohn Disease/ or exp Inflammatory Bowel Diseases/) OR (exp Coronary Disease/ or exp Myocardial Ischemia/ or chronic ischemic heart disease.mp.) OR koronare Herzkrankheit.mp. OR (Diabetes mellitus.mp. or exp Diabetes Mellitus/) OR (Vasculitis/ or Peripheral Arterial Disease/ or Arteriosklerose.mp.) OR (arteriosclerosis.mp. or exp ARTERIOSCLEROSIS/) OR (exp Cardiomyopathies/ or Kardiomyopathie.mp.) OR (Heart Failure/ or Herzinsuffizienz.mp.) OR heart insufficiency.mp. OR cardiac insufficiency.mp.) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. OR functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit yr="1998 -Current"	2.183
	((((osteoarthritis.mp. or exp OSTEOARTHRITIS, HIP/ or exp OSTEOARTHRITIS/ or exp OSTEOARTHRITIS, SPINE/ or exp OSTEOARTHRITIS, KNEE/) OR Arthrose.mp. OR arthrosis.mp. OR (joint disease*.mp. or exp Joint Diseases/) OR Rueckenschmerzen.mp. OR (back pain.mp. or exp Back Pain/) OR Spondylopathien.mp. OR (Spinal Osteophytosis/ or Cervical Vertebrae/ or spondylopathies.mp. or Spinal Diseases/) OR (exp Sciatica/ or Lumboischialgie.mp. or exp Radiculopathy/) OR lumboischialgia.mp.) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. OR functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR	1.357

	Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit yr="1998 -Current"	
	(((psychological disorder.mp. OR psychische Störung*.mp. OR affektive Störung*.mp. OR affective disorder*.mp. OR (exp Mental Disorders/ or psychiatric disorder*.mp) OR (exp Mood Disorders/ or mood disorder*.mp.) OR (Depression.mp. or exp DEPRESSION/) OR Psychische und Verhaltensstörungen durch Alkohol.mp. OR Psychological and behavioural disorders resulting from psychotropic substances.mp. OR (exp Substance-Related Disorders/ or exp Opioid-Related Disorders/ or alkohol abus*.mp. or exp Alcoholism/) OR somatoforme Störung.mp. OR (Somatoform Disorders/ or somatoform disorder*.mp.) OR (exp Psychotic Disorders/ or exp Schizophrenia/ or exp Schizophrenia, Paranoid/ or Schizophrenie.mp.) OR Angststörung.mp. OR (exp Anxiety Disorders/ or exp Anxiety/ or anxiety disorder*.mp. or exp Phobic Disorders/) OR (exp Multiple Sclerosis/ or Multiple Sklerose.mp.)) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. Or functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit yr="1998 -Current"	3.785
Web of Science Core Collection	(TS=(krebs OR *krebs OR carcinom* OR Cancer OR Neoplasms OR Neubildung* OR tumor OR tumour OR Stroke OR Brain Ischemia OR Schlaganfall OR Cerebral Infarct* OR apoplexy OR COPD OR Chronic Obstructive Pulmonary Disease* OR CED OR chronisch entzündliche Darmerkrankung* OR Ulcerative Colitis OR Crohn Disease OR Inflammatory Bowel Diseases OR Coronary Disease OR Myocard* Ischemia OR chronic ischemic heart disease OR koronare Herzkrankheit* OR Diabetes mellitus OR Vasculitis OR Peripheral Arterial Disease OR Arteriosklerose OR arteriosclerosis OR Cardiomyopath* OR Kardiomyopath* OR Heart Failure OR Herzinsuffizienz OR heart insufficienc* OR cardiac insufficienc*) AND TS=(Assess* OR Insurance Claim Review OR Work Capacity Evaluation OR claim assessment OR Disability Evaluation OR disability assessment OR employee performance appraisal OR Employee Performance Appraisal OR Work Capacity Evaluation OR functional capacity assessment OR functional capacity evaluation OR functional evaluation OR functioning-based assessment OR impairment assessment OR impairment evaluation OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR incapacity assessment OR insurance medicine OR Judgment OR medical examination OR needs assessment OR Health Services Needs and Demand OR Needs Assessment OR physical examination OR Physical Examination OR practice guideline OR Practice Guideline OR socio-medical assessment OR socio-medical evaluation OR symptom assessment OR Symptom Assessment OR symptom evaluation OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND TS=(compensation OR Compensation and Redress OR Disabled Persons OR disability benefit OR disability pension OR Retirement OR disability retirement pay OR functional capacity OR pension OR Social Security OR social security benefit OR Social Welfare OR Erwerbsminderungsrente OR Fruerente)) NOT TS=(rat OR rat* OR mouse OR mice) AND limit yr="1998 -Current"	3.639
	TS=(osteoarthritis hip OR osteoarthritis knee OR osteoarthritis spine OR osteoarthritis OR Arthrose OR Arthrosis OR joint disease* OR Rueckenschmerzen OR back pain OR Spinal Osteophytosis OR Cervical Vertebrae OR spondylopath* OR Spinal Diseas* OR Sciatica OR Lumboischialgie OR Radiculopath* OR lumboischialgia) AND TS=(Assess* OR Insurance Claim Review OR Work Capacity Evaluation OR claim assessment OR	1.173

	Disability Evaluation OR disability assessment OR employee performance appraisal OR Employee Performance Appraisal OR Work Capacity Evaluation OR functional capacity assessment OR functional capacity evaluation OR functional evaluation OR functioning-based assessment OR impairment assessment OR impairment evaluation OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR incapacity assessment OR insurance medicine OR Judgment OR medical examination OR needs assessment OR Health Services Needs and Demand OR Needs Assessment OR physical examination OR Physical Examination OR practice guideline OR Practice Guideline OR socio-medical assessment OR socio-medical evaluation OR symptom assessment OR Symptom Assessment OR symptom evaluation OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND TS=(compensation OR Compensation and Redress OR Disabled Persons OT disability benefit OR disability pension OR Retirement OR disability retirement pay OR functional capacity OR pension OR Social Security OR social security benefit OR Social Welfare OR Erwerbsminderungsrente OR Fruerente)) NOT TS=(rat OR rat* OR mouse OR mice) AND limit yr="1998 -Current"	
	TS=(psychological disorder* OR psychische Stoerung* OR affektive Stoerung* OR affective disorder* OR Mental Disorder* OR psychiatric disorder* OR Mood Disorder* OR Depression OR major depression OR Psychische und Verhaltensstoerungen durch Alkohol OR Psychological and behavioural disorders resulting from psychotropic substances OR Substance-Related Disorder* OR Opioid-Related Disorder* OR alkohol abus* OR Alcoholism OR somatoforme Stoerung OR somatoform disorder* OR Psychotic Disorder* OR Schizophrenia OR Schizophrenia OR Schizophrenie OR Angststoerung OR Anxiety Disorder* OR Anxiety OR anxiety disorder* OR Phobic Disorders* OR Multiple Sclerosis OR Multiple Sklerose) AND TS=(Assess* OR Insurance Claim Review OR Work Capacity Evaluation OR claim assessment OR Disability Evaluation OR disability assessment OR employee performance appraisal OR Employee Performance Appraisal OR Work Capacity Evaluation OR functional capacity assessment OR functional capacity evaluation OR functional evaluation OR functioning-based assessment OR impairment assessment OR impairment evaluation OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR incapacity assessment OR insurance medicine OR Judgment OR medical examination OR needs assessment OR Health Services Needs and Demand OR Needs Assessment OR physical examination OR Physical Examination OR practice guideline OR Practice Guideline OR socio-medical assessment OR socio-medical evaluation OR symptom assessment OR Symptom Assessment OR symptom evaluation OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND TS=(compensation OR Compensation and Redress OR Disabled Persons OT disability benefit OR disability pension OR Retirement OR disability retirement pay OR functional capacity OR pension OR Social Security OR social security benefit OR Social Welfare OR Erwerbsminderungsrente OR Fruerente)) NOT TS=(rat OR rat* OR mouse OR mice) AND limit yr="1998 -Current"	2.135
PsycINFO via OVID	((((krebs.mp. or exp CARCINOMA, KREBS 2/) OR (Cancer.mp. or Neoplasms/) OR Neubildungen.mp. OR Neubildung*.mp. OR tumor.mp. OR tumour.mp. OR exp Carcinoma/ OR (exp Stroke/ or exp Brain Ischemia/ or Schlaganfall.mp. or exp Cerebral Infarction/) OR apoplexy.mp. OR (COPD.mp. or exp Pulmonary Disease, Chronic Obstructive/) OR CED.mp. OR chronisch entzuendliche Darmerkrankung.mp. OR (exp Colitis, Ulcerative/ or exp Crohn Disease/ or exp Inflammatory Bowel Diseases/) OR (exp Coronary Disease/ or exp Myocardial Ischemia/ or chronic ischemic heart disease.mp.) OR koronare Herzkrankheit.mp. OR (Diabetes mellitus.mp. or exp Diabetes Mellitus/) OR (Vasculitis/ or Peripheral Arterial Disease/ or Arteriosklerose.mp.) OR (arteriosclerosis.mp. or exp ARTERIOSCLEROSIS/) OR (exp Cardiomyopathies/ or Kardiomyopathie.mp.) OR (Heart Failure/ or Herzinsuffizienz.mp.) OR heart insufficiency.mp. OR cardiac insufficiency.mp.) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. OR functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice	94

Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.) AND limit 92 to yr="1998 -Current" AND limit yr="1998 -Current"	
((osteoarthritis.mp. or exp OSTEOARTHRITIS, HIP/ or exp OSTEOARTHRITIS/ or exp OSTEOARTHRITIS, SPINE/ or exp OSTEOARTHRITIS, KNEE/) OR Arthrose.mp. OR arthrosis.mp. OR (joint disease*.mp. or exp Joint Diseases/) OR Rueckenschmerzen.mp. OR (back pain.mp. or exp Back Pain/) OR Spondylopathien.mp. OR (Spinal Osteophytosis/ or Cervical Vertebrae/ or spondylopathies.mp. or Spinal Diseases/) OR (exp Sciatica/ or Lumboischialgie.mp. or exp Radiculopathy/) OR lumboischialgia.mp.) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. Or functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.) AND limit 92 to yr="1998 -Current" AND limit yr="1998 -Current"	49
((((psychological disorder.mp. OR psychische Störung*.mp. OR affektive Störung*.mp. OR affective disorder*.mp. OR (exp Mental Disorders/ or psychiatric disorder*.mp) OR (exp Mood Disorders/ or mood disorder*.mp.) OR (Depression.mp. or exp DEPRESSION/) OR Psychische und Verhaltensstörungen durch Alkohol.mp. OR Psychological and behavioural disorders resulting from psychotropic substances.mp. OR (exp Substance-Related Disorders/ or exp Opioid-Related Disorders/ or alkohol abus*.mp. or exp Alcoholism/) OR somatoforme Störung.mp. OR (Somatoform Disorders/ or somatoform disorder*.mp.) OR (exp Psychotic Disorders/ or exp Schizophrenia/ or exp Schizophrenia, Paranoid/ or Schizophrenie.mp.) OR Angststörung.mp. OR (exp Anxiety Disorders/ or exp Anxiety/ or anxiety disorder*.mp. or exp Phobic Disorders/) OR (exp Multiple Sclerosis/ or Multiple Sklerose.mp.)) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. Or functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp.	588

	or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit 92 to yr="1998 -Current"	
PSYINDEX via OVID	((((krebs.mp. or exp CARCINOMA, KREBS 2/) OR (Cancer.mp. or Neoplasms/) OR Neubildungen.mp. OR Neubildung*.mp. OR tumor.mp. OR tumour.mp. OR exp Carcinoma/ OR (exp Stroke/ or exp Brain Ischemia/ or Schlaganfall.mp. or exp Cerebral Infarction/) OR apoplexy.mp. OR (COPD.mp. or exp Pulmonary Disease, Chronic Obstructive/) OR CED.mp. OR chronisch entzündliche Darmerkrankung.mp. OR (exp Colitis, Ulcerative/ or exp Crohn Disease/ or exp Inflammatory Bowel Diseases/) OR (exp Coronary Disease/ or exp Myocardial Ischemia/ or chronic ischemic heart disease.mp.) OR koronare Herzkrankheit.mp. OR (Diabetes mellitus.mp. or exp Diabetes Mellitus/) OR (Vasculitis/ or Peripheral Arterial Disease/ or Arteriosklerose.mp.) OR (arteriosclerosis.mp. or exp ARTERIOSCLEROSIS/) OR (exp Cardiomyopathies/ or Kardiomyopathie.mp.) OR (Heart Failure/ or Herzinsuffizienz.mp.) OR heart insufficiency.mp. OR cardiac insufficiency.mp.) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. OR functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit 92 to yr="1998 -Current" AND limit yr="1998 -Current"	10
	((((osteoarthritis.mp. or exp OSTEOARTHRITIS, HIP/ or exp OSTEOARTHRITIS/ or exp OSTEOARTHRITIS, SPINE/ or exp OSTEOARTHRITIS, KNEE/) OR Arthrose.mp. OR arthrosis.mp. OR (joint disease*.mp. or exp Joint Diseases/) OR Rueckenschmerzen.mp. OR (back pain.mp. or exp Back Pain/) OR Spondylopathien.mp. OR (Spinal Osteophytosis/ or Cervical Vertebrae/ or spondylopathies.mp. or Spinal Diseases/) OR (exp Sciatica/ or Lumboischialgie.mp. or exp Radiculopathy/) OR lumboischialgia.mp.) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. OR functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit 92 to yr="1998 -Current"	12
	((((psychological disorder.mp. OR psychische Störung*.mp. OR affektive Störung*.mp. OR affective disorder*.mp. OR (exp Mental Disorders/ or psychiatric disorder*.mp) OR	82

	(exp Mood Disorders/ or mood disorder*.mp.) OR (Depression.mp. or exp DEPRESSION/) OR Psychische und Verhaltensstörungen durch Alkohol.mp. OR Psychological and behavioural disorders resulting from psychotropic substances.mp. OR (exp Substance-Related Disorders/ or exp Opioid-Related Disorders/ or alkohol abus*.mp. or exp Alcoholism/) OR somatoforme Störung.mp. OR (Somatoform Disorders/ or somatoform disorder*.mp.) OR (exp Psychotic Disorders/ or exp Schizophrenia/ or exp Schizophrenia, Paranoid/ or Schizophrenie.mp.) OR Angststörung.mp. OR (exp Anxiety Disorders/ or exp Anxiety/ or anxiety disorder*.mp. or exp Phobic Disorders/) OR (exp Multiple Sclerosis/ or Multiple Sklerose.mp.)) AND (Assess*.mp. OR (exp "Insurance Claim Review"/ or exp Work Capacity Evaluation/ or claim assessment.mp. or exp) OR Insurance, Disability/ OR (exp Disability Evaluation/ or disability assessment.mp.) OR (employee performance appraisal.mp. or exp Employee Performance Appraisal/) OR (exp Work Capacity Evaluation/ or functional capacity assessment.mp.) OR functional capacity evaluation.mp. OR functional evaluation.mp. OR functioning-based assessment.mp. OR impairment assessment.mp. OR impairment evaluation.mp. OR impairment of earning capacity assessment.mp. OR impairment of earning capacity evaluation.mp. OR incapacity assessment.mp. OR (exp Insurance, Health/ or insurance medicine.mp) OR (exp Judgment/ or judgement.mp.) OR medical examination.mp. OR (needs assessment.mp. or exp "Health Services Needs and Demand"/ or exp Needs Assessment/) OR (physical examination.mp. or exp Physical Examination/) OR (practice guideline.mp. or Practice Guideline/) OR socio-medical assessment.mp. OR socio-medical evaluation.mp. OR (symptom assessment.mp. or exp Symptom Assessment/) OR symptom evaluation.mp. OR Sozialmedizinische Begutachtung.mp. OR Begutachtung.mp. OR Leistungsbeurteilung.mp.) AND (compensation.mp. or "Compensation and Redress"/) OR (Disabled Persons/ or disability benefit.mp.) OR disability pension.mp. OR (exp Retirement/ or disability retirement pay.mp.) OR (functional capacity.mp. OR pension.mp. or exp Pensions/) OR (exp Social Security/ or social security benefit.mp. or exp Social Welfare/) OR Erwerbsminderungsrente.mp.)) NOT (exp rat/ or rat*.mp. or exp mouse/ or mice.mp.)) AND limit 92 to yr="1998 -Current"	
CENTRAL	((("Krebs"):ti,ab,kw OR ("carcinoma"):ti,ab,kw OR MeSH descriptor: [Carcinoma] explode all trees OR ("Cancer"):ti,ab,kw OR MeSH descriptor: [Neoplasms] explode all trees OR Neoplasm* OR MeSH descriptor: [Neoplasms] explode all trees OR Neubildung* OR (tumor):ti,ab,kw OR MeSH descriptor: [Neoplasms] explode all trees OR (tumour):ti,ab,kw OR ("stroke"):ti,ab,kw OR MeSH descriptor: [Stroke] explode all trees OR (Brain Ischemia):ti,ab,kw OR MeSH descriptor: [Brain Ischemia] explode all trees OR Schlaganfall OR (Cerebral Infarct*):ti,ab,kw OR MeSH descriptor: [Cerebral Infarction] explode all trees OR (apoplexy):ti,ab,kw OR MeSH descriptor: [Stroke] explode all trees OR (COPD):ti,ab,kw OR MeSH descriptor: [Pulmonary Disease, Chronic Obstructive] explode all trees OR ("chronic obstructive pulmonary disease"):ti,ab,kw OR (CED):ti,ab,kw OR chronisch entzündliche Darmerkrankung* OR Ulcerative Colitis OR ("ulcerative colitis"):ti,ab,kw OR (Crohn Disease):ti,ab,kw OR (Inflammatory Bowel Diseases):ti,ab,kw OR MeSH descriptor: [Crohn Disease] explode all trees OR (Coronary Disease):ti,ab,kw OR MeSH descriptor: [Coronary Disease] explode all trees OR (Myocardial Ischemia):ti,ab,kw OR (chronic ischemic heart disease):ti,ab,kw OR (koronare Herzkrankheit):ti,ab,kw OR (Diabetes mellitus):ti,ab,kw OR MeSH descriptor: [Diabetes Mellitus] explode all trees OR ("vasculitis"):ti,ab,kw OR MeSH descriptor: [Vasculitis] explode all trees OR (Peripheral Arterial Disease):ti,ab,kw OR Arteriosklerose OR (arteriosclerosis):ti,ab,kw OR MeSH descriptor: [Arteriosclerosis] explode all trees OR (Cardiomyopath*):ti,ab,kw OR MeSH descriptor: [Cardiomyopathies] explode all trees OR Kardiomyopath* OR ("heart failure"):ti,ab,kw OR (Herzinsuffizienz):ti,ab,kw OR (heart insufficiency):ti,ab,kw OR (cardiac insufficiency):ti,ab,kw) AND ((Assess*):ti,ab,kw OR Insurance Claim Review OR (Insurance Claim Review):ti,ab,kw OR (Work Capacity Evaluation):ti,ab,kw OR MeSH descriptor: [Work Capacity Evaluation] explode all trees OR (claim assessment):ti,ab,kw OR (Disability Evaluation):ti,ab,kw OR MeSH descriptor: [Disability Evaluation] explode all trees OR (disability assessment):ti,ab,kw OR (employee performance appraisal):ti,ab,kw OR (functional capacity assessment):ti,ab,kw OR (functional evaluation):ti,ab,kw OR (functioning-based assessment):ti,ab,kw OR impairment assessment OR (impairment assessment):ti,ab,kw OR (impairment evaluation):ti,ab,kw OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR (insurance medicine):ti,ab,kw OR insurance medicine OR (Judgment):ti,ab,kw OR (medical examination):ti,ab,kw OR (needs assessment):ti,ab,kw OR MeSH descriptor: [Independent Medical Evaluation] explode all trees OR MeSH descriptor: [Judgment] explode all trees OR (Health Services Needs and Demand):ti,ab,kw OR MeSH descriptor: [Needs Assessment] explode all trees OR (physical examination):ti,ab,kw OR MeSH descriptor: [Physical Examination] explode all trees OR	956

(practice guideline):ti,ab,kw OR (socio-medical assessment):ti,ab,kw OR socio-medical evaluation OR (socio-medical evaluation):ti,ab,kw OR socio-medical assessment OR (symptom assessment):ti,ab,kw OR MeSH descriptor: [Symptom Assessment] explode all trees OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND ((compensation):ti,ab,kw OR MeSH descriptor: [Compensation and Redress] explode all trees OR MeSH descriptor: [Disabled Persons] explode all trees OR (disability benefit):ti,ab,kw OR (disability pension):ti,ab,kw OR disability pension OR (disability retirement):ti,ab,kw OR MeSH descriptor: [Retirement] explode all trees OR MeSH descriptor: [Pensions] explode all trees OR (functional capacity):ti,ab,kw OR ("functional capacity assessment"):ti,ab,kw OR (pension*):ti,ab,kw OR ("Social Security Disability Insurance"):ti,ab,kw OR Social Security OR (social security benefit):ti,ab,kw OR Erwerbsminderungsrente)) NOT ((rat*):ti,ab,kw OR rat* OR mouse OR mice OR (mouse):ti,ab,kw OR (mice):ti,ab,kw OR MeSH descriptor: [Rats] explode all trees OR MeSH descriptor: [Mice] explode all trees)) AND limit yr="1998 -Current"	
((osteoarthritis):ti,ab,kw OR MeSH descriptor: [Osteoarthritis] explode all trees OR (Arthrose):ti,ab,kw OR (arthrosis):ti,ab,kw OR MeSH descriptor: [Joint Diseases] explode all trees OR (Rueckenschmerzen):ti,ab,kw OR ("back pain"):ti,ab,kw OR MeSH descriptor: [Back Pain] explode all trees OR MeSH descriptor: [Low Back Pain] explode all trees OR Spondylopath* OR (Spondylopath*):ti,ab,kw OR MeSH descriptor: [Spondylarthropathies] explode all trees OR (Spinal Osteophytosis):ti,ab,kw OR (Cervical Vertebrae):ti,ab,kw OR (Spinal Diseases):ti,ab,kw OR ("sciatica"):ti,ab,kw OR MeSH descriptor: [Sciatica] explode all trees OR Lumboischialgie OR ("radiculopathy"):ti,ab,kw OR MeSH descriptor: [Radiculopathy] explode all trees) AND ((Assess*):ti,ab,kw OR Insurance Claim Review OR (Insurance Claim Review):ti,ab,kw OR (Work Capacity Evaluation):ti,ab,kw OR MeSH descriptor: [Work Capacity Evaluation] explode all trees OR (claim assessment):ti,ab,kw OR (Disability Evaluation):ti,ab,kw OR MeSH descriptor: [Disability Evaluation] explode all trees OR (disability assessment):ti,ab,kw OR (employee performance appraisal):ti,ab,kw OR (functional capacity assessment):ti,ab,kw OR (functional evaluation):ti,ab,kw OR (functioning-based assessment):ti,ab,kw OR impairment assessment OR (impairment assessment):ti,ab,kw OR (impairment evaluation):ti,ab,kw OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR (insurance medicine):ti,ab,kw OR insurance medicine OR (Judgment):ti,ab,kw OR (medical examination):ti,ab,kw OR (needs assessment):ti,ab,kw OR MeSH descriptor: [Independent Medical Evaluation] explode all trees OR MeSH descriptor: [Judgment] explode all trees OR (Health Services Needs and Demand):ti,ab,kw OR MeSH descriptor: [Needs Assessment] explode all trees OR (physical examination):ti,ab,kw OR MeSH descriptor: [Physical Examination] explode all trees OR (practice guideline):ti,ab,kw OR (socio-medical assessment):ti,ab,kw OR socio-medical evaluation OR (socio-medical evaluation):ti,ab,kw OR socio-medical assessment OR (symptom assessment):ti,ab,kw OR MeSH descriptor: [Symptom Assessment] explode all trees OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND ((compensation):ti,ab,kw OR MeSH descriptor: [Compensation and Redress] explode all trees OR MeSH descriptor: [Disabled Persons] explode all trees OR (disability benefit):ti,ab,kw OR (disability pension):ti,ab,kw OR disability pension OR (disability retirement):ti,ab,kw OR MeSH descriptor: [Retirement] explode all trees OR MeSH descriptor: [Pensions] explode all trees OR (functional capacity):ti,ab,kw OR ("functional capacity assessment"):ti,ab,kw OR (pension*):ti,ab,kw OR ("Social Security Disability Insurance"):ti,ab,kw OR Social Security OR (social security benefit):ti,ab,kw OR Erwerbsminderungsrente)) NOT ((rat*):ti,ab,kw OR rat* OR mouse OR mice OR (mouse):ti,ab,kw OR (mice):ti,ab,kw OR MeSH descriptor: [Rats] explode all trees OR MeSH descriptor: [Mice] explode all trees)) AND limit yr="1998 -Current"	578
(psychological disorder):ti,ab,kw OR (psychische Störung):ti,ab,kw OR (affektive Störung):ti,ab,kw OR (affective disorder):ti,ab,kw OR MeSH descriptor: [Mood Disorders] explode all trees OR (psychiatric disorder):ti,ab,kw OR MeSH descriptor: [Mental Disorders] explode all trees OR ("mood disorder"):ti,ab,kw OR depression OR MeSH descriptor: [Depression] explode all trees OR ("substance related disorders"):ti,ab,kw OR (Opioid-Related Disorders):ti,ab,kw OR (alkohol abus*):ti,ab,kw OR alkohol abus* OR ("alcoholism"):ti,ab,kw OR somatoforme Störung OR Somatoform Disorders OR MeSH descriptor: [Somatoform Disorders] explode all trees OR (somatoform disorder):ti,ab,kw OR (Psychotic Disorders):ti,ab,kw OR MeSH descriptor: [Psychotic Disorders] explode all trees OR ("schizophrenia"):ti,ab,kw OR MeSH descriptor: [Schizophrenia] explode all trees OR (Schizophrenie):ti,ab,kw OR (Angststörung):ti,ab,kw OR (Anxiety Disorder):ti,ab,kw OR MeSH descriptor: [Anxiety Disorders] explode all trees OR ("anxiety"):ti,ab,kw OR (Phobic Disorders):ti,ab,kw OR MeSH descriptor: [Phobic	436

	Disorders] explode all trees OR ("multiple sclerosis"):ti,ab,kw OR MeSH descriptor: [Multiple Sclerosis] explode all trees) AND ((Assess*):ti,ab,kw OR Insurance Claim Review OR (Insurance Claim Review):ti,ab,kw OR (Work Capacity Evaluation):ti,ab,kw OR MeSH descriptor: [Work Capacity Evaluation] explode all trees OR (claim assessment):ti,ab,kw OR (Disability Evaluation):ti,ab,kw OR MeSH descriptor: [Disability Evaluation] explode all trees OR (disability assessment):ti,ab,kw OR (employee performance appraisal):ti,ab,kw OR (functional capacity assessment):ti,ab,kw OR (functional evaluation):ti,ab,kw OR (functioning-based assessment):ti,ab,kw OR impairment assessment OR (impairment assessment):ti,ab,kw OR (impairment evaluation):ti,ab,kw OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR (insurance medicine):ti,ab,kw OR insurance medicine OR (Judgment):ti,ab,kw OR (medical examination):ti,ab,kw OR (needs assessment):ti,ab,kw OR MeSH descriptor: [Independent Medical Evaluation] explode all trees OR MeSH descriptor: [Judgment] explode all trees OR (Health Services Needs and Demand):ti,ab,kw OR MeSH descriptor: [Needs Assessment] explode all trees OR (physical examination):ti,ab,kw OR MeSH descriptor: [Physical Examination] explode all trees OR (practice guideline):ti,ab,kw OR (socio-medical assessment):ti,ab,kw OR socio-medical evaluation OR (socio-medical evaluation):ti,ab,kw OR socio-medical assessment OR (symptom assessment):ti,ab,kw OR MeSH descriptor: [Symptom Assessment] explode all trees OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND ((compensation):ti,ab,kw OR MeSH descriptor: [Compensation and Redress] explode all trees OR MeSH descriptor: [Disabled Persons] explode all trees OR (disability benefit):ti,ab,kw OR (disability pension):ti,ab,kw OR disability pension OR (disability retirement):ti,ab,kw OR MeSH descriptor: [Retirement] explode all trees OR MeSH descriptor: [Pensions] explode all trees OR (functional capacity):ti,ab,kw OR ("functional capacity assessment"):ti,ab,kw OR (pension*):ti,ab,kw OR ("Social Security Disability Insurance"):ti,ab,kw OR Social Security OR (social security benefit):ti,ab,kw OR Erwerbsminderungsrente) NOT ((rat*):ti,ab,kw OR rat* OR mouse OR mice OR (mouse):ti,ab,kw OR (mice):ti,ab,kw OR MeSH descriptor: [Rats] explode all trees OR MeSH descriptor: [Mice] explode all trees)) AND limit yr="1998-Current"	
CDSR	((("Krebs"):ti,ab,kw OR ("carcinoma"):ti,ab,kw OR MeSH descriptor: [Carcinoma] explode all trees OR ("Cancer"):ti,ab,kw OR MeSH descriptor: [Neoplasms] explode all trees OR Neoplasm* OR MeSH descriptor: [Neoplasms] explode all trees OR Neubildung* OR (tumor):ti,ab,kw OR MeSH descriptor: [Neoplasms] explode all trees OR (tumour):ti,ab,kw OR ("stroke"):ti,ab,kw OR MeSH descriptor: [Stroke] explode all trees OR (Brain Ischemia):ti,ab,kw OR MeSH descriptor: [Brain Ischemia] explode all trees OR Schlaganfall OR (Cerebral Infarct*):ti,ab,kw OR MeSH descriptor: [Cerebral Infarction] explode all trees OR (apoplexy):ti,ab,kw OR MeSH descriptor: [Stroke] explode all trees OR (COPD):ti,ab,kw OR MeSH descriptor: [Pulmonary Disease, Chronic Obstructive] explode all trees OR ("chronic obstructive pulmonary disease"):ti,ab,kw OR (CED):ti,ab,kw OR chronisch entzündliche Darmerkrankung* OR Ulcerative Colitis OR ("ulcerative colitis"):ti,ab,kw OR (Crohn Disease):ti,ab,kw OR (Inflammatory Bowel Diseases):ti,ab,kw OR MeSH descriptor: [Crohn Disease] explode all trees OR (Coronary Disease):ti,ab,kw OR MeSH descriptor: [Coronary Disease] explode all trees OR (Myocardial Ischemia):ti,ab,kw OR (chronic ischemic heart disease):ti,ab,kw OR (koronare Herzkrankheit):ti,ab,kw OR (Diabetes mellitus):ti,ab,kw OR MeSH descriptor: [Diabetes Mellitus] explode all trees OR ("vasculitis"):ti,ab,kw OR MeSH descriptor: [Vasculitis] explode all trees OR (Peripheral Arterial Disease):ti,ab,kw OR Arteriosklerose OR (arteriosclerosis):ti,ab,kw OR MeSH descriptor: [Arteriosclerosis] explode all trees OR (Cardiomyopath*):ti,ab,kw OR MeSH descriptor: [Cardiomyopathies] explode all trees OR Kardiomyopath* OR ("heart failure"):ti,ab,kw OR (Herzinsuffizienz):ti,ab,kw OR (heart insufficiency):ti,ab,kw OR (cardiac insufficiency):ti,ab,kw) AND ((Assess*):ti,ab,kw OR Insurance Claim Review OR (Insurance Claim Review):ti,ab,kw OR (Work Capacity Evaluation):ti,ab,kw OR MeSH descriptor: [Work Capacity Evaluation] explode all trees OR (claim assessment):ti,ab,kw OR (Disability Evaluation):ti,ab,kw OR MeSH descriptor: [Disability Evaluation] explode all trees OR (disability assessment):ti,ab,kw OR (employee performance appraisal):ti,ab,kw OR (functional capacity assessment):ti,ab,kw OR (functional evaluation):ti,ab,kw OR (functioning-based assessment):ti,ab,kw OR impairment assessment OR (impairment assessment):ti,ab,kw OR (impairment evaluation):ti,ab,kw OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR (insurance medicine):ti,ab,kw OR insurance medicine OR (Judgment):ti,ab,kw OR (medical examination):ti,ab,kw OR (needs assessment):ti,ab,kw OR MeSH descriptor: [Independent Medical Evaluation] explode all trees OR MeSH descriptor: [Judgment] explode all trees OR (Health Services Needs and Demand):ti,ab,kw OR MeSH descriptor: [Needs Assessment] explode all trees OR (physical	0

examination):ti,ab,kw OR MeSH descriptor: [Physical Examination] explode all trees OR (practice guideline):ti,ab,kw OR (socio-medical assessment):ti,ab,kw OR socio-medical evaluation OR (socio-medical evaluation):ti,ab,kw OR socio-medical assessment OR (symptom assessment):ti,ab,kw OR MeSH descriptor: [Symptom Assessment] explode all trees OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND ((compensation):ti,ab,kw OR MeSH descriptor: [Compensation and Redress] explode all trees OR MeSH descriptor: [Disabled Persons] explode all trees OR (disability benefit):ti,ab,kw OR (disability pension):ti,ab,kw OR disability pension OR (disability retirement):ti,ab,kw OR MeSH descriptor: [Retirement] explode all trees OR MeSH descriptor: [Pensions] explode all trees OR (functional capacity):ti,ab,kw OR ("functional capacity assessment"):ti,ab,kw OR (pension*):ti,ab,kw OR ("Social Security Disability Insurance"):ti,ab,kw OR Social Security OR (social security benefit):ti,ab,kw OR Erwerbsminderungsrente)) NOT ((rat*):ti,ab,kw OR rat* OR mouse OR mice OR (mouse):ti,ab,kw OR (mice):ti,ab,kw OR MeSH descriptor: [Rats] explode all trees OR MeSH descriptor: [Mice] explode all trees)) AND limit yr="1998 -Current"	
((osteoarthritis):ti,ab,kw OR MeSH descriptor: [Osteoarthritis] explode all trees OR (Arthrose):ti,ab,kw OR (arthrosis):ti,ab,kw OR MeSH descriptor: [Joint Diseases] explode all trees OR (Rueckenschmerzen):ti,ab,kw OR ("back pain"):ti,ab,kw OR MeSH descriptor: [Back Pain] explode all trees OR MeSH descriptor: [Low Back Pain] explode all trees OR Spondylopath* OR (Spondylopath*):ti,ab,kw OR MeSH descriptor: [Spondylarthropathies] explode all trees OR (Spinal Osteophytosis):ti,ab,kw OR (Cervical Vertebrae):ti,ab,kw OR (Spinal Diseases):ti,ab,kw OR ("sciatica"):ti,ab,kw OR MeSH descriptor: [Sciatica] explode all trees OR Lumboischialgie OR ("radiculopathy"):ti,ab,kw OR MeSH descriptor: [Radiculopathy] explode all trees) AND ((Assess*):ti,ab,kw OR Insurance Claim Review OR (Insurance Claim Review):ti,ab,kw OR (Work Capacity Evaluation):ti,ab,kw OR MeSH descriptor: [Work Capacity Evaluation] explode all trees OR (claim assessment):ti,ab,kw OR (Disability Evaluation):ti,ab,kw OR MeSH descriptor: [Disability Evaluation] explode all trees OR (disability assessment):ti,ab,kw OR (employee performance appraisal):ti,ab,kw OR (functional capacity assessment):ti,ab,kw OR (functional evaluation):ti,ab,kw OR (functioning-based assessment):ti,ab,kw OR impairment assessment OR (impairment assessment):ti,ab,kw OR (impairment evaluation):ti,ab,kw OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR (insurance medicine):ti,ab,kw OR insurance medicine OR (Judgment):ti,ab,kw OR (medical examination):ti,ab,kw OR (needs assessment):ti,ab,kw OR MeSH descriptor: [Independent Medical Evaluation] explode all trees OR MeSH descriptor: [Judgment] explode all trees OR (Health Services Needs and Demand):ti,ab,kw OR MeSH descriptor: [Needs Assessment] explode all trees OR (physical examination):ti,ab,kw OR MeSH descriptor: [Physical Examination] explode all trees OR (practice guideline):ti,ab,kw OR (socio-medical assessment):ti,ab,kw OR socio-medical evaluation OR (socio-medical evaluation):ti,ab,kw OR socio-medical assessment OR (symptom assessment):ti,ab,kw OR MeSH descriptor: [Symptom Assessment] explode all trees OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND ((compensation):ti,ab,kw OR MeSH descriptor: [Compensation and Redress] explode all trees OR MeSH descriptor: [Disabled Persons] explode all trees OR (disability benefit):ti,ab,kw OR (disability pension):ti,ab,kw OR disability pension OR (disability retirement):ti,ab,kw OR MeSH descriptor: [Retirement] explode all trees OR MeSH descriptor: [Pensions] explode all trees OR (functional capacity):ti,ab,kw OR ("functional capacity assessment"):ti,ab,kw OR (pension*):ti,ab,kw OR ("Social Security Disability Insurance"):ti,ab,kw OR Social Security OR (social security benefit):ti,ab,kw OR Erwerbsminderungsrente)) NOT ((rat*):ti,ab,kw OR rat* OR mouse OR mice OR (mouse):ti,ab,kw OR (mice):ti,ab,kw OR MeSH descriptor: [Rats] explode all trees OR MeSH descriptor: [Mice] explode all trees)) AND limit yr="1998 -Current"	1
(psychological disorder):ti,ab,kw OR (psychische Störung):ti,ab,kw OR (affektive Störung):ti,ab,kw OR (affective disorder):ti,ab,kw OR MeSH descriptor: [Mood Disorders] explode all trees OR (psychiatric disorder):ti,ab,kw OR MeSH descriptor: [Mental Disorders] explode all trees OR ("mood disorder"):ti,ab,kw OR depression OR MeSH descriptor: [Depression] explode all trees OR ("substance related disorders"):ti,ab,kw OR (Opioid-Related Disorders):ti,ab,kw OR (alkohol abus*):ti,ab,kw OR alkohol abus* OR ("alcoholism"):ti,ab,kw OR somatoforme Störung OR Somatoform Disorders OR MeSH descriptor: [Somatoform Disorders] explode all trees OR (somatoform disorder):ti,ab,kw OR (Psychotic Disorders):ti,ab,kw OR MeSH descriptor: [Psychotic Disorders] explode all trees OR ("schizophrenia"):ti,ab,kw OR MeSH descriptor: [Schizophrenia] explode all trees OR (Schizophrenie):ti,ab,kw OR (Angststörung):ti,ab,kw OR (Anxiety Disorder):ti,ab,kw OR MeSH descriptor: [Anxiety Disorders] explode all trees	1

OR ("anxiety"):ti,ab,kw OR (Phobic Disorders):ti,ab,kw OR MeSH descriptor: [Phobic Disorders] explode all trees OR ("multiple sclerosis"):ti,ab,kw OR MeSH descriptor: [Multiple Sclerosis] explode all trees) AND ((Assess*):ti,ab,kw OR Insurance Claim Review OR (Insurance Claim Review):ti,ab,kw OR (Work Capacity Evaluation):ti,ab,kw OR MeSH descriptor: [Work Capacity Evaluation] explode all trees OR (claim assessment):ti,ab,kw OR (Disability Evaluation):ti,ab,kw OR MeSH descriptor: [Disability Evaluation] explode all trees OR (disability assessment):ti,ab,kw OR (employee performance appraisal):ti,ab,kw OR (functional capacity assessment):ti,ab,kw OR (functional evaluation):ti,ab,kw OR (functioning-based assessment):ti,ab,kw OR impairment assessment OR (impairment assessment):ti,ab,kw OR (impairment evaluation):ti,ab,kw OR impairment of earning capacity assessment OR impairment of earning capacity evaluation OR (insurance medicine):ti,ab,kw OR insurance medicine OR (Judgment):ti,ab,kw OR (medical examination):ti,ab,kw OR (needs assessment):ti,ab,kw OR MeSH descriptor: [Independent Medical Evaluation] explode all trees OR MeSH descriptor: [Judgment] explode all trees OR (Health Services Needs and Demand):ti,ab,kw OR MeSH descriptor: [Needs Assessment] explode all trees OR (physical examination):ti,ab,kw OR MeSH descriptor: [Physical Examination] explode all trees OR (practice guideline):ti,ab,kw OR (socio-medical assessment):ti,ab,kw OR socio-medical evaluation OR (socio-medical evaluation):ti,ab,kw OR socio-medical assessment OR (symptom assessment):ti,ab,kw OR MeSH descriptor: [Symptom Assessment] explode all trees OR Sozialmedizinische Begutachtung OR Begutachtung OR Leistungsbeurteilung) AND ((compensation):ti,ab,kw OR MeSH descriptor: [Compensation and Redress] explode all trees OR MeSH descriptor: [Disabled Persons] explode all trees OR (disability benefit):ti,ab,kw OR (disability pension):ti,ab,kw OR disability pension OR (disability retirement):ti,ab,kw OR MeSH descriptor: [Retirement] explode all trees OR MeSH descriptor: [Pensions] explode all trees OR (functional capacity):ti,ab,kw OR ("functional capacity assessment"):ti,ab,kw OR (pension*):ti,ab,kw OR ("Social Security Disability Insurance"):ti,ab,kw OR Social Security OR (social security benefit):ti,ab,kw OR Erwerbsminderungsrente)) NOT ((rat*):ti,ab,kw OR rat* OR mouse OR mice OR (mouse):ti,ab,kw OR (mice):ti,ab,kw OR MeSH descriptor: [Rats] explode all trees OR MeSH descriptor: [Mice] explode all trees)) AND limit yr="1998-Current"	

Anhang 3:

PRISMA-Statement: Qualitätskriterien für systematische Reviews

Publikations- abschnitt	Zusammenfassung	Qualitäts- kriterium, aufgeführt auf Seite
Titel		
1. Titel	Identifikation als Bericht einer systematischen Übersicht, Meta-Analyse oder von beidem	Ja - S. 69
Zusammenfassung		
2. Strukturierte Zusammenfassung	Strukturierte Zusammenfassung mit den Stichworten (sofern geeignet): Hintergrund; Ziele; Datenquellen; Auswahlkriterien der Studien, Teilnehmer und Interventionen; Bewertung der Studie und Methoden der Synthese; Ergebnisse; Einschränkungen; Schlussfolgerungen und Implikation der wichtigsten Ergebnisse; Registrierungsnummer der systematischen Übersicht	Ja - S. 1-2
Einleitung		
3. Hintergrund und Rationale	Wissenschaftlicher Hintergrund und Begründung der Studie	Ja - S. 72
4. Ziele	Präzise Angabe der Fragestellungen mit Bezug auf Teilnehmer, Interventionen, Vergleiche, Zielkriterien und Studiendesign (engl.: participants, interventions, comparisons, outcomes, and study design; PICOS)	Nein
Methoden		
5. Protokoll und Registrierung	Existiert ein Studienprotokoll für die Übersichtsarbeit? Wenn ja, wo kann es gefunden bzw. wie kann es bezogen werden (z.B. Webseite); wenn verfügbar: Informationen zur Registrierung einschließlich Angabe der Registrierungsnummer	Nein
6. Auswahlkriterien	Merkmale der Studien (z.B., PICOS, Dauer der Nachbeobachtung) und der Berichte (z.B. Zeitraum der Studien, Sprache, Publikationsstatus), die als Auswahlkriterien verwendet wurden, mit Begründung	Ja - S. 74-78
7. Informationsquellen	Beschreibung aller Informationsquellen (z.B., Datenbanken mit Zeitpunkten der Berichterstattung, Kontakt mit Autoren von Studien, um zusätzliche Studien zu identifizieren), die bei der Suche verwendet wurden, einschließlich des letzten Suchdatums	Ja - S. 73 / S. 81
8. Suche	Beschreibung der vollständigen elektronischen Suchstrategie für mindestens eine Datenbank, einschließlich gewählter Limitierungen, so dass die Suche repliziert werden könnte	Ja - S. 252
9. Auswahl der Studien	Beschreibung des Auswahlprozesses von Studien (das heißt Vorauswahl, Eignung, Einschluss in die systematische Übersicht und, falls zutreffend, in die Meta-Analyse)	Ja - S. 77
10. Prozess der Datengewinnung	Beschreibung der Methode der Datenextraktion aus Berichten (z.B. Erhebungsbogen, unabhängig, doppelt) und alle Prozesse, um Daten von Untersuchern zu erhalten und zu bestätigen	Ja - S. 78

Publikations- abschnitt	Zusammenfassung	Qualitäts- kriterium, aufgeführt auf Seite
11. Datendetails	Aufzählung und Definition aller Variablen, nach denen gesucht wurde (z.B., PICOS, Finanzierungsquellen) sowie Annahmen und durchgeführte Vereinfachungen	Ja - S. 90 / S. 97 / S. 100
12. Risiko der Verzerrung in den einzelnen Studien	Methoden zur Beurteilung des Risikos von Verzerrungen der einzelnen Studien (einschließlich der Angabe, ob dieses auf der Studienebene oder für das Zielkriterium durchgeführt wurde) und wie diese Information bei der Datensynthese berücksichtigt wurde	Nein
13. Effektschätzer	Wichtigste Effektschätzer (z.B. relatives Risiko, Mittelwertsdifferenz)	Nein
14. Synthese der Ergebnisse	Beschreibung der Methoden zum Umgang mit den Daten und der Kombination der Ergebnisse der Studien; falls diese berechnet wurden, einschließlich Maßzahlen zur Homogenität der Ergebnisse (z.B. I ²) für jede Meta-Analyse	Nein
15. Risiko der Verzerrung über Studien hinweg	Beschreibung der Beurteilung des Risikos von Verzerrungen, die die kumulative Evidenz beeinflussen könnten (z.B. Publikationsverzerrung, selektives Berichten innerhalb von Studien)	Nein
16. Zusätzliche Analyse	Methoden für zusätzliche Analysen (z.B. Sensitivitätsanalysen, Subgruppenanalysen, Meta-Regression) mit Beschreibung, welche vorab spezifiziert waren	entfällt
Ergebnisse		
17. Auswahl der Studien	Anzahl der Studien, die in die Vorauswahl aufgenommen, auf Eignung geprüft und in die Übersicht eingeschlossen wurden, mit Begründung für Ausschluss in jeder Stufe, idealerweise unter Verwendung eines Fluss-diagramms	Ja S. 81 / S. 265
18. Studienmerkmale	Für jede Studie Darstellung der Merkmale, nach denen Daten extrahiert wurden (z.B. Fallzahl, PICOS, Nachbeobachtungszeitraum), Literaturstelle der Studie	Ja - S. 90 / S. 97 / S. 100
19. Risiko der Verzerrung innerhalb der Studien	Daten zum Risiko von Verzerrungen innerhalb jeder Studie und, falls verfügbar, eine Beurteilung der Güte der Zielkriterien (siehe Item 12)	Nein
20. Ergebnisse der einzelnen Studien	Für jede Studie Darstellung aller Endpunkte (Wirksamkeit und Nebenwirkungen): (a) einfache zusammenfassende Daten für jede Interventionsgruppe, (b) Effektschätzer und Konfidenzintervalle, idealerweise mit Forest Plot	Ja - S. 90 / S. 97 / S. 100
21. Ergebnissynthese	Darstellung der Meta-Analyse, einschließlich Konfidenzintervalle und Heterogenitätsmaße	entfällt
22. Risiko von Verzerrungen über Studien hinweg	Darstellung der Ergebnisse zur Beurteilung des Risikos von Verzerrungen über alle Studien hinweg (siehe Item 15)	entfällt
23. Zusätzliche Analyse	Präsentation der Ergebnisse der zusätzlichen Analysen, falls durchgeführt (z.B. Sensitivitäts- oder Subgruppenanalysen, Meta-Regression [siehe Item 16])	Nein

Publikations- abschnitt	Zusammenfassung	Qualitäts- kriterium, aufgeführt auf Seite
Diskussion		
24. Zusammen- fassung der Evidenz	Zusammenfassung der Hauptergebnisse einschließlich der Stärke der Evidenz für jedes Hauptzielkriterium; Relevanz für Zielgruppen (z.B. Gesundheitsdienstleister, Anwender, politische Entscheidungsträger)	Ja - S. 102-105
25. Einschränkungen	Einschränkungen der Studie auf Studienebene und auf Ebene der Zielkriterien diskutiert (z.B. Risiko von Verzerrungen) sowie auf Ebene der Übersicht (z.B. unvollständiges Auffinden der identifizierten Forschung, Verzerrung des Berichts)	Nein
26. Schlussfolgerung	Interpretation der Ergebnisse unter Berücksichtigung des Stands der Forschung und Schlussfolgerungen für weitere Forschung	Ja - S. 106-111
Finanzielle Unterstützung		
27. Finanzielle Unterstützung	Quellen der finanziellen Unterstützung sowie andere Unterstützung (z.B. zur Verfügung stellen von Daten); Funktion der Geldgeber für die systematische Übersicht	nein

Anhang 4:

Anhand der Volltexte von der Literaturrecherche ausgeschlossene Studien (n = 88)

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Abraham, B. P. (2012)	Disability in inflammatory bowel disease. <i>Gastroenterology clinics of North America</i> , 41 (2), 429–441.	Kein Bezug zur Erwerbsminderung
Ahola, K. (2011)	Common mental disorders and subsequent work disability: A population-based Health 2000 Study. <i>JOURNAL OF AFFECTIVE DISORDERS</i> , 134 (1–3), 365–372.	Prädiktorstudie
Al-Otaibi, S. T. (2004)	Evaluation of disability of occupational illness claims for workers' compensation. <i>Saudi medical journal</i> , 25 (2), 145–149.	Studie über Berufskrankheiten
Ananthakrishnan, A. N. (2008)	Permanent work disability in Crohn's disease. <i>The American journal of gastroenterology</i> , 103 (1), 154–161.	Prädiktorstudie
Apfel, T. (2005)	Werden psychisch Kranke zu schnell in die Rente „abgeschoben“? Querschnittserhebung anhand von 101 psychiatrischen Gutachten für die Invalidenversicherung aus dem Jahr 2002 an der Psychiatrischen Poliklinik Basel. <i>PSYCHIATRISCHE PRAXIS</i> , 32 (4), 172–176.	Kein Bezug zur Begutachtungspraxis
Bartys, S. (2017)	System influences on work disability due to low back pain: An international evidence synthesis. <i>Health policy</i> , 121 (8), 903–912.	Kein Bezug zur Begutachtungspraxis
Becattini-Oliveira, A. C. (2018)	A systematic review of a functional assessment Tool: UCSD Performance-based skill assessment (UPSA). <i>Psychiatry research</i> , 267, 12–18.	Assessment-instrument
Bethge, M. (2013)	Work Ability Index predicts application for disability pension after work-related medical rehabilitation for chronic back pain. <i>Archives of physical medicine and rehabilitation</i> , 94 (11), 2262–2268.	Prädiktorstudie
Bjorkenstam, C. (2014)	Diagnosis-specific disability pension and risk of all-cause and cause-specific mortality--a cohort study of 4.9 million inhabitants in Sweden. <i>BMC public health</i> , 14, 1247.	Korrelationsstudie
Borg, K. (2004)	Risk factors for disability pension over 11 years in a cohort of young persons initially sick-listed with low back, neck, or shoulder diagnoses. <i>Scandinavian journal of public health</i> , 32 (4), 272–278.	Prädiktorstudie
Bottcher, H. M. (2013)	Occupational stress and its association with early retirement and subjective need for occupational rehabilitation in cancer patients. <i>Psycho-oncology</i> , 22 (8), 1807–1814.	Prädiktorstudie

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Brieger, P. (2004)	Disability payments due to unipolar depressive and bipolar affective disorders. <i>PSYCHIATRISCHE PRAXIS</i> , 31 (4), 203–206.	Prädiktorstudie
Brown, C. A. (2012)	Perceived injustice in injured workers: analysis of public responses to an injured worker who took Workers' Compensation Board employees hostage. <i>SCANDINAVIAN JOURNAL OF CARING SCIENCES</i> , 26 (3), 569–578.	Kein Bezug zur Begutachtungspraxis
Busch, K. (2014)	Sick leave and disability pension in inflammatory bowel disease: a systematic review. <i>Journal of Crohn's & colitis</i> , 8 (11), 1362–1377.	Prädiktorstudie
Carey, T. S. (1999)	Disability: how successful are we in determining disability?. <i>Neurologic clinics</i> , 17 (1), 167–178.	Editorial
Chariot, P. (2013)	Evaluating the total incapacity to work: Implementing French National Authority for Health guidelines in clinical practice. <i>PRESSE MEDICALE</i> , 42 (10), 1300–1309.	Französische Sprache
Christopher, P. P. (2011)	Evaluating Psychiatric Disability: Differences by Forensic Expertise. <i>J Am Acad Psychiatry Law</i> , 39, 183-188.	Kein Bezug zur Begutachtungspraxis
Colon-Emeric, C. S. (2013)	A team approach: comprehensive evaluation for functional decline in older patients. <i>American family physician</i> , 88 (6), 397.	Editorial
Davies, P. (2000)	The effect of welfare reform on SSA's disability programs: Design of policy evaluation and early evidence. <i>SOCIAL SECURITY BULLETIN</i> , 63 (1), 3–11.	Kein Bezug zur Begutachtungspraxis
Detaille, S. I. (2009)	Common prognostic factors of work disability among employees with a chronic somatic disease: a systematic review of cohort studies. <i>Scandinavian journal of work, environment & health</i> , 35 (4), 261–281.	Prädiktorstudie
Dewa, C. S. (2016).	Employment Benefit Receipt Among Ontario Public Disability Benefit Recipients with a Disability Related to a Mental Disorder. <i>Community mental health journal</i> , 52 (2), 203–211.	Prädiktorstudie
Elbers, N. A. (2013)	Do compensation processes impair mental health? A meta-analysis. <i>INJURY-INTERNATIONAL JOURNAL OF THE CARE OF THE INJURED</i> , 44 (5), 674–683.	Studie über Unfallopfer
Eley, C. (2006)	Magnetic resonance imaging for low back injuries: appropriate use in managing workers' compensation claims. <i>AAOHN journal : official journal of the American Association of Occupational Health Nurses</i> , 54 (10), 429–433.	Kein Bezug zur Begutachtungspraxis

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Escorpizo, R. (2014)	Defining the principles of musculoskeletal disability and rehabilitation. <i>Best practice & research. Clinical rheumatology</i> , 28 (3), 367–375.	Kein Bezug zur Erwerbsminderung
Figueredo, R. (2015)	Psychiatric legal investigation for sickness benefits due to disability at the Brazilian Federal Social Security Special Court in Florianopolis, capital city of the State of Santa Catarina, southern Brazil. <i>Trends in psychiatry and psychotherapy</i> , 37 (2), 82–86.	Kein Bezug zur Begutachtungspraxis
Fink, G.	Functional evaluation in patients with chronic obstructive pulmonary disease: Pulmonary function test versus cardiopulmonary exercise test. <i>JOURNAL OF OCCUPATIONAL AND ENVIRONMENTAL MEDICINE</i> , 44 (1), 54–58.	Kein Bezug zur Erwerbsminderung
Fujii, T. (2012)	The association between compensation and chronic disabling back pain. <i>Journal of orthopaedic science : official journal of the Japanese Orthopaedic Association</i> , 17 (6), 694–698.	Prädiktorstudie
Geisser, M. E. (2003)	Psychosocial factors and functional capacity evaluation among persons with chronic pain. <i>JOURNAL OF OCCUPATIONAL REHABILITATION</i> , 13 (4), 259–276.	Prädiktorstudie
Geyh, S. (2011)	The Personal Factors of the International Classification of Functioning, Disability and Health in the literature - a systematic review and content analysis. <i>DISABILITY AND REHABILITATION</i> , 33 (13–14), 1089–1102.	Kein Bezug zur Begutachtungspraxis
Gross, D. P. (2005)	Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. <i>Physical therapy</i> , 85 (4), 315–322.	Prädiktorstudie
Haig, A. J. (2003)	The effect of order of testing in functional performance in persons with and without chronic back pain. <i>JOURNAL OF OCCUPATIONAL REHABILITATION</i> , 13 (2), 115–123.	Kein Bezug zur Erwerbsminderung
Heiland, F. (2015)	Anchoring Vignettes in the Health and Retirement Study: How Do Medical Professionals and Disability Recipients Characterize the Severity of Work Limitations? <i>PLOS ONE</i> , 10 (5), e0126218. doi:10.1371/journal.	Kein Bezug zur Begutachtungspraxis
Hendriks, S. M. (2016)	Long-term disability in anxiety disorders. <i>BMC psychiatry</i> , 16, 248.	Kein Bezug zur Begutachtungspraxis
Holwerda, A. (2013)	Work-ability assessment in young adults with disabilities applying for disability benefits. <i>DISABILITY AND REHABILITATION</i> , 35 (6), 498–505.	Prädiktorstudie
Hubertsson, J. (2013)	Risk of sick leave and disability pension in working-age women and men with knee osteoarthritis. <i>Annals of the rheumatic diseases</i> , 72 (3), 401–405.	Prädiktorstudie

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Kang, D.-M. (2014)	Compensation for occupational neurological and mental disorders. <i>Journal of Korean medical science</i> , 29 Suppl, 59-65.	Studie über Berufskrankheiten
Khan, F. (2007)	Use of international classification of functioning, disability and health (ICF) to describe patient-reported disability in multiple sclerosis and identification of relevant environmental factors. <i>JOURNAL OF REHABILITATION MEDICINE</i> , 39 (1), 63–70.	Kein Bezug zur Begutachtungspraxis
Kilgour, E. (2015)	Healing or Harming? Healthcare Provider Interactions with Injured Workers and Insurers in Workers' Compensation Systems. <i>JOURNAL OF OCCUPATIONAL REHABILITATION</i> , 25 (1), 220–239.	Kein Bezug zur Begutachtungspraxis
King, P. M. (1998)	A Critical Review of Functional Capacity Evaluation. <i>Physical Therapy</i> , 78 (8), 852-866.	Kein Bezug zur Begutachtungspraxis
Krause, N. (1999)	Alternative approaches for measuring duration of work disability after low back injury based on administrative workers' compensation data. <i>American journal of industrial medicine</i> , 35 (6), 604–618.	Kein Bezug zur Begutachtungspraxis
Krstovic-Spremo, V. (2014)	The effects of diabetes mellitus and hypertension on work productivity. <i>Acta medica academica</i> , 43 (2), 122–133.	Kein Bezug zur Begutachtungspraxis
Laisne, F. (2013)	Biopsychosocial determinants of work outcomes of workers with occupational injuries receiving compensation: A prospective study. <i>WORK-A JOURNAL OF PREVENTION ASSESSMENT & REHABILITATION</i> , 44 (2), 117–132.	Kein Bezug zur Erwerbsminderung
Langeskov-Christensen, D. (2017)	Performed and perceived walking ability in relation to the Expanded Disability Status Scale in persons with multiple sclerosis. <i>JOURNAL OF THE NEUROLOGICAL SCIENCES</i> , 382, 131–136.	Korrelationsstudie
Lederer, V. (2014)	Compensation benefits in a population-based cohort of men and women on long-term disability after musculoskeletal injuries: costs, course, predictors. <i>OCCUPATIONAL AND ENVIRONMENTAL MEDICINE</i> , 71 (11), 772–779.	Prädiktorstudie
Lippel, K. (2012)	Preserving workers' dignity in workers' compensation systems: An international perspective. <i>AMERICAN JOURNAL OF INDUSTRIAL MEDICINE</i> , 55 (6, SI), 519–536.	Kein Bezug zur Begutachtungspraxis
Lippel, K. (2010)	Access to workers' compensation benefits and other legal protections for work-related mental health problems: a Canadian overview. <i>Canadian journal of public health</i> , 101 Suppl, 16-22.	Kein Bezug zur Begutachtungspraxis
Louwerse, I. (2018)	Characteristics of individuals receiving disability benefits in the Netherlands and predictors of leaving the disability benefit scheme: a retrospective cohort study with five-year follow-up. <i>BMC public health</i> , 18 (1), 157.	Prädiktorstudie

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Luckhaupt, S. E. (2010)	Work-relatedness of selected chronic medical conditions and workers' compensation utilization: National health interview survey occupational health supplement data. <i>American journal of industrial medicine</i> , 53 (12), 1252–1263.	Kein Bezug zur Begutachtungspraxis
MacLeod, F. K. (2001)	The effect of disability claimants' coping styles on judgments of pain, disability, and compensation: A vignette study. <i>REHABILITATION PSYCHOLOGY</i> , 46 (4), 417–435.	Prädiktorstudie
Marfeo, E. E. (2013)	Conceptual Foundation for Measures of Physical Function and Behavioral Health Function for Social Security Work Disability Evaluation. <i>ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION</i> , 94 (9), 1645–1652.	Kein Bezug zur Begutachtungspraxis
Martins, E. F. (2011)	A Brazilian experience to describe functioning and disability profiles provided by combined use of ICD and ICF in chronic stroke patients at home-care. <i>Disability and rehabilitation</i> , 33 (21–22), 2064–2074.	Kein Bezug zur Erwerbsminderung
Mayer, C. U. (2003)	Schmerz und Arbeitsunfähigkeit. <i>Z Orthop</i> 141 (6), 606–615.	not available
McCauley, E. (2017)	Navigating the Disability Determination Process from the Perspective of Incarcerated Adults with Serious Mental Illnesses. <i>Community mental health journal</i> , 53 (8), 905–915.	Kein Bezug zur Erwerbsminderung
McGurk, S. R. (2018)	Cognitive Predictors of Work Among Social Security Disability Insurance Beneficiaries With Psychiatric Disorders Enrolled in IPS Supported Employment. <i>Schizophrenia bulletin</i> , 44 (1 CC-Schizophrenia), 32-37.	Prädiktorstudie
McIntosh, G. (2000)	Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. <i>Spine</i> , 25 (2), 147–157.	Prädiktorstudie
Melloh, M. (2015)	Does injury claim status and benefit status predict low back pain outcomes?. <i>The Australasian medical journal</i> , 8 (8), 268–276.	Prädiktorstudie
Michel, O. (2008)	Bewertung von Tinnitus als Gesundheitsschaden in der privaten Unfallversicherung - eine Novität. <i>LARYNGO-RHINO-OTOLOGIE</i> , 87 (7), 477–481.	Kein Bezug zur Erwerbsminderung
Moore, R. (2013)	Limitations in the Assessment of Functional Capacity in Cognitively Unimpaired Outpatients Diagnosed with Schizophrenia and Bipolar Disorder. <i>ARCHIVES OF CLINICAL NEUROPSYCHOLOGY</i> , 28 (6), 563.	Posterabstract - Korrelationsstudie
Muschalla, B. (2011)	Socio-medical aspects of mental disorders. <i>NERVENARZT</i> , 82 (7), 917–930.	Kein Bezug zur Begutachtungspraxis

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Myhr, K. M. (2001)	Disability and prognosis in multiple sclerosis: demographic and clinical variables important for the ability to walk and awarding of disability pension. <i>MULTIPLE SCLEROSIS</i> , 7 (1), 59–65.	Prädiktorstudie
Narusyte, J. (2011)	Genetic Liability to Disability Pension in Women and Men: A Prospective Population-Based Twin Study. <i>PLOS ONE</i> , 6 (8).	Kein Bezug zur Begutachtungspraxis
Nedopil, N. (2009)	Psychiatrische Begutachtung bei zivilrechtlichen Fragestellungen. <i>NERVENARZT</i> , 80 (5), 611–619.	Kein Bezug zur Erwerbsminderung
Nieuwenhuijsen, K. (2014)	Predicting Improvement of Functioning in Disability Claimants. <i>JOURNAL OF OCCUPATIONAL REHABILITATION</i> , 24 (3), 410–418.	Prädiktorstudie
Nokes, S. E. (2012)	Medical-legal issues regarding patients with low back pain. <i>Primary care</i> , 39 (3), 573–585.	not available
Park, S. (2014)	Workers' compensation for occupational respiratory diseases. <i>Journal of Korean medical science</i> , 29 Suppl, S47-51.	Kein Bezug zur Begutachtungspraxis
Pransky, G. S. (2004)	Practical Aspects of Functional Capacity Evaluations. <i>Journal of Occupational Rehabilitation</i> , 14 (3), 217-229.	Kein Bezug zur Begutachtungspraxis
Purser, K. (2015)	A therapeutic approach to assessing legal capacity in Australia. <i>International journal of law and psychiatry</i> , 38, 18–28.	Kein Bezug zur Begutachtungspraxis
Quintero, S. (2012)	The Disability Evaluation and Low Back Pain. <i>PRIMARY CARE</i> , 39 (3), 553-559.	not available
Rezvy, G. (2007)	Between health care and social security - psychiatric patients and the disability pension system in Norway and Russia. <i>BMC HEALTH SERVICES RESEARCH</i> , 7, 128.	Kein Bezug zur Begutachtungspraxis
Riberto, M. (2014)	A transversal multicenter study assessing functioning, disability and environmental factors with the comprehensive ICF core set for low back pain in Brazil. <i>European journal of physical and rehabilitation medicine</i> , 50 (2), 153–160.	Kein Bezug zur Erwerbsminderung
Rosenheck, R. A. (2017)	Incomes and Outcomes: social Security Disability Benefits in First-Episode Psychosis. <i>American journal of psychiatry</i> , 174 (9), 886-894.	Prädiktorstudie
Scheman, J. (2000)	Effects of instructions on physical capacities outcome in a workers' compensation setting. <i>PAIN MEDICINE</i> , 1 (2), 116–122.	Kein Bezug zur Begutachtungspraxis
Schönstein, E. (2000)	Diagnoses and treatment recommendations on workers compensation medical certificates. <i>The Medical journal of Australia</i> , 173 (8), 419–422.	Kein Bezug zur Begutachtungspraxis

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Schulz, S. C. (2016)	Assessing Cognitive Impairment in Patients With Schizophrenia. <i>JOURNAL OF CLINICAL PSYCHIATRY</i> , 77 (2), 3–7.	Kein Bezug zur Erwerbsminderung
Shiri, R. (2018)	A screening tool for the risk of disability retirement due to musculoskeletal disorders. <i>SCANDINAVIAN JOURNAL OF WORK ENVIRONMENT & HEALTH</i> , 44 (1), 37–46.	Assessment-instrument
Slebus, F. G. (2007)	Prognostic factors for work ability in sicklisted employees with chronic diseases. <i>Occupational and environmental medicine</i> , 64 (12), 814–819.	Prädiktorstudie
Tacci, J. A. (1999)	Clinical practices in the management of new-onset, uncomplicated, low back workers' compensation disability claims. <i>Journal of occupational and environmental medicine</i> , 41 (5), 397–404.	Kein Bezug zur Begutachtungspraxis
Tait, R. C. (2016)	Workers' Compensation Claimants with Low Back Pain: the Role of Dissatisfaction in the Transition to Disability. <i>PSYCHOLOGICAL INJURY & LAW</i> , 9 (1, SI), 16–22.	Prädiktorstudie
Topa, G. (2018)	Early Retirement: A Meta-Analysis of Its Antecedent and Subsequent Correlates. <i>FRONTIERS IN PSYCHOLOGY</i> , 8, 2157.	Korrelationsstudie
Torrey, D. B. (2018)	Recent Developments in Workers' Compensation and Employers' Liability Law. <i>Tort trial & insurance practice law journal</i> , 53 (2), 703–739.	Kein Bezug zur Begutachtungspraxis
Tousignant, M. (2000)	Occupational disability related to back pain: application of a theoretical model of work disability using prospective cohorts of manual workers. <i>American journal of industrial medicine</i> , 37 (4), 410–422.	Kein Bezug zur Erwerbsminderung
Valentin, G. H. (2016)	Prognostic factors for disability and sick leave in patients with subacute non-malignant pain: a systematic review of cohort studies. <i>BMJ open</i> , 6 (1), e007616.	Untersuchung von Korrelationsstudien
van der Bij, S. (2013)	Legal claims for malignant mesothelioma: dealing with all cases. <i>Lung cancer</i> , 80 (2), 153–158.	Korrelationsstudie
Van Eerd, D. (2006)	Capturing cases in workers' compensation databases: The example of neck pain. <i>AMERICAN JOURNAL OF INDUSTRIAL MEDICINE</i> , 49 (7), 557–568.	Kein Bezug zur Begutachtungspraxis
Vieira, E. R. (2011)	Work disability benefits due to musculoskeletal disorders among Brazilian private sector workers. <i>BMJ OPEN</i> , 1 (1), https://doi.org/10.1136/bmjopen-2011-000003	Kein Bezug zur Begutachtungspraxis
Warren, P. A. (2016)	The Usage of the AMA Guides for the Determination of Psychological Injury Within the State and Federal Workers' Compensation System. <i>Psychol Inj and Law</i> , 9, 313-340.	Kein Bezug zur Begutachtungspraxis

Erstautor (Jahr)	Quellenangabe	Grund für Ausschluss
Wind, H. (2009)	Complementary value of functional capacity evaluation for physicians in assessing the physical work ability of workers with musculoskeletal disorders. <i>INTERNATIONAL ARCHIVES OF OCCUPATIONAL AND ENVIRONMENTAL HEALTH</i> , 82 (4), 435–443.	Ergo-Kit Assessment-instrument
Wise, E. A. (2016)	Psychological Injuries, Workers' Compensation Insurance, and Mental Health Policy Issues. <i>PSYCHOLOGICAL INJURY & LAW</i> , 9 (4), 283–297.	Kein Bezug zur Begutachtungspraxis

Anhang 5:

Manual zum Peer Review-Verfahren: Arbeitsversion 2010

Qualitätssicherung der sozialmedizinischen Begutachtung

Manual

Arbeitsversion vom

06.12.2010

Deutsche Rentenversicherung

Wissenschaftlicher Beratung durch den Arbeitsbereich Rehabilitations-
wissenschaften der Universität Würzburg
(Gerlich, Löffler, Meng, Vogel)

A. Einzelkriterium „Formale Gestaltung“

Das Einzelkriterium „*Formale Gestaltung*“ untergliedert sich in zwei Prüffragen. Das Kriterium erfordert die adäquate Berücksichtigung des Anforderungsprofils für das sozialmedizinische Gutachten. So sind die formalen Anforderungen an das Gutachten gemäß den Gliederungsvorgaben der DRV-Schrift Band 21 einzuhalten. Des Weiteren ist der einheitliche Formularsatz der Deutschen Rentenversicherung, bestehend aus Deckblatt und Schlussblatt Teil 1 und Teil 2, zu verwenden.

Die Prüfung der Qualität bezieht sich in den folgenden zwei Prüffragen ausschließlich auf rein formale Aspekte. Inhaltliche Vollständigkeit und Nachvollziehbarkeit der Darstellung sind hier nicht qualitätsrelevant; sie werden an anderer Stelle geprüft.

Einzelkriterium A: Formale Gestaltung

Prüffrage A1

A1: Inwieweit werden die Gliederungsvorgaben gemäß den Anforderungen eingehalten?

Folgende Überschriften, die derzeit im DRV Bd. 21 (auf Seite 19) dargelegt sind, müssen vorhanden sein:

- Anamnese
- Untersuchungsbefunde
- Diagnosen
- Epikrise
- Sozialmedizinische Leistungsbeurteilung

Die Überschriften (bezogen auf das Layout) sind im freien Teil des Gutachtens aufzuführen und klar voneinander abzugrenzen.

In Abgrenzung zu den Prüffragen des Einzelkriteriums "Vollständigkeit" geht es hier um das rein formale Vorhandensein der Gliederungspunkte. Inhaltliche Aspekte, wie z. B. Art, Umfang, Angemessenheit oder Vollständigkeit der Gliederungspunkte sind hier nicht Gegenstand der Bewertung.

Leichte Mängel

Die Reihenfolge der Gliederungspunkte wird nicht eingehalten.

Deutliche Mängel

Ein Gliederungspunkt fehlt.

Gravierende Mängel

Mehr als ein Gliederungspunkt fehlt.

Welche Mängel wurden festgestellt?

Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren.

Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.

Einzelkriterium A: Formale Gestaltung Prüffrage A2	
A2: Inwieweit wird der einheitliche Formularsatz bestehend aus Deck- und Schlussblatt verwendet? Bei der Beurteilung der Prüffragen dieses Einzelkriteriums ist zu berücksichtigen, inwieweit die Vorgaben der DRV-Schrift 21 eingehalten werden. An dieser Stelle wird geprüft, ob das einheitliche Deck- und Schlussblatt verwendet werden. Es wird nicht geprüft, ob die Aussagen des Deck- und Schlussblattes an anderer Stelle im Gutachten vorhanden ist bzw. ob das Gutachten bei fehlendem Deck- oder Schlussblatt verwertbar ist. Die Frage entfällt bei Gutachten mit Auslandsberührung (z. B. E213).	
Leichte Mängel	Das Deckblatt ist beigelegt. Es ist aber gegenüber dem einheitlichen Formularsatz abgewandelt, alle wesentlichen Elemente sind vorhanden. ODER Das Schlussblatt ist beigelegt. Es ist aber gegenüber dem einheitlichen Formularsatz abgewandelt, alle wesentlichen Elemente sind vorhanden.
Deutliche Mängel	Es fehlt das Deckblatt des einheitlichen Formularsatzes. ODER Es fehlt das Schlussblatt des einheitlichen Formularsatzes.
Gravierende Mängel	Weder das Deckblatt noch das Schlussblatt des einheitlichen Formularsatzes werden verwendet.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

B. Einzelkriterium „Verständlichkeit“

Das Einzelkriterium „*Verständlichkeit*“ untergliedert sich in vier Prüffragen. Geprüft wird die sprachliche Darstellung des Gutachtens sowie der Umgang mit sozialrechtlichen Begriffen und Aussagen. Im Gutachten wird eine allgemeinverständliche und prägnante Sprache erwartet. Fachsprachliche Abkürzungen sind in den Diagnosen und der Epikrise möglichst zu vermeiden bzw. allgemeinsprachlich zu übersetzen. Die sozialmedizinische Terminologie ist richtig anzuwenden. Sozialrechtliche Ausführungen/Folgerungen sind zu unterlassen.

Einzelkriterium B: Verständlichkeit Prüffrage B1	
B1: Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich? Die Frage bezieht sich auf das gesamte Gutachten. Sprache, Grammatik und Satzkonstruktion sind zu beurteilen, nicht jedoch der persönliche Schreibstil des Gutachters. Gefordert ist eine knappe, prägnante Ausdrucksweise. Unnötige Längen, z. B. durch reine Wiederholung von bereits beschriebenen Sachverhalten oder ausführliche Diskussionen irrelevanter Befunde, sind zu vermeiden. Zentral ist die Verständlichkeit der Aussagen.	
Leichte Mängel	Das Gutachten enthält unnötige Längen. ODER Im Gutachten werden unnötig komplizierte Formulierungen verwendet.
Deutliche Mängel	Inhaltliche Aussagen sind vereinzelt nicht eindeutig bzw. nicht verständlich.
Gravierende Mängel	Inhaltliche Aussagen sind überwiegend nicht eindeutig bzw. nicht verständlich.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium B: Verständlichkeit Prüffrage B2	
B2: Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt? Die Frage bezieht sich nur auf Diagnose und Epikrise. Erklärungsbedürftig sind Fachausdrücke und Abkürzungen, welche von fachkundigen Laien nicht sicher vorausgesetzt werden können (z. B. Spondylolisthesis, PTBS, FPI). Ausnahmen sind allgemeinverständliche Fachausdrücke und Abkürzungen (z. B. EKG, HIV). Nicht allgemein verständliche Abkürzungen sind mindestens einmal auszuschreiben bzw. zu erklären. Weiterhin sind Abkürzungen, die für unterschiedliche medizinische Begrifflichkeiten verwendet werden (z. B. HWI: Hinterwandinfarkt, Harnwegsinfekt; BAA: Bauchaortenaneurysma, bradykarde absolute Arrhythmie) zu vermeiden.	
Leichte Mängel	Eine Abkürzung wird nicht mindestens einmal ausgeschrieben, oder es erfolgt keine Erläuterung eines nicht allgemein verständlichen Fachbegriffs.
Deutliche Mängel	Mehrfach werden Abkürzungen nicht mindestens einmal ausgeschrieben, oder es erfolgt mehrfach keine Erläuterung nicht allgemein verständlicher medizinische Fachbegriffe.
Gravierende Mängel	Überwiegend werden Abkürzungen nicht mindestens einmal ausgeschrieben, oder es erfolgt überwiegend keine Erläuterung nicht allgemein verständlicher Fachbegriffe.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium B: Verständlichkeit Prüffrage B3	
B3: Inwieweit wird die sozialmedizinische Terminologie richtig angewendet? Geprüft wird die richtige Anwendung sozialmedizinischer Termini. Sie sind gemäß der Definitionen im Sozialmedizinischen Glossar der Deutschen Rentenversicherung (entsprechend DRV-Schrift Band 81) anzuwenden. Beispiel für eine nicht-richtige Anwendung sozialmedizinischer Termini: dreimal zeitweise für die Arbeitshaltung (dies ist als deutlicher Mangel zu kodieren).	
Leichte Mängel	Sozialmedizinische Termini werden nicht immer korrekt angewendet. ABER Dies lässt keine missverständlichen oder falschen Schlussfolgerungen zu.
Deutliche Mängel	Sozialmedizinische Termini werden nicht immer korrekt angewendet. UND Dies lässt missverständliche Schlussfolgerungen zu.
Gravierende Mängel	Sozialmedizinische Termini werden nicht immer korrekt angewendet. UND Daraus ergeben sich zwangsläufig falsche Schlussfolgerungen.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium B: Verständlichkeit Prüffrage B4	
B4: Inwieweit werden sozialrechtliche Ausführungen/Folgerungen unterlassen? Die Aufgabe des Gutachters ist es, die persönlichen medizinischen Voraussetzungen von Sozialleistungen zu beschreiben, ohne daraus sozialrechtliche Folgerungen zu ziehen. Ein Beispiel für eine nicht zulässige sozialrechtliche Ausführung ist: Der Sozialmediziner äußert sich zur Wegefähigkeit anstelle zur Gehstrecke. Weitere unzulässige sozialrechtliche Folgerungen durch den Gutachter sind: Feststellung einer Erwerbsminderung, Empfehlung einer Rente. Zulässig ist die Wiedergabe sozialanamnestischer Angaben (z.B. MdE, GdB), ohne dass daraus sozialrechtliche Folgerungen durch den Gutachter getätigt werden.	
Leichte Mängel	Einzelne sozialrechtliche Ausführungen sind vorhanden, die die Entscheidung über die Sozialleistung nicht beeinflussen.
Deutliche Mängel	Einzelne sozialrechtliche Ausführungen sind vorhanden, die die Entscheidung über die Sozialleistung beeinflussen könnten (z.B. Äußerung zur Wegefähigkeit statt zur Gehstrecke, Äußerung zu betriebsunüblichen Pausen).
Gravierende Mängel	Im Gutachten finden sich unzulässige präjudizierende sozialrechtliche Folgerungen (z.B. Feststellung einer Erwerbsminderung oder Empfehlung einer Erwerbsminderungsrente).
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

C. Einzelkriterium „Transparenz“

Das Einzelkriterium „*Transparenz*“ untergliedert sich in drei Prüffragen. Erfasst wird, ob die verwendeten Verfahren, Methoden und Instrumente beschrieben wurden und die Kriterien zur Beurteilung der Befunde im Gutachten dargelegt sind. Die Herkunft von anamnestischen Informationen und Befunden muss erkennbar sein.

Einzelkriterium C: Transparenz

Prüffrage C1

C1: Inwieweit ist die Herkunft von medizinischen Informationen erkennbar?

Medizinische Informationen sind:

- Angaben, die der Versicherte selbst gemacht hat
- Angaben von dritten Personen
- Ärztliche Berichte
- Krankenhausentlassungsberichte
- Reha-Entlassungsberichte
- Gutachten der Rentenversicherungsträger, Gutachten im Klageverfahren
- Gutachten anderer Sozialleistungsträger (z. B. Agentur für Arbeit, MDK, BG), Betreuungsgutachten
- Selbst erhobene und in Auftrag gegebene Befunde
- Vor- bzw. Fremdbefunde (z. B. vom Versicherten mitgebrachter Koronarangiographie-Befund durch vorbehandelnden Kardiologen).

Erwartet wird, dass Herkunft (Quelle) und zeitliche Einordnung (Datum) der vorliegenden Informationen zu erkennen sind.

Leichte Mängel

Die Herkunft (Quelle) oder die zeitliche Einordnung (Datum) fehlt bei einer medizinischen Information, die für die sozialmedizinische Leistungsbeurteilung nicht relevant ist.

Deutliche Mängel

Die Herkunft (Quelle) oder die zeitliche Einordnung (Datum) fehlt bei mehreren medizinischen Informationen, die für die sozialmedizinische Leistungsbeurteilung nicht relevant sind.

Gravierende Mängel

Die Herkunft (Quelle) oder die zeitliche Einordnung (Datum) fehlt bei mindestens einer medizinischen Information, die für die sozialmedizinische Leistungsbeurteilung relevant ist.

Welche Mängel wurden festgestellt?

Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren.

Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.

Einzelkriterium C: Transparenz Prüffrage C2	
C2: Inwieweit wird dargestellt, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben werden? Die Prüffrage bezieht sich auf die selbst erhobenen Befunde. Erwartet wird, dass bei selbst erhobenen Befunden die angewendeten Verfahren, Instrumente und Methoden einschließlich der äußeren Rahmenbedingungen dargelegt werden. Beispielsweise ist anzugeben, ob eine Ergometrie im Liegen oder in sitzender Position durchgeführt wurde oder Röntgenaufnahmen im Liegen oder Stehen etc. Auch sind die Rahmenbedingungen der testpsychologischen Untersuchungen zu benennen (z.B. welche Skala erhoben wurde und die Bedingungen der Untersuchungsdurchführung). Wird auf Befunde verwiesen, die in einer nicht vorliegenden Anlage dokumentiert sind, führt dies nicht zu einem Mangel, da die Prüffrage nicht entschieden werden kann.	
Leichte Mängel	Für einen sozialmedizinisch nicht relevanten Befund fehlt die Darstellung der Verfahren, Methoden und Instrumente einschließlich der äußeren Rahmenbedingungen.
Deutliche Mängel	Für mehrere sozialmedizinisch nicht relevante Befunde fehlt die Darstellung der Verfahren, Methoden und Instrumente einschließlich der äußeren Rahmenbedingungen.
Gravierende Mängel	Für mindestens einen sozialmedizinisch relevanten Befund fehlt die Darstellung der Verfahren, Methoden und Instrumente einschließlich der äußeren Rahmenbedingungen.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium C: Transparenz Prüffrage C3	
C3: Inwieweit wird dargestellt, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind? Die Prüffrage bezieht sich auf selbst erhobene apparativ-technische Untersuchungen einschließlich psychometrischer Testverfahren. Unauffällige Befundergebnisse sind als unauffällig zu kennzeichnen. Bei pathologischen Ausprägungen sind Laborwerte mit Referenzbereichen exakt anzugeben. Bei diesen Laborwerten reicht die Kenntlichmachung +/- (im Sinne von über bzw. unter dem Referenzwert) nicht aus. Bei apparativ-technischen Untersuchungen sind die erhobenen Meßwerte und Referenzbereiche anzugeben. Hinweis: Die Neutral-Null-Methode ist keine apparativ-technische Untersuchung, die Angabe von Referenzwerten ist daher nicht erforderlich. Wird auf nicht vorliegende Anlagen verwiesen, führt dies nicht zu einem Mangel, da die Prüffrage nicht entschieden werden kann.	
Leichte Mängel	Es fehlt der exakte Wert/Referenzbereich für einen sozialmedizinisch nicht relevanten Messwert.
Deutliche Mängel	Es fehlen exakte Werte/Referenzbereiche für mehrere sozialmedizinisch nicht relevante Messwerte.
Gravierende Mängel	Es fehlt der exakte Wert/Referenzbereich für mindestens einen sozialmedizinisch relevanten Messwert.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

D. Einzelkriterium „Vollständigkeit“

Das Einzelkriterium „*Vollständigkeit*“ untergliedert sich in neun Prüffragen. Dabei ist zu prüfen, inwieweit Anamnese, Befunde, Diagnosen und Epikrise vollständig dargestellt und die sozialmedizinischen Fragestellungen vollständig beantwortet sind. Hier ist die Bewertung der Vollständigkeit nicht aber die der Richtigkeit der Beantwortung der sozialmedizinischen Fragestellung gefordert.

Die dem Gutachten zugrunde liegenden Befunde sind vom Gutachter ausreichend zu dokumentieren. Befunde, die für die sozialmedizinische Beurteilung eine untergeordnete Rolle spielen, können als nicht relevant gekennzeichnet werden; für diese ist dann eine weitergehende Auseinandersetzung nicht erforderlich.

Der Gutachter muss sich auch mit den vom Versicherten geltend gemachten Beschwerden im Gutachten auseinandersetzen. Bei umfangreichen Beschwerdeschilderungen können diese gegebenenfalls in einer zusammenfassenden Würdigung dargestellt werden. Auch hat vom Gutachter eine Stellungnahme zu bisherigen Therapien des Antragstellers und zu weiteren therapeutischen Optionen zu erfolgen. Wesentliche Abweichungen in der sozialmedizinischen Beurteilung gegenüber Vor-Gutachten sind darzulegen.

Einzelkriterium D: Vollständigkeit Prüffrage D1	
D1: Inwieweit wird die Anamnese vollständig dargestellt? Für die Darstellung der Anamnese werden folgende Inhalte erwartet: Anamnese medizinischer Daten - Eigenanamnese (ggf. mit Darstellung der Tagesstruktur, soweit für die sozialmedizinische Leistungsbeurteilung relevant) - Bei psychiatrischen Gutachten: Biographische Anamnese, soweit störungsbezogen für die sozialmedizinische Beurteilung relevant - Jetzige Beschwerden (Beschwerden, mit denen eine Leistungsminderung begründet wird) - Therapie - Fachrichtung der behandelnden Ärzte, sofern nicht auf dem Formblatt aufgeführt Arbeits- und Sozialanamnese - Berufs- und Arbeitsanamnese - Familiäre Situation - Weitere soziale Kontextfaktoren (z. B. Wohnsituation), Umstände der Antragstellung, bisherige Sozialleistungen	
Leichte Mängel	Einer oder mehrere der folgenden drei Bestandteile der Anamnese fehlen: - Fachrichtung der behandelnden Ärzte - Familiäre Situation - Weitere soziale Kontextfaktoren (z. B. Wohnsituation), Umstände der Antragstellung, bisherige Sozialleistungen
Deutliche Mängel	Es fehlt maximal einer der folgenden Bestandteile der Anamnese: - Eigenanamnese - Biographische Anamnese bei psychiatrischen Gutachten, soweit störungsbezogen für die sozialmedizinische Beurteilung relevant - Jetzige Beschwerden - Therapie - Berufs- und Arbeitsanamnese

Gravierende Mängel	Es fehlen mehrere der folgenden Bestandteile der Anamnese: - Eigenanamnese - Biographische Anamnese bei psychiatrischen Gutachten, soweit störungsbezogen für die sozialmedizinische Beurteilung relevant - Jetzige Beschwerden - Therapie - Berufs- und Arbeitsanamnese
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium D: Vollständigkeit Prüffrage D2	
D2: Inwieweit sind die bei der Beantwortung der sozialmedizinischen Fragestellung zugrunde gelegten Befunde dokumentiert? Alle der Beantwortung der sozialmedizinischen Fragestellung zugrunde gelegten Befunde sind zu dokumentieren. Befunde sind alle selbst erhobenen Befunde sowie Vorbefunde und Vor-Gutachten, die der Versicherte mitbringt oder die bereits in den Unterlagen vorhanden sind: - Ärztliche Berichte - Krankenhausentlassungsberichte - Reha-Entlassungsberichte - Gutachten der Rentenversicherungsträger, Gutachten im Klageverfahren - Gutachten anderer Sozialleistungsträger (z. B. Agentur für Arbeit, MDK, BG), Betreuungsgutachten - Selbst erhobene und in Auftrag gegebene Befunde - Vor- bzw. Fremdbefunde (z. B. vom Versicherten mitgebrachter Koronarangiographie-Befund durch vorbehandelnden Kardiologen). Die Kernaussagen der der sozialmedizinischen Beurteilung zugrunde liegenden Befunde, Vorbefunde und Vor-Gutachten sind darzustellen, bei Vor-Gutachten insbesondere die sozialmedizinische Beurteilung. Die Angabe "Siehe Befundbericht Dr. XXX" ist nicht ausreichend.	
Leichte Mängel	Es fehlt die Dokumentation eines Befundes, der der sozialmedizinischen Beurteilung zugrunde gelegt wird.
Deutliche Mängel	Es fehlt die Dokumentation von mehreren Befunden, die der sozialmedizinischen Beurteilung zugrunde gelegt werden.
Gravierende Mängel	Es fehlt die Dokumentation der überwiegenden Zahl von Befunden, die der sozialmedizinischen Beurteilung zugrunde gelegt werden.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium D: Vollständigkeit Prüffrage D3	
<i>D3: Inwieweit werden die Diagnosen dargestellt und mit Funktions-einschränkungen (sog. Funktionsdiagnosen) formuliert?</i> Bei der Beurteilung dieser Prüffrage ist zu berücksichtigen, inwieweit die diesbezüglichen Vorgaben der DRV-Schrift 21 eingehalten wurden. Diagnosen sind danach gesondert im freien Text aufzulisten, die ersten drei sozialmedizinisch relevanten Diagnosen sind zusätzlich im Formular anzugeben (Schlussblatt). Sie sind - nach Schweregrad (ggf. mit Klassifikation und/oder Stadieneinteilung) ihrer funktionellen Einschränkungen zu ordnen - mit Funktionseinschränkungen anzugeben (z. B. medikamentös eingestellter Diabetes mellitus Typ 2 ohne wesentliche Folgeschäden, z. B. chronisch obstruktive Lungenerkrankung mit dauerhaft schwerwiegender Einschränkung der Atemfunktion) Die Diagnosen sind nach der jeweils gültigen ICD zu verschlüsseln. Die ICD Verschlüsselung ist - sofern möglich - vierstellig (einschließlich Stelle nach dem Punkt) z. B. J45.1 gefordert. Die Verschlüsselung auf dem Schlussblatt ist dazu ausreichend. Inwieweit eine Diagnose aus der Anamnese bzw. der klinischen Untersuchung ableitbar ist, soll hier nicht bewertet werden. Es ist auch nicht zu bewerten, ob die Diagnose gesichert ist.	
Leichte Mängel	Bei mindestens einer der ersten drei Diagnosen fehlt die ICD-Verschlüsselung
Deutliche Mängel	Die sprachliche Formulierung der Diagnosen (z. B. Asthma bronchiale) fehlt. ODER Die Funktionseinschränkungen sind nicht benannt.
Gravierende Mängel	Die sprachliche Formulierung der Diagnosen (z. B. Asthma bronchiale) fehlt. UND Die Funktionseinschränkungen sind nicht benannt.

Welche Mängel wurden festgestellt?

Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren.

Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.

Einzelkriterium D: Vollständigkeit Prüffrage D4	
<i>D4: Inwieweit werden die vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen in der Epikrise erfasst?</i> Erwartet wird eine Auseinandersetzung mit der medizinischen Relevanz der vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen. Eine reine Wiederholung der anamnestischen Angaben des Versicherten ist nicht ausreichend. Im Falle einer sehr umfangreichen Beschwerdeschilderung durch den Versicherten kann diese in der Epikrise zusammenfassend gewürdigt werden. Beschwerden, die vom Versicherten geltend gemacht wurden, aber für die sozialmedizinische Beurteilung eine untergeordnete Rolle spielen, können als nicht relevant gekennzeichnet werden; für diese ist dann eine weitergehende Auseinandersetzung nicht erforderlich.	
Leichte Mängel	Die Epikrise setzt sich nicht mit allen* vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen auseinander. Eine Auseinandersetzung erfolgt aber für alle sozialmedizinisch relevanten Beschwerden, Erkrankungen und Funktionseinschränkungen. * Im Falle einer sehr umfangreichen Beschwerdeschilderung durch den Versicherten kann diese in der Epikrise zusammenfassend gewürdigt werden. Weiterhin ist als für sozialmedizinisch nicht relevant gekennzeichnete Beschwerden keine weitergehende Auseinandersetzung erforderlich
Deutliche Mängel	Für einzelne, vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen, die sozialmedizinisch relevant sind, fehlt die Auseinandersetzung.
Gravierende Mängel	Für die überwiegende Zahl der vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen, die sozialmedizinisch relevant sind, fehlt die Auseinandersetzung.

Welche Mängel wurden festgestellt?

Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren.

Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.

Einzelkriterium D: Vollständigkeit Prüffrage D5	
D5: Inwieweit werden die im Gutachten aufgeführten Befunde in der Epikrise erfasst? Erwartet wird eine Auseinandersetzung mit der medizinischen Relevanz von Fremdbefunden und selbst erhobenen Befunden. Befunde, die für die sozialmedizinische Beurteilung eine untergeordnete Rolle spielen, können als nicht relevant gekennzeichnet werden; für diese ist dann eine weitergehende Auseinandersetzung nicht erforderlich.	
Leichte Mängel	Das Gutachten setzt sich nicht mit allen* im Gutachten aufgeführten Befunden auseinander. Eine Auseinandersetzung erfolgt aber mit allen sozialmedizinisch relevanten Befunden. * Für als sozialmedizinische nicht relevant gekennzeichnete Befunde ist keine weitergehende Auseinandersetzung erforderlich
Deutliche Mängel	Für einzelne im Gutachten aufgeführte und sozialmedizinisch relevante Befunde fehlt die Auseinandersetzung.
Gravierende Mängel	Für die überwiegende Zahl der im Gutachten aufgeführten und sozialmedizinisch relevanten Befunde fehlt die Auseinandersetzung.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium D: Vollständigkeit Prüffrage D6	
<i>D6: Inwieweit werden die festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit im Erwerbsleben bewertet?</i> Zu prüfen ist, inwieweit die im Gutachten aus den Befunden und Diagnosen abgeleiteten Funktionsstörungen auch sozialmedizinisch bewertet werden. Die Prüffrage bezieht sich auf die Vollständigkeit, nicht aber auf die Richtigkeit der sozialmedizinischen Bewertung der festgestellten Funktionsstörungen sowohl in den Ausführungen als auch auf dem Schlussblatt.	
Leichte Mängel	Im Gutachten werden nicht alle der festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit im Erwerbsleben bewertet. Eine Bewertung erfolgt aber für alle festgestellten sozialmedizinisch relevanten Funktionsstörungen.
Deutliche Mängel	Für einzelne der festgestellten sozialmedizinisch relevanten Funktionsstörungen fehlt die Bewertung.
Gravierende Mängel	Für die überwiegende Zahl der festgestellten sozialmedizinisch relevanten Funktionsstörungen fehlt die Bewertung.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium D: Vollständigkeit Prüffrage D7	
<i>D7: Inwiefern erfolgt eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen?</i> Erwartet wird in der Epikrise eine Stellungnahme zu bisherigen Behandlungen und gegebenenfalls zu Therapien, die geeignet sind, die Leistungsfähigkeit deutlich zu steigern. Ist ersichtlich, ob ein Behandlungsfall vorliegt?	
Leichte Mängel	Es fehlt eine Stellungnahme zu bisherigen Behandlungen.
Deutliche Mängel	Es fehlt eine Stellungnahme zu Therapien, die geeignet sind, die Leistungsfähigkeit deutlich zu steigern.
Gravierende Mängel	Es fehlt eine Stellungnahme zu bisherigen Behandlungen. UND Es fehlt eine Stellungnahme zu Therapien, die geeignet sind, die Leistungsfähigkeit deutlich zu steigern.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium D: Vollständigkeit Prüffrage D8	
<i>D8: Inwieweit werden wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten dargelegt?</i> Hierbei geht es um abweichende Aussagen zur Leistungsfähigkeit, die eine andere sozialrechtliche Entscheidung nach sich ziehen könnten. Es wird eine Auseinandersetzung mit solchen Abweichungen erwartet. Unter Vor-Gutachten wird dabei verstanden: Reha-Entlassungsbericht, Gutachten der Rentenversicherungsträger, Gutachten im Klageverfahren, Gutachten anderer Sozialleistungsträger (z. B. der Agentur für Arbeit, des MDK, der BG), Betreuungsgutachten) Liegen aus Vor-Gutachten keine Aussagen zur Leistungsfähigkeit vor, so führt dies nicht zu einem Mangel.	
Leichte Mängel	Es gibt wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vor-Gutachten. Diese werden im Gutachten aufgezeigt, es unterbleibt jedoch eine Bewertung von Teilaspekten der Abweichung.
Deutliche Mängel	Es gibt wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vor-Gutachten. Diese werden im Gutachten aufgezeigt, es unterbleibt jedoch eine Bewertung dieser Abweichungen.
Gravierende Mängel	Es gibt wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vor-Gutachten. Die Abweichungen werden im Gutachten nicht aufgezeigt.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium D: Vollständigkeit Prüffrage D9	
D9: Inwieweit werden die sozialmedizinischen Fragestellungen vollständig beantwortet? Fragestellungen an das sozialmedizinische Gutachten der Deutschen Rentenversicherung sind: - quantitatives Leistungsvermögen - qualitatives Leistungsvermögen - Beginn/Dauer einer Leistungsminderung - Reha-Bedarf. Diese Fragestellungen sind formuliert im – bundesweit konsentierten – Schlussblatt zum Gutachten (siehe DRV Schrift 21). Darüber hinaus können sich Fragestellungen aus den Angaben des Versicherten oder aus den Befunden und Diagnosen im Einzelfall ergeben: z. B. Gehfähigkeit, Gebrauchsfähigkeit beider Hände, zusätzliche Arbeitsunterbrechungen. Unabdingbar ist eine Aussage zum Bedarf von Leistungen zur Teilhabe. Es reicht nicht aus, wenn die fehlende Notwendigkeit einer Leistung zur Teilhabe durch Ankreuzen auf dem Schlussblatt dokumentiert wird. Bei einer Leistungsfähigkeit im Erwerbsleben von unter 3h ist es nicht erforderlich, die qualitative Leistungsfähigkeit im Erwerbsleben darzustellen. Im Peer-Review wird nicht geprüft, ob eine Beurteilung der Leistungsfähigkeit in Bezug auf die zuletzt ausgeübte Tätigkeit durch den Gutachter erfolgt ist.	
Leichte Mängel	Es fehlen Aussagen zum Reha-Bedarf.
Deutliche Mängel	Es fehlen Aussagen zu Beginn/Dauer einer Leistungsminderung.
Gravierende Mängel	Es fehlen Aussagen zur qualitativen Leistungsfähigkeit. ODER Es fehlen Aussagen zur quantitativen Leistungsfähigkeit.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

E. Einzelkriterium „Medizinisch-wissenschaftliche Grundlagen“

Das Einzelkriterium „*Medizinisch-wissenschaftliche Grundlagen*“ untergliedert sich in zwei Prüffragen. Zu prüfen ist, inwieweit der allgemein anerkannte Stand medizinischer Erkenntnisse eingehalten und inwieweit die Begutachtungsliteratur der Deutschen Rentenversicherung berücksichtigt werden.

Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen Prüffrage E1	
E1: Inwieweit wird der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet? Geprüft wird, ob bei der Gewinnung der medizinischen Informationen, der Prognose und den Therapieoptionen der allgemein anerkannte Stand medizinischer Erkenntnisse eingehalten wird. Die Schlüssigkeit der Diagnose und deren Ableitung werden hier nicht geprüft.	
Leichte Mängel	Fallbezogen werden Außenseitermethoden verwendet, jedoch nicht zur sozialmedizinischen Beurteilung herangezogen.
Deutliche Mängel	Außenseitermethoden werden zusätzlich zur Herleitung der sozialmedizinischen Beurteilung herangezogen.
Gravierende Mängel	Die sozialmedizinische Beurteilung wird mit Außenseitermethoden begründet.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen

Prüffrage E2

E2: Inwieweit wird die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet?

Geprüft wird, inwieweit grobe Abweichungen von der Begutachtungsliteratur zu erkennen sind. Hier ist die von der Rentenversicherung veröffentlichte sozialmedizinische Begutachtungsliteratur gemeint, insbesondere die Buchveröffentlichung „Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung“ und die Leitlinien zur Begutachtung der Deutschen Rentenversicherung (veröffentlicht auf den DRV-Internetseiten).

Leitlinien müssen nicht zitiert werden.

Begründete Abweichungen von der Begutachtungsliteratur führen nicht zu einer Mängelbewertung. Die Richtigkeit der Begründung ist nicht Gegenstand der Prüfung.

Leichte Mängel

Es liegt eine Abweichung von der Begutachtungsliteratur vor, die keine Auswirkung auf die sozialmedizinische Beurteilung hat.

Deutliche Mängel

Es liegt mehr als eine Abweichung von der Begutachtungsliteratur vor, die keine Auswirkung auf die sozialmedizinische Beurteilung haben.

Gravierende Mängel

Es liegt mindestens eine Abweichung von der Begutachtungsliteratur vor, die Auswirkungen auf die sozialmedizinische Beurteilung hat.

Welche Mängel wurden festgestellt?

Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren.

Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.

F. Einzelkriterium „Wirtschaftlichkeit“

Das Einzelkriterium „*Wirtschaftlichkeit*“ umfasst zwei Prüffragen. Diese beziehen sich auf die Diagnostik und orientieren sich am Wirtschaftlichkeitsgebot des SGB V. Verfahren, Methoden und Instrumente sind dementsprechend auszurichten. Der Peer prüft, ob ein Gutachter Sinn und Zweck einer Untersuchung ersichtlich gemacht hat und in der Epikrise auch Schlüsse daraus zieht.

Einzelkriterium F: Wirtschaftlichkeit Prüffrage F1	
F1: Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig? Erwartet wird, dass die eingesetzte medizinisch-technische Diagnostik und eventuelle testpsychologische Untersuchungen zur Beantwortung der Fragestellung geeignet sind und das Maß des Notwendigen nicht überschreiten. Dabei ist zu prüfen, inwieweit die angewandte Diagnostik tatsächlich erforderlich ist, um zu einer sozialmedizinischen Beurteilung zu kommen, z. B. wird bei Vorliegen aktueller Fremdbefunde auf Diagnostik verzichtet. Diagnosesicherung und Differentialdiagnostik sind nicht vorrangig Aufgabe der sozialmedizinischen Begutachtung. Die Indikation für die eingesetzte Diagnostik muss sich aus dem Gutachten ergeben.	
Leichte Mängel	Die eingesetzte Diagnostik ist für die Beantwortung der sozialmedizinischen Fragestellung geeignet. Vereinzelt wird das Maß des Notwendigen überschritten.
Deutliche Mängel	Die eingesetzte Diagnostik ist für die Beantwortung der sozialmedizinischen Fragestellung geeignet. Mehrfach wird das Maß des Notwendigen überschritten.
Gravierende Mängel	Die eingesetzte Diagnostik ist für die Beantwortung der sozialmedizinischen Fragestellung nicht geeignet und damit auch nicht notwendig.
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

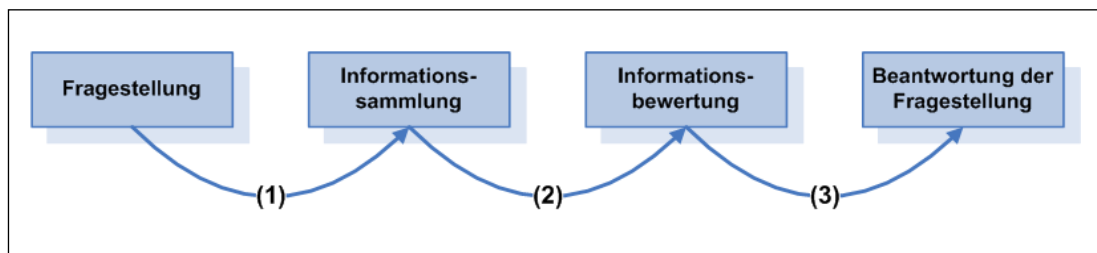
Einzelkriterium F: Wirtschaftlichkeit Prüffrage F2	
F2: Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend? Erwartet wird, dass die eingesetzte medizinisch-technische Diagnostik und eventuelle testpsychologische Untersuchungen zur Beantwortung der Fragestellung ausreichend sind. Bei nicht ausreichenden Vorbefunden ist eine eigene Diagnostik durch den Gutachter einzusetzen. Unter Wirtschaftlichkeitsaspekten kann auch das Fehlen von Diagnostik relevant sein, wenn dadurch zusätzlicher Bearbeitungsaufwand, weitere Diagnostik oder Gutachten notwendig werden.	
Leichte Mängel	Eine zur Stützung der sozialmedizinischen Beurteilung indizierte Untersuchung fehlt (z. B. fehlende Blutgasanalyse bei COPD). Zusätzlicher Bearbeitungsaufwand ist jedoch nicht notwendig.
Deutliche Mängel	Einzelne zur Unterlegung der sozialmedizinischen Beurteilung indizierte Untersuchungen fehlen und erschweren die Beurteilung (z. B. bei COPD nur Untersuchung der Lungenfunktion (LuFu) durchgeführt, Bodyplethysmographie fehlt).
Gravierende Mängel	Es fehlen wesentliche Untersuchungen, die zur Beantwortung der sozialmedizinischen Fragestellungen erforderlich gewesen wären. Dies zieht zusätzlichen Bearbeitungsaufwand nach sich (z. B. bei COPD nur körperliche Untersuchung durchgeführt, Lungenfunktion (LuFu) und Bodyplethysmographie fehlen).
Welche Mängel wurden festgestellt? Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen (in Stichworten) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Übergeordnetes Kriterium „Nachvollziehbarkeit des Gutachtens“

Im übergeordneten Qualitätskriterium wird das gesamte Gutachten im Zusammenhang betrachtet und die Nachvollziehbarkeit des Gutachtens anhand der Argumentationskette bewertet.

Übergeordnetes Kriterium: Nachvollziehbarkeit des Gutachtens**Bewerten Sie die Nachvollziehbarkeit des Gutachtens anhand der Argumentationsführung.**

Dieses Kriterium bildet die Nachvollziehbarkeit, d. h. innere Logik des Gutachtens als Ganzes ab. Gefragt wird nach der Argumentationskette, d. h. der sinnvollen Verknüpfung der verschiedenen Begutachtungsschritte und schlüssigen Herleitung der sozialmedizinischen Beurteilung der Leistungsfähigkeit. Die Zielgrößen der Überprüfung der Argumentationskette sind die sozialmedizinische Beurteilung der quantitativen und qualitativen Leistungsfähigkeit sowie Beginn und Dauer der Leistungsminderung. Abhängig von der Fallkonstellation können Mängel in den Gliedern der Argumentationskette unterschiedlich gewichtet werden. Diese Gewichtung im Einzelfall vorzunehmen ist Aufgabe des Peers. Beispielsweise wäre zu besonderen qualitativen Einschränkungen Stellung zu nehmen, wenn diese aus dem Krankheitsbild zu erwarten sind wie bei der arteriellen Verschlusskrankheit eine Einschränkung der Gehstrecke oder bei der rheumatoiden Arthritis eine Einschränkung der Gebrauchsfähigkeit der Hände.

Gutachten-
auftragAnamnese
BefundFunktions-)
Diagnosen,
EpikriseSozialmedizinische
Beurteilung der
Leistungsfähigkeit
Reha-Bedarf

Grundlage der Bewertung ist die Schlüssigkeit der abgebildeten Argumentationskette mit den aufgeführten Begutachtungsschritten. Hier sind bei der Bewertung eines Gutachtens Mängel d. h. Brüche der Argumentationskette vom Peer zu benennen und anhand des Gutachtens konkret zu belegen.

Die Bewertung der Argumentationskette wird über ein Ampelsystem vorgenommen:

Die Ampelbewertung GRÜN kennzeichnet, dass keine nennenswerten Brüche in der Argumentationskette vorliegen.

Bei der Ampelbewertung GELB liegen Brüche in der Argumentationskette vor, jedoch ermöglichen die im Gutachten enthaltenen sozialmedizinischen Informationen dem Peer, diese durch entsprechende Neu-Bewertung zu überbrücken und das Gutachtenergebnis damit nachzuvollziehen.

Bei der Ampelbewertung ROT liegen Brüche in der Argumentationskette vor, die vom Sozialmediziner auch mittels Neu-Bewertung der sozialmedizinischen Informationen nicht mehr überbrückt werden können.

Bei der Bewertung des übergeordneten Kriteriums kommt es nicht darauf an, ob der Peer das Gutachtenergebnis teilt oder er selbst zu einem anderen Ergebnis kommt. Da eine gravierende Verletzung des Neutralitätsgebotes die Nachvollziehbarkeit des gesamten Gutachtens grundsätzlich in Frage stellt, ist an dieser Stelle auch die Neutralität zu überprüfen. Eine Verletzung der Neutralität liegt vor, wenn offensichtlich herabsetzende und beleidigende Äußerungen vom Gutachter gegenüber einer Person oder Bevölkerungsgruppe gemacht werden. Die Voreingenommenheit und Parteilichkeit muss aus dem Text des Gutachters eindeutig belegbar sein. Ein gravierender Verstoß gegen das Neutralitätsgebot führt zu der Ampelbewertung „Rot“.	
Grün	Die Argumentationskette für die sozialmedizinische Beurteilung ist in sich schlüssig. Es sind keine nennenswerten Brüche in der Argumentationskette vorhanden.
Gelb	Die Argumentationskette für die sozialmedizinische Beurteilung ist lückenhaft, vom Sozialmediziner können jedoch Zusammenhänge selbst hergestellt und festgestellte Brüche in der Argumentationskette überbrückt werden.
Rot	Die Argumentationskette für die sozialmedizinische Beurteilung ist unterbrochen, die Brüche können auch vom Sozialmediziner nicht überbrückt werden. ODER Es liegt ein gravierender Verstoß gegen das Neutralitätsgebot vor.
Welche Mängel wurden festgestellt? Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, (sind in Stichworten auf dem Prüfbogen) zu dokumentieren. Die Dokumentation soll dazu beitragen, den Begutachtungsprozess Transparenz zu verleihen und nachvollziehbar zu gestalten, insbesondere für die Verfasser von Gutachten.	

Anhang 6:

Prüffragenspezifische Codes der qualitativen Inhaltsanalyse der Einzelkriterien

Prüffragenspezifische Codes der Prüffrage A1

Prüffrage A1	bezieht sich auf das rein formale Vorhandensein der Gliederungsvorgabe der Überschriften für sozialmedizinische Gutachten. Als <i>Einschränkung</i> wird erwähnt, dass eine inhaltliche Bewertung (Angemessenheit, Vollständigkeit) nicht vorgesehen ist.

Code	Codebeschreibung
A1L	Die Reihenfolge der Gliederungspunkte wird nicht eingehalten.
A1D	Ein Gliederungspunkt, der im DRV-Band 21 dargelegten Überschriften fehlt.
A1G	Mehr als ein Gliederungspunkt der im DRV-Band 21 dargelegten Überschriften fehlt.

Prüffragenspezifische Codes der Prüffrage A2

Prüffrage A2	bezieht sich auf die Verwendung und das Vorhandensein des einheitlichen Deck- und Schlussblattes. Als <i>Einschränkungen</i> wird aufgeführt, dass die Prüffrage bei Gutachten mit Auslandsberührung entfällt. Ferner soll seitens der Peers nicht geprüft werden, ob Aussagen des Deck- und Schlussblattes an andere Stelle im Gutachten vorhanden sind oder ob das Gutachten aufgrund des Fehlens von Deck- und Schlussblatt noch verwertbar ist.

Code	Codebeschreibung
A2L1	Das dem Gutachten beigelegte Deckblatt ist abgewandelt, verfügt jedoch gegenüber dem einheitlichen Formularsatz über alle wesentlichen Elemente
A2L2	Das dem Gutachten beigelegte Schlussblatt ist abgewandelt, verfügt jedoch gegenüber dem einheitlichen Formularsatz über alle wesentlichen Elemente
A2D1	Das Deckblatt des einheitlichen Formularsatzes fehlt.
A2D2	Das Schlussblatt des einheitlichen Formularsatzes fehlt.
A2G	Das Schluss- und das Deckblatt des einheitlichen Formularsatzes fehlen.

Prüffragenspezifische Codes der Prüffrage B1

Prüffrage B1	überprüft die Sprache, Grammatik und Satzkonstruktion im gesamten Gutachten, um unnötige Längen und komplizierte Formulierungen zu vermeiden. Als <i>Einschränkung</i> wird angegeben, dass der persönliche Schreibstil des Gutachters nicht Gegenstand der Beurteilung sein soll.

Code	Codebeschreibung
B1L1	Das Gutachten enthält unnötige Längen.
B1L2	Im Gutachten werden unnötig komplizierte Formulierungen verwendet.
B1D	Vereinzelte inhaltliche Aussagen nicht eindeutig bzw. verständlich.
B1G	Überwiegend sind inhaltliche Aussagen nicht eindeutig bzw. verständlich.

Prüffragenspezifische Codes der Prüffrage B2

Prüffrage B2	Überprüft, inwieweit Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt wurden. Als <i>Einschränkung</i> wird angegeben, dass die Prüffrage ausschließlich auf die Gutachtenbereiche Diagnose und Epikrise zu beziehen ist.

Code	Codebeschreibung
B2L1	Eine Abkürzung wird nicht mindestens einmal ausgeschrieben.
B2L2	Ein Fachbegriff wird nicht allgemeinverständlich erläutert.
B2D1	Mehrfach werden Abkürzungen nicht mindestens einmal ausgeschrieben.
B2D2	Mehrere Fachbegriffe werden nicht allgemeinverständlich erläutert.
B2G1	Überwiegend werden Abkürzungen nicht mindestens einmal ausgeschrieben.
B2G2	Überwiegende Zahl der Fachbegriffe wird nicht allgemeinverständlich erläutert.

Prüffragenspezifische Codes der Prüffrage B3

Prüffrage B3	untersucht, inwieweit die sozialmedizinische Terminologie, gemäß der Definition im Sozialmedizinischen Glossar der Deutschen Rentenversicherung (DRV Schrift 81), richtig angewendet wird. Die Prüffrage enthält <i>keine Einschränkungen</i>.

Code	Codebeschreibung
B3L	Die sozialmedizinischen Termini werden nicht immer korrekt angewandt, dies lässt jedoch keine missverständlichen oder falschen Schlussfolgerungen zu.
B3D	Die sozialmedizinischen Termini werden nicht immer korrekt angewandt, dies lässt missverständliche Schlussfolgerungen zu.
B3G	Die sozialmedizinischen Termini werden nicht immer korrekt angewandt, dies lässt zwangsläufig falsche Schlussfolgerungen zu.

Prüffragenspezifische Codes der Prüffrage B4

Prüffrage B4	überprüft, inwieweit der Gutachter bei der Beschreibung der persönlich medizinischen Voraussetzungen von Sozialleistungen sozialrechtliche Ausführungen bzw. Folgerungen unterlässt. Die Prüffrage enthält <i>keine Einschränkungen</i>.

Code	Codebeschreibung
B4L	Einige vorhandene sozialrechtliche Folgerungen beeinflussen die Entscheidung über die Sozialleistung nicht.
B4D	Einige vorhandene sozialrechtliche Folgerungen könnten die Entscheidung über die Sozialleistung beeinflussen.
B4G	Das Gutachten enthält unzulässige präjudizierende sozialrechtliche Folgerungen.

Prüffragenspezifische Codes der Prüffrage C1

Prüffrage C1	untersucht, inwieweit die Herkunft von medizinischen Informationen im Gutachten erkennbar ist. Die Prüffrage enthält <i>keine Einschränkungen</i>.

Code	Codebeschreibung
C1L1	Die Herkunft/Quelle einer für die sozialmedizinische Leistungsbeurteilung <i>nicht relevanten</i> medizinischen Information fehlt /ist nicht angegeben.
C1L2	Die zeitliche Einordnung einer für die sozialmedizinische Leistungsbeurteilung <i>nicht relevanten</i> medizinischen Information fehlt.
C1D1	Die Herkunft/Quelle mehrerer, für die sozialmedizinische Leistungsbeurteilung <i>nicht relevanten</i> medizinischen Informationen fehlt.
C1D2	Die zeitliche Einordnung mehrerer, für die sozialmedizinische Leistungsbeurteilung <i>nicht relevanten</i> medizinischen Informationen fehlt.
C1G1	Die Herkunft/Quelle einer für die sozialmedizinische Leistungsbeurteilung <i>relevanten</i> medizinischen Information fehlt.
C1G2	Die zeitliche Einordnung einer für die sozialmedizinische Leistungsbeurteilung <i>relevanten</i> medizinischen Information fehlt.

Prüffragenspezifische Codes der Prüffrage C2

Prüffrage C2	untersucht, inwieweit dargestellt wird, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben wurden. Als <i>Einschränkung</i> wird angegeben, dass sich die Prüffrage ausschließlich auf selbst erhobene Befunde bezieht. Darüber hinaus führt ein Verweis auf Befunde im Anhang nicht zu einem Mangel.

Code	Codebeschreibung
C2L	Die Erhebungsdarstellung eines sozialmedizinisch <i>nicht relevanten</i> Befundes fehlt.
C2D	Die Erhebungsdarstellung mehrerer sozialmedizinisch nicht relevanter Befunde fehlt.
C2G	Die Erhebungsdarstellung mindestens eines sozialmedizinisch <i>relevanten</i> Befundes fehlt.

Prüffragenspezifische Codes der Prüffrage C3

Prüffrage C3	überprüft, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind. Als <i>Einschränkung</i> wird angegeben, dass sich die Prüffrage ausschließlich auf selbst erhobene apparativ-technische Befunde bezieht. Darüber hinaus führt ein Verweis auf Befunde im Anhang nicht zu einem Mangel.

Code	Codebeschreibung
C3L	Ein exakter Referenzbereich für einen sozialmedizinisch <i>nicht relevanten</i> Messwert fehlt.
C3D	Mehrere exakte Referenzbereiche für sozialmedizinisch <i>nicht relevante</i> Messwerte fehlen.
C3G	Mindestens ein exakter Referenzbereich für einen sozialmedizinisch <i>relevanten</i> Messwert fehlt.

Prüffragenspezifische Codes der Prüffrage D1

Prüffrage D1	überprüft, inwieweit die Anamnese vollständig dargestellt wird. Die Prüffrage enthält <i>keine Einschränkungen</i>.

Code	Codebeschreibung
D1L	Es fehlen ein oder mehrere folgender Bestandteile: Fachrichtung Arzt, Familiäre Situation, soziale Kontextfaktoren.
D1D	Es fehlt maximal einer dieser Bestandteile: Eigenanamnese, Biographische Anamnese, Jetzige Beschwerden, Therapie, Berufs- und Arbeitsanamnese fehlen.
D1G	Es fehlen mehrere dieser Bestandteile: Eigenanamnese, Biographische Anamnese, Jetzige Beschwerden, Therapie, Berufs- und Arbeitsanamnese fehlen.

Prüffragenspezifische Codes der Prüffrage D2

Prüffrage D2	überprüft, inwieweit die bei der Beantwortung der sozialmedizinischen Fragestellung zugrunde gelegten Befunde dokumentiert sind. Die Prüffrage enthält <i>keine Einschränkungen</i> .

Code	Codebeschreibung
D2L	Die Dokumentation eines Befundes fehlt.
D2D	Die Dokumentation mehrerer Befunde fehlt.
D2G	Die Dokumentation der überwiegenden Zahl der Befunde fehlt.

Prüffragenspezifische Codes der Prüffrage D3

Prüffrage D3	überprüft, inwieweit die Diagnosen dargestellt und mit entsprechenden Funktionseinschränkungen (Funktionsdiagnosen) formuliert wurden. Als <i>Einschränkungen</i> wird angegeben, dass nicht Gegenstand der Prüfung ist, ob eine Diagnose aus der Anamnese oder der Untersuchung ableitbar ist. So ist auch die Sicherheit der Diagnose nicht Gegenstand des Reviews.

Code	Codebeschreibung
D3L	Bei mindestens einer der ersten drei Diagnosen fehlt die ICD-Verschlüsselungen.
D3D1	Die sprachliche Formulierungen der Diagnose fehlt
D3D2	Die Funktionseinschränkungen sind nicht benannt.
D3G	Die sprachliche Formulierung der Diagnosen fehlt und die Funktionseinschränkungen sind nicht benannt.

Prüffragenspezifische Codes der Prüffrage D4

Prüffrage D4	untersucht, inwieweit die vom Versicherten geltend gemachten Beschwerden, Erkrankungen und Funktionseinschränkungen in der Epikrise erfasst wurden. Als <i>Einschränkung</i> wird abgegeben, dass die Prüffrage ausschließlich auf die Epikrise zu beziehen ist.

Code	Codebeschreibung
D4L	Die Auseinandersetzung mit sozialmedizinisch <i>nicht relevanten</i> Beschwerden, Erkrankungen und Funktionseinschränkungen ist nicht vollständig.
D4D	Die Auseinandersetzung mit einzelnen sozialmedizinisch <i>relevanten</i> Beschwerden, Erkrankungen und Funktionseinschränkungen fehlt.
D4G	Die Auseinandersetzung mit den überwiegend sozialmedizinisch <i>relevanten</i> Beschwerden, Erkrankungen und Funktionseinschränkungen fehlt.

Prüffragenspezifische Codes der Prüffrage D5

Prüffrage D5	überprüft, inwieweit die im Gutachten aufgeführten Befunde in der Epikrise erfasst sind. Als <i>Einschränkung</i> wird abgegeben, sich die Prüffrage ausschließlich auf die Epikrise zu beziehen ist.

Code	Codebeschreibung
D5L	Die Auseinandersetzung mit sozialmedizinisch <i>nicht relevanten</i> Befunden ist nicht vollständig.
D5D	Die Auseinandersetzung mit einzelnen sozialmedizinisch <i>relevanten</i> Befunden fehlt.
D5G	Die Auseinandersetzung mit den überwiegend sozialmedizinisch <i>relevanten</i> Befunden fehlt.

Prüffragenspezifische Codes der Prüffrage D6

Prüffrage D6	überprüft, inwieweit die festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit im Erwerbsleben bewertet werden. Als <i>Einschränkung</i> wird abgegeben, dass sich die Prüffrage auf die Vollständigkeit, jedoch nicht auf die Richtigkeit der sozialmedizinischen Bewertung bezieht. Diese ist nicht Bestandteil der Prüffrage.

Code	Codebeschreibung
D6L	Die Bewertung sozialmedizinisch <i>nicht relevanter</i> festgestellter Funktionsstörungen in Bezug auf die Leistungsfähigkeit ist nicht vollständig.
D6D	Die Auseinandersetzung mit einzelnen sozialmedizinisch <i>relevanten</i> festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit fehlt.
D6G	Die Auseinandersetzung mit der überwiegenden Zahl der sozialmedizinisch <i>relevanten</i> festgestellten Funktionsstörungen in Bezug auf die Leistungsfähigkeit fehlt.

Prüffragenspezifische Codes der Prüffrage D7

Prüffrage D7	untersucht, inwiefern eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen erfolgte. Als <i>Einschränkung</i> wird abgegeben, dass die Prüffrage ausschließlich auf die Epikrise zu beziehen ist.

Code	Codebeschreibung
D7L	Eine Stellungnahme zur bisherigen Behandlung fehlt.
D7D	Eine Stellungnahme zu Therapien zur Steigerung der Leistungsfähigkeit fehlt.
D7G	Eine Stellungnahme zur bisherigen Behandlung und möglichen Therapien zur Steigerung der Leistungsfähigkeit fehlt.

Prüffragenspezifische Codes der Prüffrage D8

Prüffrage D8	prüft, inwieweit wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber dem Vorgutachten dargelegt wurden. Als <i>Einschränkung</i> wird aufgeführt, dass das Fehlen von Aussagen zur Leistungsfähigkeit aus Vor-Gutachten nicht zu einem Mangel führt.

Code	Codebeschreibung
D8L	Es gibt Abweichungen gegenüber dem Vor-Gutachten, die aufgezeigt wurden. Es unterbleibt jedoch die Bewertung von Teilaspekten der Abweichung.
D8D	Es gibt Abweichungen gegenüber dem Vor-Gutachten, die aufgezeigt wurden. Es unterbleibt jedoch die vollständige Bewertung dieser Abweichungen.
D8G	Es gibt Abweichungen gegenüber dem Vor-Gutachten. Diese Abweichungen werden im Gutachten nicht aufgezeigt.

Prüffragenspezifische Codes der Prüffrage D9

Prüffrage D9	untersucht, inwieweit sozialmedizinische Fragestellungen vollständig beantwortet werden. Als <i>Einschränkung</i> wird aufgeführt, dass das Vorhandensein der Beurteilung der Leistungsfähigkeit in Bezug auf die letzte Tätigkeit nicht Gegenstand der Prüffrage sein soll.

Code	Codebeschreibung
D9L	Aussagen zum Reha-Bedarf fehlen.
D9D	Aussagen zum Beginn bzw. zur Dauer der Leistungsminderung fehlen.
D9G1	Es fehlen Aussagen zur quantitativen Leistungsfähigkeit.
D9G2	Es fehlen Aussagen zur qualitativen Leistungsfähigkeit.

Prüffragenspezifische Codes der Prüffrage E1

Prüffrage E1	untersucht, inwieweit der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet wird. Als <i>Einschränkung</i> wird aufgeführt, dass die Schlüssigkeit der Diagnose und deren Ableitung nicht Gegenstand der Prüffrage sind.

Code	Codebeschreibung
E1L	Es werden Außenseitermethoden verwendet, jedoch nicht zur sozialmedizinischen Beurteilung herangezogen.
E1D	Es werden zusätzlich Außenseitermethoden zur Herleitung der sozialmedizinischen Beurteilung herangezogen.
E1G	Die sozialmedizinische Beurteilung wird durch Außenseitermethoden begründet.

Prüffragenspezifische Codes der Prüffrage E2

Prüffrage E2	untersucht, inwieweit die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet wird. Als <i>Einschränkung</i> wird aufgeführt, dass eine begründete Abweichung von der Literatur nicht zu einem Mangel führt.

Code	Codebeschreibung
E2L	Eine sozialmedizinisch <i>nicht relevante</i> Abweichung von der Begutachtungsliteratur liegt vor.
E2D	Mehr als eine sozialmedizinisch <i>nicht relevante</i> Abweichung von der Begutachtungsliteratur liegt vor.
E2G	Mindestens eine sozialmedizinisch <i>relevante</i> Abweichung der Begutachtungsliteratur liegt vor.

Prüffragenspezifische Codes der Prüffrage F1

Prüffrage F1	untersucht, inwieweit die Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig ist. Die Prüffrage enthält <i>keine Einschränkungen</i>.

Code	Codebeschreibung
F1L	Die eingesetzte Diagnostik ist zur Beantwortung der Fragestellung geeignet, überschreitet jedoch das notwendige Maß vereinzelt.
F1D	Die eingesetzte Diagnostik ist zur Beantwortung der Fragestellung geeignet, überschreitet jedoch das notwendige Maß mehrfach.
F1G	Die eingesetzte Diagnostik ist zur Beantwortung der Fragestellung nicht geeignet und daher nicht notwendig.

Prüffragenspezifische Codes der Prüffrage F2

Prüffrage F2	untersucht, inwieweit die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend ist. Die Prüffrage enthält <i>keine Einschränkungen</i>.

Code	Codebeschreibung
F2L	Eine zur Stützung der sozialmedizinischen Beurteilung indizierte Untersuchung fehlt, jedoch ist kein zusätzlicher Bearbeitungsaufwand notwendig.
F2D	Einzelne zur Unterlegung der sozialmedizinischen Beurteilung indizierte Untersuchungen fehlen und erschweren die Beurteilung.
F2G	Wesentliche Untersuchungen zur Beantwortung der sozialmedizinischen Fragestellung fehlen, was zusätzlichen Bearbeitungsaufwand nach sich zieht.

Anhang 7:

Prüffragenübergreifende Codes der qualitativen Inhaltsanalyse der Einzelkriterien

Prüffragenübergreifende Codes „*Neue Aspekte*“

Code	Codebeschreibung
X1	Die freitextliche Begründung enthält einen neuen Gesichtspunkt, der über die Ausprägung der Mängelkategorien hinaus geht (z.B. Generalisierung / Pauschalisierung der Epikrise).
X2	Die Begründung ist ausführlich und enthält mehr Aspekte als ursprünglich von der Prüffrage zu erwarten sind.

Prüffragenübergreifende Codes „*Keine Beantwortung der Prüffrage*“

Code	Codebeschreibung
X3	Ein erkannter Mangel wird über die in der Bewertungsmatrix hinterlegte Mängeldefinition nicht abgedeckt, ist aber im erläuternden Text der Prüffrage aufgeführt.
X4	Die Mängelbegründung verstößt gegen eine definierte Einschränkung des Geltungsbereichs einer Prüffrage.
X5	Einer Bewertungsmatrix liegt für die Graduierung der Mängelstufen eine Strukturierungsdimension zugrunde, z.B. Prüffragen bei denen die sozialmedizinische Relevanz (C3) oder die Häufigkeit eines dokumentierten Mangels (D2) ausschlaggebend für die Beurteilung ist. Die freitextliche Mängelbegründung ist nicht eindeutig der jeweiligen Mängelstufe zuzuordnen, da ein konkreter Verweis auf die Graduierungsdimension fehlt.
X6	Die freitextliche Begründung zur Erklärung des Qualitätsdefizits beantwortet die Prüffrage nicht.
X7	Obwohl eine der drei definierten Mängelkategorien vergeben wurde, wurden die festgestellten Qualitätsdefizite nicht dokumentiert.
X8	Es wird als Begründung lediglich auf eine andere Kategorie verwiesen (Bsp.: siehe B2).
X9	Die Begründung Antwort entspricht der Mängeldefinitionen einer anderen Prüffrage.

Prüffragenübergreifende Codes „*Unvollständig Beantwortung*“

Code	Codebeschreibung
X10	Die Beantwortung der Prüffrage wurde in einem sehr hohen Detaillierungsgrad vorgenommen. Die Beantwortung der Prüffrage im Sinne des Manuals geht dabei allerdings verloren.
X11	Die Beantwortung der Frage wurde auf einem zu niedrigen Detaillierungsgrad vorgenommen. Festgestellte Mängel werden nicht konkret aus dem Gutachten belegt.
X12	Es werden Aspekte zur Begründung herangezogen, die über die Einzelkriterien hinausgehen und sich eher auf die Kriterien beziehen, die in der Ampel beurteilt werden sollen.

X13	Die Mängelbegründung ist ohne Kontext nicht nachvollziehbar.
X14	Ein Qualitätsdefizit wird erkannt und bemängelt. Dieser Mangel scheint jedoch nicht in die Wahl der Mängelstufe eingeflossen zu sein.

Prüffragenübergreifende Codes „Indirekte Beantwortung“

Code	Codebeschreibung
X15	Aus der gegebenen Begründung kann nur indirekt der Mangel und die Mangelkategorie abgeleitet werden. Die eigentliche Begründung ergibt sich aus einem logischen Schluss aus der gegebenen Begründung.
X16	Die Prüffrage kann indirekt anhand eines Beispiels und der angegebenen Mängelkategorie beantwortet werden.

Prüffragenübergreifende Codes „Kritik“

Code	Codebeschreibung
X17	Es wird eine allgemeine Kritik am Manual zum Peer Review-Verfahren zur sozialmedizinischen Begutachtung gegeben.
X18	Es wird eine allgemeine Kritik am Gutachten bzw. Vorgehen des Sozialmediziners im Gutachten gegeben.
X19	Es wird eine allgemeine Kritik an der Durchführung der für das Peer-Review notwendigen Verfahrensabläufe geäußert, z.B. Anonymisierungsfehler.

Prüffragenübergreifende Codes „Positiv“

Code	Codebeschreibung
X20	Angabe einer nicht erforderlichen freitextlichen Begründung, z.B. Anmerkungen der Peer-Gutachter

Anhang 8: Codes der qualitativen Inhaltsanalyse des übergeordneten Kriteriums

Prüffragenspezifische Codes des übergeordneten Kriteriums

Code	Codebeschreibung
ÜK1	Es wird verallgemeinernd ein Bruch in der Argumentationskette seitens des Peers angemerkt, jedoch nicht weiter ausdifferenziert, an welcher Stelle dieser zu finden ist.
ÜK2	Es wird ein Bruch in der Argumentationskette in Form einer unvollständigen Anamneseerhebung beschrieben.
ÜK3	Es wird ein Bruch in der Argumentationskette in Form einer fehlenden Darstellung eigener Untersuchungsergebnisse beschrieben.
ÜK4	Es wird ein Bruch in der Argumentationskette in Form von fehlenden Befunden beschrieben.
ÜK5	Es wird ein Bruch in der Argumentationskette in Form einer fehlenden Befundbewertung seitens des Gutachters beschrieben.
ÜK6	Es wird ein Bruch in der Argumentationskette in Form einer unvollständigen Epikrise beschrieben.
ÜK7	Es wird ein Bruch in der Argumentationskette in Form fehlender oder unvollständiger Angaben zur Funktionsfähigkeit beschrieben.
ÜK8	Es wird ein Bruch in der Argumentationskette in Form fehlender oder unvollständiger Angaben zur Leistungsfähigkeit beschrieben.
ÜK9	Es wird ein Bruch in der Argumentationskette in Form fehlender oder unvollständiger Angaben zur qualitativen Leistungsfähigkeit beschrieben.
ÜK10	Es wird ein Bruch in der Argumentationskette in Form fehlender oder unvollständiger Angaben zur quantitativen Leistungsfähigkeit beschrieben.
ÜK11	Es wird ein Bruch in der Argumentationskette in Form von fehlender oder unvollständiger Angaben zu therapeutischen Maßnahmen beschrieben
ÜK12	Es wird ein Bruch in der Argumentationskette in Form einer fehlenden oder falschen seitens des Gutachters beschrieben
ÜK13	Im Gutachten befinden sich Diskrepanzen und Widersprüche zwischen den geschilderten Beschwerden und der allg. Leistungsbeurteilung, was zu Brüchen in der Argumentationskette führt.
ÜK14	Es fehlt ein Zusatzgutachten im Anhang, das essenziell zum Verständnis des vorliegenden sozialmedizinischen Gutachtens ist, was zu Brüchen in der Argumentationskette führt.
ÜK15	Der Aufbau des Gutachtens ist fehlerhaft, infolgedessen die Struktur des Gutachtens defizitär ist, was zu Brüchen in der Argumentationskette führt.
ÜK16	Es wird verallgemeinernd dargelegt, dass das Gutachten undifferenziert sei, was zu Brüchen in der Argumentationskette führt. Ein Hinweis auf konkrete Textpassagen im Gutachten wird nicht gegeben
ÜK17	Es wird verallgemeinernd dargelegt, dass das Gutachten unvollständig sei, was zu Brüchen in der Argumentationskette führt. Ein Hinweis auf konkrete Textpassagen im Gutachten wird nicht gegeben

ÜK18	Die Begründung hat keinen Bezug zum Review-Verfahren und wird als irrelevant eingestuft. Ein Rückschluss auf Brüche in der Argumentationsketten kann nicht gezogen werden.
ÜK19	Es wird als Begründung lediglich auf eine andere Prüffrage verwiesen (Bsp.: siehe E1).
ÜK20	Angabe einer nicht erforderlichen freitextlichen Begründung bei der Codierung von „grün“
ÜK21	Obwohl eine „gelbe“ oder „rote“ Ampelbewertung vergeben wurde, wurden die festgestellten Qualitätsdefizite nicht dokumentiert.

Anhang 9:

Prüffragenkatalog des Manuals zum Peer Review-Verfahren 2013

Qualitätssicherung der sozialmedizinischen Begutachtung

Manual zum Peer Review-Verfahren

28. Januar 2013



Erstet unter wissenschaftlicher Begleitung durch die Forschungsgruppe Dr. Vogel
Arbeitsbereich Rehabilitationswissenschaften, Institut für Psychotherapie und
Medizinische Psychologie, Universität Würzburg

A. Einzelkriterium „Formale Gestaltung“

Das Einzelkriterium „Formale Gestaltung“ untergliedert sich in zwei Prüffragen. Das Kriterium erfordert die adäquate Berücksichtigung des Anforderungsprofils für das sozialmedizinische Gutachten. So sind die formalen Anforderungen an das Gutachten gemäß den Gliederungsvorgaben der DRV-Schrift Band 21 einzuhalten. Des Weiteren ist der einheitliche Formularsatz der Deutschen Rentenversicherung, bestehend aus Deckblatt und Schlussblatt Teil 1 und Teil 2, zu verwenden.

Die Prüfung der Qualität bezieht sich in den folgenden zwei Prüffragen ausschließlich auf rein formale Aspekte. Inhaltliche Vollständigkeit und Nachvollziehbarkeit der Darstellung sind hier nicht qualitätsrelevant; sie werden an anderer Stelle geprüft.

Einzelkriterium A: Formale Gestaltung

Prüffrage A1

A1: Inwieweit werden die Gliederungsvorgaben gemäß den Anforderungen eingehalten?

Folgende Überschriften, die derzeit im DRV Bd. 21 dargelegt sind, müssen vorhanden sein:

- Anamnese
- Untersuchungsbefunde
- Diagnosen
- Epikrise
- Sozialmedizinische Leistungsbeurteilung

Die Überschriften sind im freien Teil des Gutachtens aufzuführen und klar voneinander abzugrenzen.

In Abgrenzung zu den Prüffragen des Einzelkriteriums "Vollständigkeit" geht es hier um das rein formale Vorhandensein der Gliederungspunkte. Inhaltliche Aspekte, wie z. B. Art, Umfang, Angemessenheit oder Vollständigkeit der Gliederungspunkte sind hier nicht Gegenstand der Bewertung.

Leichte Mängel

Die Reihenfolge der Gliederungspunkte wird nicht eingehalten.

Deutliche Mängel

Eine Überschrift fehlt.

Gravierende Mängel

Mehr als eine Überschrift fehlt.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren

Einzelkriterium A: Formale Gestaltung

Prüffrage A2

A2: Inwieweit wird der einheitliche Formularsatz bestehend aus Deck- und Schlussblatt verwendet?

Bei der Beurteilung der Prüffragen dieses Einzelkriteriums ist zu berücksichtigen, inwieweit die Vorgaben der DRV-Schrift 21 eingehalten werden.

An dieser Stelle wird geprüft, ob das einheitliche Deck- und Schlussblatt verwendet wird. Es wird nicht geprüft, ob die Aussagen des Deck- und Schlussblattes an anderer Stelle im Gutachten vorhanden sind bzw. ob das Gutachten bei fehlendem Deck- oder Schlussblatt verwertbar ist.

Die Frage entfällt bei Gutachten mit Auslandsberührung (z. B. E213).

Leichte Mängel

Das Deckblatt ist beigelegt. Es ist aber gegenüber dem einheitlichen Formularsatz abgewandelt, alle wesentlichen Elemente sind vorhanden.

ODER

Das Schlussblatt ist beigelegt. Es ist aber gegenüber dem einheitlichen Formularsatz abgewandelt, alle wesentlichen Elemente sind vorhanden.

Deutliche Mängel

Es fehlt das Deckblatt des einheitlichen Formularsatzes.

ODER

Es fehlt das Schlussblatt des einheitlichen Formularsatzes.

Gravierende Mängel

Weder das Deckblatt noch das Schlussblatt des einheitlichen Formularsatzes werden verwendet.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren

B. Einzelkriterium „Verständlichkeit“

Das Einzelkriterium „*Verständlichkeit*“ untergliedert sich in vier Prüffragen. Geprüft wird die sprachliche Darstellung des Gutachtens sowie der Umgang mit sozialrechtlichen Begriffen und Aussagen. Im Gutachten wird eine allgemeinverständliche und prägnante Sprache erwartet. Fachsprachliche Abkürzungen sind in den Diagnosen und der Epikrise möglichst zu vermeiden bzw. allgemeinsprachlich zu übersetzen. Die sozialmedizinische Terminologie ist richtig anzuwenden. Sozialrechtliche Ausführungen/Folgerungen sind zu unterlassen.

Einzelkriterium B: Verständlichkeit Prüffrage B1	
<i>B1: Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich?</i> Die Frage bezieht sich auf das gesamte Gutachten. Sprache, Grammatik und Satzkonstruktion sind zu beurteilen, nicht jedoch der persönliche Schreibstil des Gutachters. Gefordert ist eine knappe, prägnante Ausdrucksweise. Unnötige Längen, z. B. durch reine Wiederholung von bereits beschriebenen Sachverhalten oder ausführliche Diskussionen irrelevanter Befunde, sind zu vermeiden. Zentral ist die Verständlichkeit der Aussagen.	
Leichte Mängel	Das Gutachten enthält unnötige Längen. ODER Im Gutachten werden unnötig komplizierte Formulierungen verwendet.
Deutliche Mängel	Durch die sprachliche Ausdrucksweise sind Aussagen vereinzelt nicht eindeutig bzw. nicht verständlich.
Gravierende Mängel	Durch die sprachliche Ausdrucksweise sind Aussagen überwiegend nicht eindeutig bzw. nicht verständlich.
Welche Mängel wurden festgestellt? Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren	

Einzelkriterium B: Verständlichkeit Prüffrage B2	
<i>B2: Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt?</i> Die Frage bezieht sich nur auf Diagnose und Epikrise. Erklärungsbedürftig sind Fachausdrücke und Abkürzungen, welche von fachkundigen Laien nicht sicher vorausgesetzt werden können (z. B. Spondylolisthesis, PTBS, FPI). Ausnahmen sind allgemeinverständliche Fachausdrücke und Abkürzungen (z. B. EKG, HIV). Nicht allgemein verständliche Abkürzungen sind mindestens einmal auszuschreiben bzw. zu erklären. Weiterhin sind Abkürzungen, die für unterschiedliche medizinische Begrifflichkeiten verwendet werden (z. B. HWI: Hinterwandinfarkt, Harnwegsinfekt; BAA: Bauchaortenaneurysma, bradykarde absolute Arrhythmie) zu vermeiden.	
Leichte Mängel	Eine Abkürzung wird nicht mindestens einmal ausgeschrieben. ODER Es erfolgt keine Erläuterung eines nicht allgemein verständlichen Fachbegriffs.
Deutliche Mängel	Mehrfach werden Abkürzungen nicht mindestens einmal ausgeschrieben ODER Es erfolgt mehrfach keine Erläuterung nicht allgemein verständlicher medizinische Fachbegriffe.
Gravierende Mängel	Überwiegend werden Abkürzungen nicht mindestens einmal ausgeschrieben ODER Es erfolgt überwiegend keine Erläuterung nicht allgemein verständlicher Fachbegriffe.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

Einzelkriterium B: Verständlichkeit

Prüffrage B3

B3: Inwieweit wird die sozialmedizinische Terminologie richtig angewendet?

Die Aufgabe des Gutachters ist es, die persönlichen medizinischen Voraussetzungen für Sozialleistungen zu beschreiben, ohne daraus sozialrechtliche Folgerungen zu ziehen. Es wird erwartet, dass der Gutachter mit sozialmedizinisch relevanten Begrifflichkeiten korrekt umzugehen weiß.

Geprüft wird die richtige Anwendung sozialmedizinischer Termini. Sie sind gemäß den Definitionen im Sozialmedizinischen Glossar der Deutschen Rentenversicherung (entsprechend DRV-Schrift Band 81) anzuwenden.

Beispiel für eine nicht richtige Anwendung sozialmedizinischer Termini: dreimal „zeitweise“ für die Arbeitshaltung (dies ist als deutlicher Mangel zu kodieren).

Beispiele für (nicht zulässige) sozialrechtliche Folgerungen: Der Sozialmediziner äußert sich zur Wegefähigkeit statt (nur) zur Gehstrecke, stellt eine Erwerbsminderung fest oder empfiehlt eine Erwerbsminderungsrente.

Leichte Mängel

Sozialmedizinische Termini werden nicht immer korrekt angewendet.

ABER

Dies lässt keine missverständlichen oder falschen Schlussfolgerungen zu.

Deutliche Mängel

Sozialmedizinische Termini werden nicht immer korrekt angewendet.

UND

Dies lässt missverständliche Schlussfolgerungen zu.

Gravierende Mängel

Im Gutachten finden sich mindestens eine sozialrechtliche Folgerung des Gutachters

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren

C. Einzelkriterium „Transparenz“

Das Einzelkriterium „*Transparenz*“ untergliedert sich in drei Prüffragen. Erfasst wird, inwieweit die verwendeten Verfahren, Methoden und Instrumente beschrieben wurden und die Kriterien zur Beurteilung der Befunde im Gutachten dargelegt sind. Die Herkunft von anamnestischen Informationen und Befunden muss erkennbar sein.

Einzelkriterium C: Transparenz

Prüffrage C1

C1: Inwieweit ist die Herkunft von medizinischen Informationen erkennbar?

Alle im Gutachten erwähnten medizinischen Informationen sind zu dokumentieren. Dabei wird erwartet, dass Herkunft (Quelle), zeitliche Einordnung (Datum) und Kernaussagen der vorliegenden medizinischen Informationen erkennbar sind

Medizinische Informationen sind:

- Angaben, die der Versicherte selbst gemacht hat
- Angaben von dritten Personen
- Ärztliche Berichte
- Krankenhausentlassungsberichte
- Reha-Entlassungsberichte
- Gutachten der Rentenversicherungsträger, Gutachten im Klageverfahren
- Gutachten anderer Sozialleistungsträger (z. B. Agentur für Arbeit, Medizinischer Dienst der Krankenkassen, Berufsgenossenschaft, Versorgungsamt), Betreuungsgutachten
- Selbst erhobene und in Auftrag gegebene Befunde
- Vor- bzw. Fremdbefunde (z. B. vom Versicherten mitgebrachter Koronarangiographie-Befund durch vorbehandelnden Kardiologen).

Als Kernaussage von Vorgutachten ist insbesondere die sozialmedizinische Beurteilung zu berücksichtigen. Die Angabe "Siehe Befundbericht Dr. XXX" ist nicht ausreichend.

Leichte Mängel

Herkunft, zeitliche Einordnung oder Kernaussagen der medizinischen Informationen fehlen vereinzelt.

Deutliche Mängel

Herkunft, zeitliche Einordnung oder Kernaussagen fehlen bei mehreren medizinischen Informationen.

Gravierende Mängel

Herkunft, zeitliche Einordnung oder Kernaussagen der medizinischen Informationen fehlen überwiegend.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

Einzelkriterium C: Transparenz Prüffrage C2	
<i>C2: Inwieweit wird dargestellt, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben werden?</i> Die Prüffrage bezieht sich auf die selbst erhobenen Befunde. Erwartet wird, dass bei selbst erhobenen Befunden die angewendeten Verfahren, Instrumente und Methoden einschließlich der äußeren Rahmenbedingungen dargelegt werden. Beispielsweise ist anzugeben, ob eine Ergometrie im Liegen oder in sitzender Position durchgeführt wurde oder Röntgenaufnahmen im Liegen oder Stehen etc. Auch sind die Rahmenbedingungen der testpsychologischen Untersuchungen zu benennen (z. B. welche Skala erhoben wurde und die Bedingungen der Untersuchungsdurchführung).	
Leichte Mängel	Für einen sozialmedizinisch nicht relevanten Befund fehlt die Darstellung der Verfahren, Methoden und Instrumente einschließlich der äußeren Rahmenbedingungen.
Deutliche Mängel	Für mehrere sozialmedizinisch nicht relevante Befunde fehlt die Darstellung der Verfahren, Methoden und Instrumente einschließlich der äußeren Rahmenbedingungen.
Gravierende Mängel	Für mindestens einen sozialmedizinisch relevanten Befund fehlt die Darstellung der Verfahren, Methoden und Instrumente einschließlich der äußeren Rahmenbedingungen.
Welche Mängel wurden festgestellt? Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.	

Einzelkriterium C: Transparenz

Prüffrage C3

C3: Inwieweit wird dargestellt, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind?

Es wird erwartet, dass bei selbst erhobenen apparativ-technischen Untersuchungen einschließlich psychometrischer Testverfahren die erhobenen Messwerte und Referenzbereiche aufgeführt werden.

Unauffällige Befundergebnisse sind als unauffällig zu kennzeichnen. Bei pathologischen Ausprägungen sind Laborwerte mit Referenzbereichen exakt anzugeben. Bei diesen Laborwerten reicht die Kenntlichmachung +/- (im Sinne von über bzw. unter dem Referenzwert) nicht aus.

Hinweis: Die Neutral-Null-Methode ist zwar keine apparativ-technische Untersuchung, jedoch ist die Angabe von Referenzbereichen grundsätzlich für alle erhobenen Messwerte zu fordern.

Leichte Mängel

Es fehlt der exakte Wert/Referenzbereich für einen sozialmedizinisch nicht relevanten Messwert.

Deutliche Mängel

Es fehlen exakte Werte/Referenzbereiche für mehrere sozialmedizinisch nicht relevante Messwerte.

Gravierende Mängel

Es fehlt der exakte Wert/Referenzbereich für mindestens einen sozialmedizinisch relevanten Messwert.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

D. Einzelkriterium „Vollständigkeit“

as Einzelkriterium „Vollständigkeit“ untergliedert sich in fünf Prüffragen. Dabei ist zu prüfen, inwieweit Anamnese, Befunde, Diagnosen und Epikrise vollständig dargestellt und die sozialmedizinischen Fragestellungen vollständig beantwortet sind. Hier ist die Bewertung der Vollständigkeit, nicht aber die der Richtigkeit der Beantwortung der sozialmedizinischen Fragestellung gefordert.

Die dem Gutachten zugrunde liegenden Befunde sind vom Gutachter ausreichend zu dokumentieren. Befunde, die für die sozialmedizinische Beurteilung eine untergeordnete Rolle spielen, können als nicht relevant gekennzeichnet werden; für diese ist dann eine weitergehende Auseinandersetzung nicht erforderlich.

Der Gutachter muss sich auch mit den vom Versicherten geltend gemachten Beschwerden im Gutachten auseinandersetzen. Bei umfangreichen Beschwerdeschilderungen können diese gegebenenfalls in einer zusammenfassenden Würdigung dargestellt werden. Auch hat vom Gutachter eine Stellungnahme zu bisherigen Therapien des Antragstellers und zu weiteren therapeutischen Optionen zu erfolgen. Wesentliche Abweichungen in der sozialmedizinischen Beurteilung gegenüber Vorgutachten sind darzulegen.

Einzelkriterium D: Vollständigkeit

Prüffrage D1

D1: Inwieweit wird die Anamnese vollständig dargestellt?

Für die Darstellung der Anamnese werden folgende Inhalte erwartet:
Anamnese medizinischer Daten

- Eigenanamnese (ggf. mit Darstellung der Tagesstruktur, soweit für die sozialmedizinische Leistungsbeurteilung relevant)
- Bei psychiatrischen Gutachten: Biographische Anamnese, soweit störungsbezogen für die sozialmedizinische Beurteilung relevant
- Jetzige Beschwerden (Beschwerden, mit denen eine Leistungsminderung begründet wird)
- Therapie
- Fachrichtung der behandelnden Ärzte, sofern nicht auf dem Formblatt aufgeführt

Arbeits- und Sozialanamnese

- Berufs- und Arbeitsanamnese
- Familiäre Situation
- Weitere soziale Kontextfaktoren (z. B. Wohnsituation), Umstände der Antragstellung, bisherige Sozialleistungen

Leichte Mängel

Einer oder mehrere der folgenden drei Bestandteile der Anamnese fehlen:

- Fachrichtung der behandelnden Ärzte
- Familiäre Situation
- Weitere soziale Kontextfaktoren (z. B. Wohnsituation), Umstände der Antragstellung, bisherige Sozialleistungen

Deutliche Mängel

Es fehlt maximal einer der folgenden Bestandteile der Anamnese:

- Eigenanamnese
- Biographische Anamnese bei psychiatrischen Gutachten, soweit störungsbezogen für die sozialmedizinische Beurteilung relevant
- Jetzige Beschwerden
- Therapie
- Berufs- und Arbeitsanamnese

Gravierende Mängel	Es fehlen mehrere der folgenden Bestandteile der Anamnese: - Eigenanamnese - Biographische Anamnese bei psychiatrischen Gutachten, soweit störungsbezogen für die sozialmedizinische Beurteilung relevant - Jetzige Beschwerden - Therapie - Berufs- und Arbeitsanamnese
Welche Mängel wurden festgestellt? Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.	

Einzelkriterium D: Vollständigkeit

Prüffrage D2

D2: Inwieweit werden die Diagnosen dargestellt und mit Funktions-einschränkungen (sog. Funktionsdiagnosen) formuliert?

Entsprechend den Vorgaben der DRV-Schrift Band 21 sind die Diagnosen

- sprachlich zu formulieren (z. B. Asthma bronchiale) und nach der jeweils gültigen ICD zu verschlüsseln
- gesondert im freien Text aufzulisten, die ersten drei sozialmedizinisch relevanten Diagnosen zusätzlich im Schlussblatt
- mit Funktionseinschränkungen anzugeben (z. B. medikamentös eingestellter Diabetes mellitus Typ 2 ohne wesentliche Folgeschäden, z. B. chronisch obstruktive Lungenerkrankung mit dauerhaft schwerwiegender Einschränkung der Atemfunktion)
- nach Schweregrad (ggf. mit Klassifikation und/oder Stadieneinteilung) ihrer funktionellen Einschränkungen zu ordnen

Die ICD Verschlüsselung ist - sofern möglich - vierstellig (einschließlich Stelle nach dem Punkt) z. B. J45.1 gefordert. Die Verschlüsselung auf dem Schlussblatt ist dazu ausreichend.

Inwieweit eine Diagnose aus der Anamnese bzw. der klinischen Untersuchung ableitbar ist, soll hier nicht bewertet werden. Es ist auch nicht zu bewerten, ob die Diagnose gesichert ist.

Leichte Mängel

Bei mindestens einer der ersten drei Diagnosen fehlt die ICD-Verschlüsselung

Deutliche Mängel

Die sprachliche Formulierung der Diagnosen (z. B. Asthma bronchiale) fehlt.

ODER

Die Funktionseinschränkungen sind nicht benannt.

Gravierende Mängel

Die sprachliche Formulierung der Diagnosen (z. B. Asthma bronchiale) fehlt.

UND

Die Funktionseinschränkungen sind nicht benannt.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

Einzelkriterium D: Vollständigkeit Prüffrage D3	
<i>D3: Inwiefern erfolgt eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen?</i> Erwartet wird in der Epikrise eine Stellungnahme zu bisherigen Behandlungen und gegebenenfalls zu Therapien, die geeignet sind, die Leistungsfähigkeit deutlich zu steigern. Ist ersichtlich, ob ein Behandlungsfall vorliegt?	
Leichte Mängel	Es fehlt eine Stellungnahme zu bisherigen Behandlungen.
Deutliche Mängel	Es fehlt eine Stellungnahme zu Therapien, die geeignet sind, die Leistungsfähigkeit deutlich zu steigern.
Gravierende Mängel	Es fehlt eine Stellungnahme zu bisherigen Behandlungen. UND Es fehlt eine Stellungnahme zu Therapien, die geeignet sind, die Leistungsfähigkeit deutlich zu steigern.
Welche Mängel wurden festgestellt? Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.	

Einzelkriterium D: Vollständigkeit
Prüffrage D4
D4: Inwieweit werden wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten dargelegt?

Hierbei geht es um abweichende Aussagen zur Leistungsfähigkeit, die eine andere sozialrechtliche Entscheidung nach sich ziehen könnten. Es wird erwartet, dass Abweichungen hinsichtlich des quantitativen und qualitativen Leistungsvermögens, von Beginn und Dauer der Leistungsminderung oder des Reha-Bedarfs in der Epikrise aufgezeigt und bewertet werden.

Unter Vorgutachten wird dabei verstanden: Reha-Entlassungsbericht, Gutachten der Rentenversicherungsträger, Gutachten im Klageverfahren, Gutachten anderer Sozialleistungsträger (z. B. der Agentur für Arbeit, des Medizinischen Dienstes Der Krankenkassen, der Berufsgenossenschaft, des Versorgungsamtes), Betreuungsgutachten.

Liegen aus Vorgutachten keine Aussagen zur Leistungsfähigkeit vor, so ist kein Mangel zu kodieren.

Leichte Mängel

Es gibt wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten. Diese werden im Gutachten aufgezeigt, es unterbleibt jedoch eine Bewertung von Teilaspekten der Abweichungen in der Epikrise.

Deutliche Mängel

Es gibt wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten. Diese werden im Gutachten aufgezeigt, es unterbleibt jedoch eine Bewertung dieser Abweichungen in der Epikrise.

Gravierende Mängel

Es gibt wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten. Die Abweichungen werden in der Epikrise nicht aufgezeigt.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

inzelkriterium D: Vollständigkeit

Prüffrage D5

D5: Inwieweit werden die sozialmedizinischen Fragestellungen vollständig beantwortet?

Fragestellungen an das sozialmedizinische Gutachten der Deutschen Rentenversicherung sind formuliert im bundesweit konsentierten Schlussblatt zum Gutachten (siehe DRV Schrift 21):

- quantitatives Leistungsvermögen
- qualitatives Leistungsvermögen
- Beginn/Dauer einer Leistungsminderung
- Reha-Bedarf.

Darüber hinaus können sich Fragestellungen aus den Angaben des Versicherten oder aus den Befunden und Diagnosen im Einzelfall ergeben: z. B. Gehfähigkeit, Gebrauchsfähigkeit beider Hände, zusätzliche Arbeitsunterbrechungen.

Erwartet werden zu jeder der vier Fragestellungen an das Gutachten Aussagen im freien Teil des Gutachtens und die entsprechende Dokumentation auf dem Schlussblatt. Es reicht nicht aus, die getroffenen Feststellungen allein durch Ankreuzen auf dem Schlussblatt zu dokumentieren.

Bei einer Leistungsfähigkeit im Erwerbsleben von unter drei Stunden ist es nicht erforderlich, die qualitative Leistungsfähigkeit im Erwerbsleben darzustellen.

Im Peer Review-Verfahren wird nicht geprüft, ob eine Beurteilung der Leistungsfähigkeit in Bezug auf die zuletzt ausgeübte Tätigkeit durch den Gutachter erfolgt ist.

Leichte Mängel	Es fehlen Aussagen zum Reha-Bedarf oder es bestehen Diskrepanzen zwischen Aussagen zum Reha-Bedarf in Freitext und Schlussblatt
Deutliche Mängel	Es fehlen Aussagen zu Beginn/Dauer einer Leistungsminderung oder es bestehen Diskrepanzen zwischen Aussagen zu Beginn/Dauer einer Leistungsminderung in Freitext und Schlussblatt.
Gravierende Mängel	Es fehlen Aussagen zur Leistungsfähigkeit (qualitativ/quantitativ) oder es bestehen Diskrepanzen zwischen Aussagen zur Leistungsfähigkeit (qualitativ/quantitativ) in Freitext und Schlussblatt

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

E. Einzelkriterium „Medizinisch-wissenschaftliche Grundlagen“

Das Einzelkriterium „*Medizinisch-wissenschaftliche Grundlagen*“ untergliedert sich in zwei Prüffragen. Zu prüfen ist, inwieweit der allgemein anerkannte Stand medizinischer Erkenntnisse eingehalten und inwieweit die Begutachtungsliteratur der Deutschen Rentenversicherung berücksichtigt werden.

Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen Prüffrage E1	
E1: Inwieweit wird der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet? Geprüft wird, ob bei der Gewinnung der medizinischen Informationen, der Prognose und den Therapieoptionen der allgemein anerkannte Stand medizinischer Erkenntnisse eingehalten wird. Außenseitermethoden entsprechen nicht dem allgemein anerkannten Stand medizinischer Erkenntnisse. Die Schlüssigkeit der Diagnose und deren Ableitung werden hier nicht geprüft.	
Leichte Mängel	Fallbezogen werden Außenseitermethoden verwendet, jedoch nicht zur sozialmedizinischen Beurteilung herangezogen.
Deutliche Mängel	Außenseitermethoden werden zusätzlich zur Herleitung der sozialmedizinischen Beurteilung herangezogen.
Gravierende Mängel	Die sozialmedizinische Beurteilung wird mit Außenseitermethoden begründet.
Welche Mängel wurden festgestellt? Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren	

Einzelkriterium E: Medizinisch-wissenschaftliche Grundlagen

Prüffrage E2

E2: Inwieweit wird die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet?

Geprüft wird, inwieweit grobe Abweichungen von der Begutachtungsliteratur zu erkennen sind. Hier ist die von der Rentenversicherung veröffentlichte sozialmedizinische Begutachtungsliteratur gemeint, insbesondere die Buchveröffentlichung „Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung“ und die „Leitlinien für die sozialmedizinische Begutachtung“ der Deutschen Rentenversicherung (veröffentlicht auf den DRV-Internetseiten).

Bei Mängeln ist vom Peer ein konkreter Bezug zur Begutachtungsliteratur herzustellen, z. B. indem die Leitlinie oder das (Kapitel im) Begutachtungsbuch benannt werden.

Begründete Abweichungen des Gutachters von der Begutachtungsliteratur führen nicht zu einer Mängelbewertung. Die Richtigkeit der Begründung ist nicht Gegenstand der Prüfung.

Leichte Mängel

Es liegt eine Abweichung von der Begutachtungsliteratur vor, die keine Auswirkung auf die sozialmedizinische Beurteilung hat.

Deutliche Mängel

Es liegt mehr als eine Abweichung von der Begutachtungsliteratur vor, die keine Auswirkung auf die sozialmedizinische Beurteilung haben.

Gravierende Mängel

Es liegt mindestens eine Abweichung von der Begutachtungsliteratur vor, die Auswirkungen auf die sozialmedizinische Beurteilung hat.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

F. Einzelkriterium „Wirtschaftlichkeit“

Das Einzelkriterium „*Wirtschaftlichkeit*“ umfasst zwei Prüffragen. Diese beziehen sich auf die Diagnostik und orientieren sich am Wirtschaftlichkeitsgebot des SGB V. Verfahren, Methoden und Instrumente sind dementsprechend auszurichten. Der Peer prüft, ob ein Gutachter Sinn und Zweck einer Untersuchung ersichtlich gemacht hat und in der Epikrise auch Schlüsse daraus zieht.

Einzelkriterium F: Wirtschaftlichkeit

Prüffrage F1

F1: Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig?

Erwartet wird, dass die vom Gutachter eingesetzte medizinisch-technische Diagnostik und eventuelle testpsychologische Untersuchungen zur Beantwortung der Fragestellung geeignet sind und das Maß des Notwendigen nicht überschreiten. Dabei ist zu prüfen, inwieweit die vom Gutachter eingesetzte Diagnostik tatsächlich erforderlich ist, um zu einer sozialmedizinischen Beurteilung zu kommen, z. B. wird bei Vorliegen aktueller Fremdbefunde auf Diagnostik verzichtet. Diagnosesicherung und Differentialdiagnostik sind nicht vorrangig Aufgabe der sozialmedizinischen Begutachtung.

Hierbei sind auch die Vorgaben der aktuellen Strahlenschutzverordnung zu beachten.

Ferner ist zu berücksichtigen, dass manche Laboruntersuchungen in Kombination (z. B. Leberfunktionsparameter im Paket) nicht teurer als mehrere einzelne Untersuchungen sind.

Die Indikation für die eingesetzte Diagnostik muss sich aus dem Gutachten ergeben.

Leichte Mängel

Die vom Gutachter eingesetzte Diagnostik ist für die Beantwortung der sozialmedizinischen Fragestellung geeignet. Vereinzelt wird das Maß des Notwendigen überschritten.

Deutliche Mängel

Die vom Gutachter eingesetzte Diagnostik ist für die Beantwortung der sozialmedizinischen Fragestellung geeignet. Mehrfach wird das Maß des Notwendigen überschritten.

Gravierende Mängel

Die vom Gutachter eingesetzte Diagnostik ist für die Beantwortung der sozialmedizinischen Fragestellung nicht geeignet und damit auch nicht notwendig.

Welche Mängel wurden festgestellt?

Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

Einzelkriterium F: Wirtschaftlichkeit Prüffrage F2	
F2: Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend? Erwartet wird, dass die eingesetzte medizinisch-technische Diagnostik und eventuelle testpsychologische Untersuchungen zur Beantwortung der Fragestellung ausreichend sind. Bei nicht ausreichenden Vorbefunden ist eine eigene Diagnostik durch den Gutachter einzusetzen. Unter Wirtschaftlichkeitsaspekten kann auch das Fehlen von Diagnostik relevant sein, wenn dadurch zusätzlicher Bearbeitungsaufwand, weitere Diagnostik oder Gutachten notwendig werden.	
Leichte Mängel	Eine zur Stützung der sozialmedizinischen Beurteilung indizierte Untersuchung fehlt (z. B. fehlende Blutgasanalyse bei COPD). Zusätzlicher Bearbeitungsaufwand ist jedoch nicht notwendig.
Deutliche Mängel	Einzelne zur Unterlegung der sozialmedizinischen Beurteilung indizierte Untersuchungen fehlen und erschweren die Beurteilung (z. B. bei COPD nur Untersuchung der Lungenfunktion (LuFu) durchgeführt, Bodyplethysmographie fehlt).
Gravierende Mängel	Es fehlen wesentliche Untersuchungen, die zur Beantwortung der sozialmedizinischen Fragestellungen erforderlich gewesen wären. Dies zieht zusätzlichen Bearbeitungsaufwand nach sich (z. B. bei COPD nur körperliche Untersuchung durchgeführt, Lungenfunktion (LuFu) und Bodyplethysmographie fehlen).
Welche Mängel wurden festgestellt? Die Bewertung ist ausschließlich auf den erfragten Prüfinhalt der Prüffrage zu beziehen und in den angegebenen Bewertungskategorien vorzunehmen. Wenn keine der obigen drei Mängelkategorien zutrifft, ist die Kategorie „kein Mangel“ zu kodieren. Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.	

Übergeordnetes Kriterium „Nachvollziehbarkeit des Gutachtens“

Im übergeordneten Qualitätskriterium wird das gesamte Gutachten im Zusammenhang betrachtet und die Nachvollziehbarkeit des Gutachtens anhand der Argumentationskette bewertet.

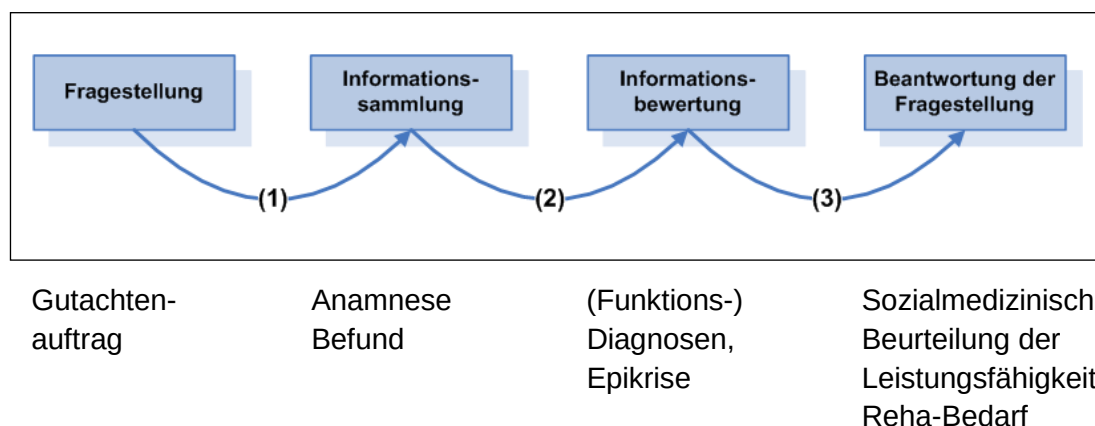
Übergeordnetes Kriterium: Nachvollziehbarkeit des Gutachtens

Ampel

Bewerten Sie die Nachvollziehbarkeit des Gutachtens anhand der Argumentationsführung.

Beim übergeordneten Kriterium wird das gesamte Gutachten im Zusammenhang betrachtet und die Nachvollziehbarkeit des Gutachtens anhand der Argumentationskette (s. Abb.) bewertet.

Module der Argumentationskette



Der Fragestellung schließt sich in der Argumentationskette die Informationssammlung an. Zu dieser Informationssammlung gehören die Anamnese (medizinische Anamnese, Arbeits- und Sozialanamnese inklusive AU-Zeiten und Kontextfaktoren) sowie Fremdinformationen und durch den Gutachter erhobene Untersuchungsbefunde. Anschließend erfolgt in der Argumentationskette eine Informationsbewertung der vom Gutachter gesammelten bzw. dokumentierten Informationen mittels Formulierung von Diagnosen (nach ICD) mit daraus folgenden Funktionseinschränkungen und einer kritischen Bewertung der Informationssammlung in der Epikrise. Hierbei sind die Funktionseinschränkungen, die Kompensationsmöglichkeiten des Antragstellers, mögliche Therapie-Optionen wie auch differentialdiagnostische Überlegungen vom Gutachter darzulegen und zu gewichten. Den Abschluss der Argumentationskette bildet die Beantwortung der Fragestellung, die im Gutachtenauftrag formuliert ist. Dies geschieht über die sozialmedizinische Beurteilung des Leistungsvermögens mit Aussagen zum quantitativen und qualitativen Leistungsvermögen, zu Beginn und Dauer der Leistungsminderung sowie zum Reha-Bedarf.

Mit der Bewertung dieses Kriteriums gibt der Peer ein Gesamturteil zur Nachvollziehbarkeit des von ihm geprüften Gutachtens ab. Grundlage der Bewertung ist die Schlüssigkeit der Argumentationskette, d. h. die sinnvolle Verknüpfung der einzelnen Begutachtungsschritte und schlüssige Herleitung der sozialmedizinischen Beurteilung der Leistungsfähigkeit. Abhängig von der Fallkonstellation können Mängel in den Gliedern der Argumentationskette unterschiedlich gewichtet werden. Diese Gewichtung im Einzelfall vorzunehmen, ist Aufgabe des Peers. Beispielsweise wäre zu

besonderen qualitativen Einschränkungen Stellung zu nehmen, wenn diese aus dem Krankheitsbild zu erwarten sind wie z. B. bei der arteriellen Verschlusskrankheit eine Einschränkung der Gehstrecke oder bei der rheumatoiden Arthritis eine Einschränkung der Gebrauchsfähigkeit der Hände.

Bei der Bewertung eines Gutachtens sind Mängel, d. h. Brüche der Argumentationskette, vom Peer zu benennen und anhand des Gutachtens konkret zu belegen.

Die Bewertung der Argumentationskette wird über ein Ampelsystem vorgenommen:

- Die Ampelbewertung GRÜN kennzeichnet, dass keine nennenswerten Brüche in der Argumentationskette vorliegen.
- Bei der Ampelbewertung GELB liegen Brüche in der Argumentationskette vor, jedoch ermöglichen die im Gutachten enthaltenen sozialmedizinischen Informationen dem Peer, diese durch entsprechende Neubewertung zu überbrücken und das Gutachtenergebnis damit nachzuvollziehen.
- Bei der Ampelbewertung ROT liegen Brüche in der Argumentationskette vor, die vom Sozialmediziner auch mittels Neubewertung der sozialmedizinischen Informationen nicht mehr überbrückt werden können.

Bei der Bewertung des Übergeordneten Kriteriums kommt es nicht darauf an, ob der Peer das Gutachtenergebnis teilt oder er selbst zu einem anderen Ergebnis kommt.

Da eine gravierende Verletzung des Neutralitätsgebotes die Nachvollziehbarkeit des gesamten Gutachtens grundsätzlich in Frage stellt, ist an dieser Stelle auch die Neutralität zu überprüfen. Eine Verletzung der Neutralität liegt vor, wenn offensichtlich herabsetzende und beleidigende Äußerungen vom Gutachter gegenüber einer Person oder Bevölkerungsgruppe gemacht werden. Die Voreingenommenheit und Parteilichkeit muss aus dem Text des Gutachters eindeutig belegbar sein. Ein gravierender Verstoß gegen das Neutralitätsgebot führt zu der Ampelbewertung ROT.

Grün	Die Argumentationskette für die sozialmedizinische Beurteilung ist in sich schlüssig. Es sind keine nennenswerten Brüche in der Argumentationskette vorhanden.
Gelb	Die Argumentationskette für die sozialmedizinische Beurteilung ist lückenhaft, vom Sozialmediziner können jedoch Zusammenhänge selbst hergestellt und festgestellte Brüche in der Argumentationskette überbrückt werden.
Rot	Die Argumentationskette für die sozialmedizinische Beurteilung ist unterbrochen, die Brüche können auch vom Sozialmediziner nicht überbrückt werden. ODER Es liegt ein gravierender Verstoß gegen das Neutralitätsgebot vor.

Welche Mängel wurden festgestellt?

Die festgestellten Qualitätsdefizite, die der Mängelbewertung zugrunde liegen, sind auf dem Prüfbogen zu dokumentieren.

Anhang 10:

Dokumentation der freitextlichen Mängelbegründungen während der standardisierten Peer-Schulung (Lernziele 2.2, 2.3, 2.5)

Erläuterung zu den Tabellen: Das ‚x‘ beschreibt für welche Mängelkategorie sich eine Murrelgruppe entschieden hat. Die in der ersten Spalte ‚fett+kursiv‘ formatierte Mängelkategorie zeigt die zu dieser Prüfrage im Vorfeld von den sozialmedizinischen Experten konsentierende Lösung. Im darunter liegenden Tabellenbereich ‚Begründungen‘ werden die dokumentierten Mängelbegründungen der Murrelgruppen dargestellt. Im Tabellenkopf finden sich die einzelnen Kleingruppen entsprechend des Schulungsortes, d.h. Würzburg 1 bedeutet Kleingruppe 1 am Schulungsort Würzburg usw.

Prüfrage A1. Inwieweit werden die Gliederungsvorgaben gemäß den Anforderungen eingehalten?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	xxx	x	xx	xxx
Leichte Mängel				
Deutliche Mängel				
Gravierende Mängel		xx	x	
Begründungen				
Würzburg 1				
Würzburg 2	Überschriften „Anamnese“ und „Befund“ fehlen Layout nicht abgegrenzt			
Berlin 1				
Berlin 2	„Anamnese“ „Untersuchungsbefunde“ sind anders benannt			

Prüfrage A2. Inwieweit wird der einheitliche Formulareinsatz bestehend aus Deck- und Schlussblatt verwendet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	xxx		xx	xx
Leichte Mängel		xx	x	
Deutliche Mängel				
Gravierende Mängel				x
Begründungen				
Würzburg 1				
Würzburg 2	abgewandelt			
Berlin 1				
Berlin 2				

Prüfrage B1. Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel				
Leichte Mängel	xxx	xxx	xx	xxx
Deutliche Mängel			x	
Gravierende Mängel				
Begründungen				
Würzburg 1	Unnötige Längen (PSA-Verlauf) Wiederholung (Anamnese & Epikrise)			
Würzburg 2				
Berlin 1	Wiederholungen			
Berlin 2				

Prüfrage B2. Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel				
Leichte Mängel		x		x
Deutliche Mängel	xx	x	xx	xx
Gravierende Mängel	x	x	x	
Begründungen				
Würzburg 1	Beispiele: „bipolar“, „PSA“, „erektiler Dysfunktion“, „radikale Prostatovesikulektomie“, „Kataraktexzision“ Abkürzung „LMR“			
Würzburg 2	„PSA“, „bipolare Störung“, arthrotomische Eingriffe; „erektiler Dysfunktion“, „radikale Prostatovesikulektomie“			
Berlin 1	Beispiele: „PSA“, „Osteoarthrotomie“, „Kataraktexzision“			
Berlin 2	„PSA“, „LMR“, primäre Operation, sekundäre Bestrahlung → mehrfach nicht erklärte Fachbegriffe: „bipolare Störung“, „Kataraktexzision“			

Prüfrage B3. Inwieweit wird die sozialmedizinische Terminologie richtig angewendet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel		xx	xx	xx
Leichte Mängel			x	x
Deutliche Mängel	xx	x		
Gravierende Mängel				
Begründungen				
Würzburg 1	Beispiele: „Belastungsfähigkeit“, „Anmarschweg (500m/Zeit)“, „nervliche Belastung“			
Würzburg 2	Belastungsfähigkeit Zeitungsfähigkeit 3-6 Stunden → Tagesbezug fehlt			
Berlin 1	„Wegefähigkeit“, „nervliche Belastung“			
Berlin 2	missverständliche Terminologie			

Prüffrage C1. Inwieweit sind Herkunft, zeitliche Einordnung und Kernaussagen von medizinischen Informationen erkennbar?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	x			
Leichte Mängel	xx		x	
Deutliche Mängel		x	xx	xxx
Gravierende Mängel		xx		
Begründungen				
Würzburg 1	Herkunft „bipolare Störung“ „urologische Kontrollen“ „Arthroskopie“			
Würzburg 2	affektive Störung bildet sich nicht ab Kniegelenk Schweregrad keine klare Darstellung über Herkunft der med. Information Blutzuckerwerte Abklärung Neuropathie Herkunft der Information in Epikrise			
Berlin 1	Herkunft psychische Störung warum Kniegelenk wann/wie PSA-Wert Inkontinenz (Ausprägung) keine Abtrennung eigene Befunde/ Anamnese Details zur Diabetes Kernaussage E-Bericht Rententeil bis Blatt 2C			
Berlin 2				

Prüffrage C2. Inwieweit wird dargestellt, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben werden?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	xx	xxx	x	x
Leichte Mängel			x	
Deutliche Mängel			x	xx
Gravierende Mängel	x			
Begründungen				
Würzburg 1	Urinbefund → kein Verfahren beschrieben			
Würzburg 2				
Berlin 1				
Berlin 2	psychiatrischer Befund Laborwerte in der Epikrise (Urinzucker)			

Prüffrage C3. Inwieweit wird dargestellt, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel				
Leichte Mängel				
Deutliche Mängel	x		xx	
Gravierende Mängel	xx	xxx	x	xxx
Begründungen				
Würzburg 1	Angaben zu Knie sind relevant Angabe PSA-Wert Spirometrie			
Würzburg 2	Kniegelenk Referenzbereich→ sozialmed. Relevant Urinzucker-Referenzwerte			
Berlin 1				
Berlin 2	Lungenfunktion ohne Werte, depressive Störung, Kniegelenke, Diabetes-Werte → Referenzbereiche fehlen Neutral-Null			

Prüffrage D1. Inwieweit wird die Anamnese vollständig dargestellt?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel			x	xx
Leichte Mängel	xx	x		x
Deutliche Mängel	x	xx	x	
Gravierende Mängel			x	
Begründungen				
Würzburg 1	Tagesstruktur			
Würzburg 2	Tagesstruktur fehlt (relevant, weil psych. Diagnose; aber internistisches GA)			
Berlin 1				
Berlin 2	Tagesstruktur Vollständigkeit d. „behandelnden Ärzte“			

Prüffrage D2. Inwieweit sind die Diagnosen dargestellt und mit unktionseinschränkungen (sog. Funktionsdiagnosen) formuliert?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel				
Leichte Mängel				
Deutliche Mängel	xxx	xxx	xxx	xxx
Gravierende Mängel				
Begründungen				
Würzburg 1	keine Funktionsdiagnosen bei Knie und Inkontinenz			
Würzburg 2	keine Funktionseinschränkungen			
Berlin 1				
Berlin 2	Kniegelenk keine Funktionsdiagnose			

Prüffrage D3. Inwiefern erfolgt eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel				X
Leichte Mängel				
Deutliche Mängel			X	X
Gravierende Mängel	XX	XX	XX	X
Begründungen				
Würzburg 1	Medikamente weitere Therapien			
Würzburg 2	bisherige Behandlung keine Stellungnahme zu Therapien			
Berlin 1				
Berlin 2	bisherige Behandlung (Depression, ??) Stellungnahme zukünftige Therapieerfolge Therapieoptionen fehlen			

Prüffrage D4. Inwieweit werden wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten dargelegt?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	X	XX		X
Leichte Mängel	X			XX
Deutliche Mängel		X	XX	
Gravierende Mängel	X		X	
Begründungen				
Würzburg 1	nicht in Epikrise			
Würzburg 2	Rehaentlassbericht			
Berlin 1				
Berlin 2	fehlende Darstellung der Abweichung zur Rehabilitation (Reha-E-Bericht) → Teilaspekte			

Prüffrage D5. Inwieweit werden die sozialmedizinischen Fragestellungen vollständig beantwortet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel			X	X
Leichte Mängel				
Deutliche Mängel	XX	X	X	
Gravierende Mängel	X	XX	X	XX
Begründungen				
Würzburg 1	Beginn fehlt Diskrepanz zwischen Freitext und „Kreuzchen“			
Würzburg 2	Leistungsminderung fehlt			
Berlin 1				
Berlin 2	Diskrepanz qual. Leistung in Epikrise zu Schlussblatt fehlende Aussage Leistungsfähigkeit letzter Beruf Beginn Leistungseinschränkung Diskrepanz Dauer der Leistungsminderung			

Prüffrage E1. Inwieweit wird der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	XXX	XXX	XXX	XXX
Leichte Mängel				
Deutliche Mängel				
Gravierende Mängel				

Prüffrage E2. Inwieweit wird die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	x			x
Leichte Mängel				
Deutliche Mängel				
Gravierende Mängel	xx	xxx	xx	xx
Begründungen				
Würzburg 1	psychiatrische Diagnostik Prognose			
Würzburg 2	keine Erwähnung Literatur Abweichung psychische Störung → relevant für sozialmed. Urteil psychiatr. Diagnose nicht leitliniengerecht			
Berlin 1				
Berlin 2	psychiatrischer Befund fachärztlicher Befund/Betreuung Kniebefund, weil bei 3 Hauptdiagnosen			

Prüffrage F1. Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	xxx	xxx	xxx	xxx
Leichte Mängel				
Deutliche Mängel				
Gravierende Mängel				
Begründungen				
Würzburg 1				
Würzburg 2				
Berlin 1				
Berlin 2	Spirometrie ohne Erkenntnisgewinn			

Prüffrage F2. Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	xx		xx	
Leichte Mängel		x		
Deutliche Mängel		x		x
Gravierende Mängel	x	x		xx
Begründungen				
Würzburg 1	Psychometrie fehlt (eventuell Zusatzgutachten?) Röntgen Knie			
Würzburg 2	zusätzliche Laboruntersuchung (Tumorgeschehen) testpsychologische Untersuchung (besser psychopathologischer Befund) Atemnot: weitere Diagnostik weitere Diagnostik KHK psychiatrische Diagnosen belegen			
Berlin 1				
Berlin 2	psychopathologischer Befund Kniegelenk Anamnese links, Abschluss „die“ Kniegelenke Inkontinenz fehlerhafte „Diagnostik“			

Ampel. Bewerten Sie die Nachvollziehbarkeit des Schulungsgutachtens 1

Ampel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Grün				
Gelb	x	x		xxx
rot	xx	xx	xxx	

Begründungen

Würzburg 1

Würzburg 2

Brüche in der Argumentation
Inkongruenzen
Gewichtung der Diagnosen
führt sehr viel auf → Brüche überbrückbar

Berlin 1

Berlin 2

Vorgeschichte zur Psyche dünn
aktuelle Therapie fehlt
dürftiger psychopathologischer Befund
Verknüpfung der Argumentationskette: Vielzahl von Störungen werden
aufgeführt, aber nicht verknüpft

Ampel. Bewerten Sie die Nachvollziehbarkeit des Schulungsgutachtens 2

Ampel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Grün				
Gelb				
Rot	xxx	xxx	xxx	xxx

Begründungen

Würzburg 1

Würzburg 2

keine Befundlage, die Leistungsbeurteilung rechtfertigt
keine spezifizierte Depressivität
psychopathologischer Befund fehlt

Berlin 1

Berlin 2

Erläuterungen zu den Tabellen zum Lernziel 2.5 der standardisierten Peer-Schulung. Im Tabellenkopf finden sich die einzelnen Kleingruppen entsprechend des Schulungsortes abgebildet, d.h. Würzburg 1 bedeutet Kleingruppe 1 am Schulungsort Würzburg. Die Zahlen beschreiben die Menge der Peers, die sich für die entsprechende Mängelkategorie ausgesprochen haben. Die in der ersten Spalte **fett+kursiv** formatierte Mängelkategorie zeigt wiederum die zu dieser Prüffrage im Vorfeld konsentierende Lösung. Die Dokumentation der festgestellten Qualitätsdefizite erfolgte in diesem Lernziel mündlich und wurde nicht auf Flipchart festgehalten.

Prüffrage A1. Inwieweit werden die Gliederungsvorgaben gemäß den Anforderungen eingehalten?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	2	4		
Leichte Mängel				
Deutliche Mängel	3	1		2
Gravierende Mängel	5	4	10	8

Prüffrage A2. Inwieweit wird der einheitliche Formularsatz bestehend aus Deck- und Schlussblatt verwendet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	1	1	4	2
Leichte Mängel	9	6	5	6
Deutliche Mängel			1	1
Gravierende Mängel		2		1

Prüffrage B1. Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	8	6	6	9
Leichte Mängel	2	3	2	1
Deutliche Mängel			2	
Gravierende Mängel				

Prüffrage B2. Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	5	3	5	3
Leichte Mängel	3	6	4	3
Deutliche Mängel	2	1	1	3
Gravierende Mängel				

Prüffrage B3. Inwieweit wird die sozialmedizinische Terminologie richtig angewendet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	9	6	7	7
Leichte Mängel		2	1	
Deutliche Mängel	1	2	2	3
Gravierende Mängel				

Prüfrage C1. Inwieweit sind Herkunft, zeitliche Einordnung und Kernaussagen von medizinischen Informationen erkennbar?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel			3	2
Leichte Mängel	5	4	2	3
Deutliche Mängel	1	2	4	1
Gravierende Mängel	4	2	1	4

Prüfrage C2. Inwieweit wird dargestellt, mit welchen Verfahren, Methoden und Instrumenten die Befunde erhoben werden?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	8	8	8	10
Leichte Mängel			2	
Deutliche Mängel				
Gravierende Mängel	2	1		

Prüfrage C3. Inwieweit wird dargestellt, welche Messwerte, Referenzbereiche und Graduierungen die Grundlage der Beurteilung sind?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	10	10	9	10
Leichte Mängel			1	
Deutliche Mängel				
Gravierende Mängel				

Prüfrage D1. Inwieweit wird die Anamnese vollständig dargestellt?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	1	3	1	3
Leichte Mängel		2	4	3
Deutliche Mängel	5	3	5	1
Gravierende Mängel	4	1		3

Prüfrage D2. Inwieweit sind die Diagnosen dargestellt und mit Funktionseinschränkungen (sog. Funktionsdiagnosen) formuliert?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel		1	2	1
Leichte Mängel				
Deutliche Mängel	9	8	2	8
Gravierende Mängel	1			1

Prüfrage D3. Inwiefern erfolgt eine Stellungnahme zur bisherigen Therapie und zu weiteren therapeutischen Optionen?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel		1	3	5
Leichte Mängel	1	2		
Deutliche Mängel	4	5	3	4
Gravierende Mängel	5	1	4	1

Prüfrage D4. Inwieweit werden wesentliche Abweichungen in der sozialmedizinischen Beurteilung der Leistungsfähigkeit gegenüber Vorgutachten dargelegt?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	3	7	7	6
Leichte Mängel	1			
Deutliche Mängel	2		2	1
Gravierende Mängel	4	3	1	2

Prüfrage D5. Inwieweit werden die sozialmedizinischen Fragestellungen vollständig beantwortet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	2	1	1	
Leichte Mängel	1		3	1
Deutliche Mängel	4	4	4	4
Gravierende Mängel	3	3	2	4

Prüfrage E1. Inwieweit wird der allgemein anerkannte Stand medizinischer Erkenntnisse beachtet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	10	10	10	10
Leichte Mängel				
Deutliche Mängel				
Gravierende Mängel				

Prüfrage E2. Inwieweit wird die vorhandene Begutachtungsliteratur der Deutschen Rentenversicherung beachtet?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	2	3	5	1
Leichte Mängel	2	2		4
Deutliche Mängel	4			3
Gravierende Mängel	2	3	5	2

Prüfrage F1. Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung geeignet und notwendig?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	10		10	10
Leichte Mängel				
Deutliche Mängel				
Gravierende Mängel				

Prüfrage F2. Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der sozialmedizinischen Fragestellung ausreichend?

Mängel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Keine Mängel	4		7	5
Leichte Mängel				2
Deutliche Mängel	3			1
Gravierende Mängel	3		3	1

„Ampel-Bewertung“ des übergeordneten Kriteriums. Bewerten Sie die Nachvollziehbarkeit des Schulungsgutachtens 3

Ampel	Würzburg 1	Würzburg 2	Berlin 1	Berlin 2
Grün		5		3
Gelb	9	3	7	6
rot	1	1	3	1

Abbildungsverzeichnis

Abbildung 1:	Überblick der Rentenarten (modifiziert nach Götz, Roth & Chojetzki, 2011, S. 14).	12
Abbildung 2:	EM-Rentenzugänge nach ausgewählten Diagnosegruppen im Zeitverlauf (modifiziert nach DRV, 2017b, S. 2).	14
Abbildung 3:	Szenarien für die Entwicklung von Morbidität und Langlebigkeit (modifiziert nach Fries, 2003, S. 456).	23
Abbildung 4:	Modell zur Erklärung gesundheitlicher Ungleichheit (modifiziert nach Mielck, 2012, S. 136).	26
Abbildung 5:	ICF-Modell der Funktionsfähigkeit und Behinderung mit Wechselwirkungen zwischen den Komponenten (modifiziert nach DIMDI/WHO, 2005, S. 23).....	29
Abbildung 6:	Gutachterliche Argumentationskette innerhalb eines sozialmedizinischen Gutachtens (entnommen aus Cibis et al., 2011, S. 111).	41
Abbildung 7:	Operationalisierung der Ergebnisqualität (modifiziert nach DRV, 2009c, S. 8)	58
Abbildung 8:	Struktur der Dissertation nach dem Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).....	67
Abbildung 9:	Darstellung der Sensitivität und Präzision bei der Literaturrecherche.....	75
Abbildung 10:	Schematische Darstellung der Suchstrategie nach Boeker und Motschall (modifiziert nach Gechter et al., 2013, S. 36).	76
Abbildung 11:	Flussdiagramm der verschiedenen Phasen der systematischen Recherche (modifiziertes PRISMA-Diagramm nach Moher, Liberati, Tetzlaff & Altman, 2011, S. e11).	81
Abbildung 12:	Schema zur Beurteiler-Übereinstimmung und Beurteiler-Reliabilität.	126
Abbildung 13:	Schema des Vorgehens bei der qualitativen Inhaltsanalyse.....	130
Abbildung 14:	Varianz der Ampelbewertungen (übergeordnetes Kriterium) für die 20 von allen Peers beurteilten Gutachten (in Prozent; modifiziert nach Strahl et al., 2016, S.158).....	135
Abbildung 15:	Varianz der Peerbeurteilungen des übergeordneten Kriteriums (in Prozent).	136

Abbildung 16: Subjektive Beurteilung der Relevanz von Prüffragen und übergeordnetem Kriterium (n = 16)	163
Abbildung 17: Veranschaulichungsbeispiel aus dem Schulungscurriculum (Entwurf)..	186
Abbildung 18: Ebenen der Reliabilitätsberechnungen im Rahmen der summativen Evaluation.....	190
Abbildung 19: Varianz der Peerbeurteilungen des übergeordnetes Kriteriums (in Prozent) der Hausaufgabengutachten.....	199
Abbildung 20: Entwurf des Lernziels 1.6 der standardisierten Peer-Schulung	215
Abbildung 21: Einordnung der Teilstudie 1 in den Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).....	222
Abbildung 22: Einordnung der Teilstudie 2 in den Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).....	223
Abbildung 23: Einordnung der Teilstudie 3 in den Public Health Action Cycle (modifiziert nach Rosenbrock, 1995).....	224

Tabellenverzeichnis

Tabelle 1:	Grundprinzipien der Sozialversicherung (modifiziert nach Horn & Schuchardt, 2015, S. 5).....	8
Tabelle 2:	Reformziele des Gesetzes zur Organisationsreform in der gesetzlichen Rentenversicherung (RVOrgG; modifiziert nach Göbel, 2015, S. 84)	10
Tabelle 3:	12-Monats-Prävalenzen psychischer Störungen in der erwachsenen Allgemeinbevölkerung (modifiziert nach Jacobi et al., 2014, S. 80)	17
Tabelle 4:	Erbrachte Leistungen der Deutschen Rentenversicherung bei Krankheiten des Muskel-Skelett-Systems und des Bindegewebes unterteilt nach Diagnosegruppen (Stand: 2016; modifiziert nach DRV, 2017c, S. 4-5).....	20
Tabelle 5:	Anwendungsgebiete der ICF (modifiziert nach DIMDI/WHO, 2005, S. 11)...	28
Tabelle 6:	Definition der ICF Komponenten (modifiziert nach DIMDI/WHO, 2005, S. 17-19; Ostholt-Corsten & Schuntermann, 2011, S. 69)	30
Tabelle 7:	EUMASS Core Sete: ICF Core Set für die „Begutachtung von Beeinträchtigungen in der Sozialversicherung“ (modifiziert nach Brage, Donceel & Falez, 2008. S. 1395).....	33
Tabelle 8:	Gliederung und Anforderungsprofil des ärztlichen Gutachtens für die gesetzliche Rentenversicherung (modifiziert nach VDR, 2001. S. 21).	39
Tabelle 9:	Einzelkriterien zur Erfassung der Qualität von sozialmedizinischen Gutachten (modifiziert nach Cibis et al., 2011. S. 111)	60
Tabelle 10:	Operationalisierung der Mängelkategorien am Beispiel der Prüffrage B1: „Inwieweit ist die sprachliche Ausdrucksweise korrekt und unmissverständlich?“ (entnommen aus DRV, 2013. S. 15).....	60
Tabelle 11:	Fragestellungen der systematischen Literaturrecherche zur Identifikation wissenschaftlich fundierter Literatur zur Evidenzbasierung in der Sozialmedizinischen Begutachtung.	72
Tabelle 12:	PICO-Schema der systematischen Literaturrecherche zur Evidenzbasierung in der Sozialmedizinischen Begutachtung (Schema nach Schmucker, Motschall, Antes & Meerpohl, 2013, S.1392).....	76
Tabelle 13:	Schlagworte und Synonyme der elektronischen Datenbankrecherche	77

Tabelle 14: Adaptierte Evidenzklassen in PGF (modifiziert nach Kliche et al., 2006, S. 147)	80
Tabelle 15: Inhaltliche Charakteristika von Studien mit praktischen Hinweise zur Sozialmedizinischen Begutachtung.	90
Tabelle 16: Inhaltliche Charakteristika von Studien mit Bezug zur Anwendung der ICF in der Sozialmedizinischen Begutachtung.	97
Tabelle 17: Inhaltliche Charakteristika von Studien zur Sichtweise der Versicherungsmedizin auf die Sozialmedizinische Begutachtung	100
Tabelle 18: Evidenzklassen der eingeschlossenen Studien	102
Tabelle 19: Qualitätsbeurteilung der eingeschlossenen Studien anhand des Mixed Methods Appraisal Tools nach Hong et al. (2018)	103
Tabelle 20: Fragestellungen zur Analyse der Interrater-Reliabilität des Peer Review Verfahrens.....	117
Tabelle 21: Ausschnitt aus Schema für die Gutachtenverteilung auf die Peers.....	121
Tabelle 22: Überblick Prüffragenkatalog des Manuals zum Peer Review-Verfahren (Löffler et al., 2010, Anhang 12, S.4)	123
Tabelle 23: Anzahl der freitextlichen Mängelbegründungen der Einzelkriterien.....	131
Tabelle 24: Mittelwerte der Beurteiler-Konkordanz (325 Reviews)	132
Tabelle 25: Mittelwerte der Beurteiler-Konkordanz der Einzelkriterien (325 Reviews)..	133
Tabelle 26: Mittelwerte der Beurteiler-Konkordanz des übergeordneten Kriteriums (325 Reviews)	134
Tabelle 27: Gutachtenspezifische prozentuale Übereinstimmung des übergeordneten Kriteriums (325 Reviews).....	135
Tabelle 28: Codierung der Prüffrage A1.....	137
Tabelle 29: Codierung der Prüffrage A2.....	138
Tabelle 30: Codierung der Prüffrage B1.....	139
Tabelle 31: Codierung der Prüffrage B2.....	140
Tabelle 32: Beispiel für gleiche Mängelbegründung bei unterschiedlicher Peer-Beurteilung	140
Tabelle 33: Beispiel für gleiche Mängelbegründung bei unterschiedlicher Peer-Beurteilung	140

Tabelle 34: Codierung der Prüffrage B3.....	141
Tabelle 35: Codierung der Prüffrage B4.....	142
Tabelle 36: Codierung der Prüffrage C1	143
Tabelle 37: Codierung der Prüffrage C2	144
Tabelle 38: Codierung der Prüffrage C3	145
Tabelle 39: Codierung der Prüffrage D1	146
Tabelle 40: Codierung der Prüffrage D2	147
Tabelle 41: Codierung der Prüffrage D3	148
Tabelle 42: Codierung der Prüffrage D4	149
Tabelle 43: Codierung der Prüffrage D5	150
Tabelle 44: Codierung der Prüffrage D6	151
Tabelle 45: Codierung der Prüffrage D7	152
Tabelle 46: Codierung der Prüffrage D8	153
Tabelle 47: Codierung der Prüffrage D9	154
Tabelle 48: Codierung der Prüffrage E1.....	155
Tabelle 49: Codierung der Prüffrage E2.....	156
Tabelle 50: Codierung der Prüffrage F1.....	157
Tabelle 51: Codierung der Prüffrage F2.....	158
Tabelle 52: Codierung des übergeordneten Kriteriums	159
Tabelle 53: Übersicht über die Qualitätsbewertung, dokumentierte Qualitätsdefizite und den dazugehörigen inhaltsanalytische Codes am Beispielgutachten 1	161
Tabelle 54: Fragestellungen der formativen und summativen Evaluation eines strukturierten Peer-Schulungsprogramms.	179
Tabelle 55: Adaptierte Bloom'scheTaxonomie der kognitiven Lernziele (modifiziert nach Krathwohl, 2002, S. 215; Baumgartner, 2014, S. 43-46).	182
Tabelle 56: Lernzielinhalte und Vermittlungsmethoden.....	191
Tabelle 57: Aufbau und Lernziele des Schulungscurriculums (modifiziert nach Strahl et al., 2018, S. 4)	192

Tabelle 58: Sicherheit im Umgang mit den Bestandteilen des Manuals in Prozent (n = 36) (nach Strahl et al., 2018)	195
Tabelle 59: Beurteilung der didaktischen Methoden und Schulungsmaterialien (n = 36).....	196
Tabelle 60: Qualitative Beurteilung des Schulungscurriculums und dessen Durchführung (n = 36)	196
Tabelle 61: Allgemeine Beurteilung der Peer-Schulung (n = 36)	197
Tabelle 62: Gutachtenspezifische prozentuale Übereinstimmung in Prozent (119 Reviews)	198
Tabelle 63: Gutachtenspezifische prozentuale Übereinstimmung in Prozent (119 Reviews)	199
Tabelle 64: Beurteilerübereinstimmung mittels Kappa-Koeffizient von Fleiss (κ_m) für mehrere Beurteiler (117 Reviews)	200
Tabelle 65: Prozentuale Übereinstimmung mit dem Peer-Konsens (Modus) in Prozent (119 Reviews)	201
Tabelle 66: Prozentuale Übereinstimmung mit dem Experten-Konsens in Prozent (119 Reviews)	202
Tabelle 67: Beurteilerübereinstimmung mittels Kappa-Koeffizient von Fleiss (κ_m) im Vergleich von IG und KG (IG: 78 Reviews; KG: 54 Reviews).....	203

Abkürzungsverzeichnis

Abb.	Abbildung
Abs.	Absatz
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
BAS-G	Bath Ankylosing Spondylitis patient Global score
BASMI	Bath Ankylosing Spondylitis Metrology Index
BASRI	Bath Ankylosing Spondylitis Radiology Index
CDSR	Cochrane Database of Systematic Reviews
CED	chronisch entzündliche Darmerkrankung
CENTRAL	Cochrane Central Register of Controlled Trials
CIDI	Composite International Diagnostic Interview
COPD	chronic obstructive pulmonary disease (Chronisch obstruktive Lungen- erkrankung)
CT	Computertomographie
DAS28	Disease Activity Score 28
DIPS	Diagnostischem Interview bei psychischen Störungen
DRV	Deutsche Rentenversicherung
EbM	Evidenz-basierte Medizin
EEG	Elektroenzephalographie
EKG	Elektrokardiogramm
EM-Rente	Erwerbsminderungsrenten / Renten infolge verminderter Erwerbsfähigkeit
EvaBegut	Projekt Evaluation und Aufwandschätzung zum Qualitätssicherungs- verfahren der Sozialmedizinischen Begutachtung
FFbH-RA	Funktionsfragebogen Hannover – Rheumatoide Arthritis
GAF	Global Assessment of Functioning

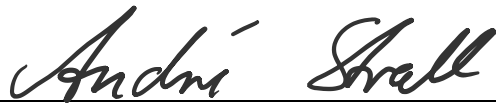
GAS	Global Assessment Scale
GdB	Grad der Behinderung
GKV	gesetzliche Krankenversicherung
GSDS-II	Goningen Social Disability Schedule
HAQ	Health Assessment Questionnaire
(I)ADL	(Instrumental) Activities of Daily Living
I. d. R.	in der Regel
IMET	Index zur Messung von Einschränkungen der Teilhabe
ICC	Intra-Klassen-Korrelation
ICD	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
ICF	Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit
IG	Interventionsgruppe
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
Kap.	Kapitel
KG	Kontrollgruppe
KHK	koronare Herzkrankheit
LitBegut	Projekt Systematische Literaturrecherche zur sozialmedizinischen Begutachtung bei Erwerbsminderungsrenten und Reha-Zugang
LuFu	Lungenfunktionsdiagnostik
MeSH	Medical Subject Headings
M	Mittelwert
MMAT	Mixed Methods Appraisal Tool
MMPI-2	Minnesota Multiphasic Personality Inventory-2
MRT	Magnetresonanztomographie
PGF	Prävention und Gesundheitsförderung
PrävG	Gesetz zur Stärkung der Gesundheitsförderung und der Prävention (Präventionsgesetz)

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROs	Peer Review Organisations
PSROs	Professional Standards Review Organisations
QUOROM	Quality of Reporting of Meta-analyses
RA	rheumatoide Arthritis
RCT	randomisiert kontrollierte Studie
RKI	Robert Koch Institut
RV-Träger	Rentenversicherungsträger
RVOrgG	Gesetzes zur Organisationsreform in der gesetzlichen Rentenversicherung
SchuP	Projekt Entwicklung eines Peer-Schulungsprogramms für das trägerübergreifende Qualitätssicherungsverfahren der Sozialmedizinischen Begutachtung
SD	Standardabweichung
SGB	Sozialgesetzbuch
SIRS	Structured Interview of Reported Symptoms
SKID	Strukturiertes Klinisches Interview für DSM-IV
SOFAS	Social and Occupational Functioning Assessment Scale
SOMEKO	Kommission zur Weiterentwicklung der Sozialmedizin in der gesetzlichen Rentenversicherung
s.	siehe
Std.	Stunde
Tab.	Tabelle
TSI	Trauma Symptom Inventory
vgl.	vergleiche
WHO	World Health Organisation (Weltgesundheitsorganisation)
WHODAS 2.0	World Health Organization Disability Assessment Schedule II
WoS CC	Web of Science Core Collection
z. B.	zum Beispiel
ZPID	Leibniz-Zentrum für Psychologische Information und Dokumentation

Persönliche Versicherung des Bewerbers

„Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und andere als in der Dissertation angegebene Hilfsmittel nicht benutzt habe; die aus fremden Quellen (einschließlich elektronischer Quellen, dem Internet und mündlicher Kommunikation) direkt oder indirekt übernommenen Gedanken sind ausnahmslos unter genauer Quellenangabe als solche kenntlich gemacht. Zentrale Inhalte der Dissertation sind nicht schon zuvor für eine andere Qualifikationsarbeit verwendet worden. Insbesondere habe ich nicht die Hilfe sogenannter Promotionsberaterinnen bzw. Promotionsberater in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar Geld oder geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Auf die Bedeutung einer eidesstattlichen Versicherung und die strafrechtlichen Folgen einer, auch fahrlässigen, falschen oder unvollständigen eidesstattlichen Versicherung und die Bestimmungen der §§ 156, 161 StGB bin ich hingewiesen worden.“

Hamburg, den 22.03.2019



André Strahl