



Bericht der Pilotstudie zum Peer Review Verfahren (PRV)

zur Qualitätsbeurteilung von medizinischen Gutachten der IV

Erstellt von der Fachstelle EKQMB und verabschiedet von der EKQMB an ihrer Sitzung vom 25. Oktober 2023

Zusammenfassung

Das primäre Ziel der Pilotstudie war die Validierung des Prüffragenkatalogs des Peer-Review-Verfahrens (PRV). Die Studie wurde erfolgreich durchgeführt, wobei die Schulung der Reviewer effektiv etabliert und gut umgesetzt wurde. Die verwendete elektronische Online-Plattform hat sich während der Studie als zuverlässig erwiesen, und es traten keine nennenswerten technischen Probleme auf.

Die Reviewer bewerteten die Definitionen der Kriterien und des Schweregrades der Mängel (Ankerdefinitionen) der Prüffragen als verständlich und nützlich, was die Schulung bestätigte, da nur minimale Präzisierungen notwendig waren. Dies unterstreicht die bereits jetzt erreichte Qualität der Schulungsmaterialien. Der durchschnittliche Zeitaufwand für die Überprüfung eines Teilgutachtens lag bei 95 Minuten, während mangelhafte Gutachten mit 107 Minuten einen signifikant höheren Zeitaufwand erforderten. In diesem Kontext wird deutlich, dass die Qualität des Gutachtens erheblichen Einfluss auch auf den Aufwand des PRV hat.

Eine Überprüfung des Fragebogens auf mögliche Reduktionen zeigte, dass das Entfernen formaler Prüffragen in der PRV (Vollversion) keine substantielle Zeitersparnis zur Folge hätte. Dies legt nahe, dass alle aktuellen Prüffragen wesentlich für die gründliche Bewertung der Gutachten sind und beibehalten werden sollten. Hingegen soll die PRV-Kurzversion um drei zusätzliche Prüffragen erweitert werden, um die Erfassung kritischer Qualitätsaspekte zu verbessern. Gleichzeitig wird vorgeschlagen, zwei der weniger relevanten Prüffragen aus der Kurzversion zu entfernen, um die Effizienz des Verfahrens zu steigern und den Fokus stärker auf die wesentlichen Elemente der Bewertung zu legen.

Peer Review Verfahren (PRV-CH)

M. Braun^a, M. Hermelink^a, R. Jaquier^a, M. Liebrez^{a,d} & R. Schleifer^a

^a Fachstelle der Eidgenössischen Kommission für Qualitätssicherung in der medizinischen Begutachtung
^b Bundesamt für Sozialversicherungen
^c Eidgenössische Kommission für Qualitätssicherung in der medizinischen Begutachtung
^d Forensisch-Psychiatrischer Dienst der Universität Bern

Einleitung

Bisher haben die internen medizinischen Dienste der Schweizerischen Invalidenversicherung die Qualität von Gutachten unsystematisch und uneinheitlich von Fall zu Fall beurteilt¹. Eine Arbeitsgruppe des Bundesamtes für Sozialversicherungen hat deshalb das Schweizerische Peer-Review-Verfahren (PRV-CH) zur Begutachtung von medizinischen Gutachten entwickelt. Die vom Gesetzgeber im Jahr 2022 neu geschaffene Eidgenössische Kommission für Qualitätssicherung in der medizinischen Begutachtung hat das PRV-CH als Online-Tool implementiert und dabei das von der deutschen Rentenversicherung etablierte Verfahren berücksichtigt^{2,3}.

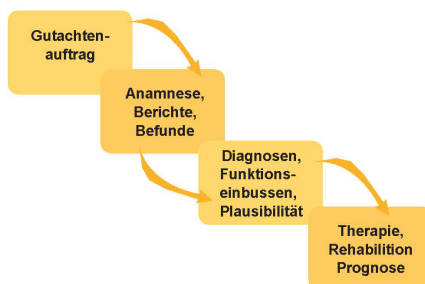
Beschreibung des Instruments zur Beurteilung medizinischer Gutachten

Das Instrument enthält 24 Prüffragen zur Messung von 6 Kriterien.

Jede Prüffrage ist detailliert formuliert und enthält Ankerdefinitionen für die Bewertung auf einer 4-Punkte-Skala.

	Qualitätskriterien
A:	Formale Gestaltung
B:	Verständlichkeit
C:	Transparenz
D:	Vollständigkeit
E:	Fachliche Kompetenzen
F:	Wirtschaftlichkeit

Übergeordnete Nachvollziehbarkeit der Argumentation



Zukünftige Anwendungen des Instruments

- Allgemeine, nicht fallbezogene Qualitätssicherung (repräsentative Stichprobe, Baseline)
- Periodische Überprüfung des aktuellen Standes der Gutachtenqualität (Veränderungsmessung)
- Audits und Akkreditierungsprüfungen von Gutachterstellen (Defizite: Abweichung von der Baseline)
- Aus- und Fortbildung von Gutachtern (Feedback durch einen doppelverblindeten Online-Dienst)

Beispiel einer Prüffrage des PRV-CH

Prüffrage D3	Kriterium D: Vollständigkeit
Inwieweit werden wesentliche Abweichungen in der versicherungsmedizinischen Beurteilung des Leistungsvermögens gegenüber relevanten Vorberichten beschrieben und bewertet?	

Hierbei geht es um abweichende Aussagen zum Leistungsvermögen gegenüber Vorberichten, die eine andere versicherungsrechtliche Entscheidung nach sich ziehen könnten. Es wird erwartet, dass Abweichungen hinsichtlich des quantitativen und qualitativen Leistungsvermögens, von Beginn und Dauer der Leistungsminderung oder des Behandlungsbedarfs in der Beurteilung aufgezeigt und bewertet werden. Die Schlüssigkeit der Bewertung wird hier nicht beurteilt.

Unter Vorberichten wird dabei z. B. verstanden: Reha-Austrittsbericht, Gutachten der Sozialversicherungsträger, Gerichtsgutachten, Gutachten anderer Versicherer, Berichte von Behandlern und Spitalern.

Ergeben sich aus dem Gutachten keine Hinweise auf abweichende Vorberichte, wird hier kein Mangel kodiert.

<input type="checkbox"/>	Kein Mangel	
<input type="checkbox"/>	Leichte Mängel	Es gibt wesentliche Abweichungen in der versicherungsmedizinischen Beurteilung des Leistungsvermögens gegenüber Vorberichten. Diese werden im Gutachten aufgezeigt, es unterbleibt jedoch eine Bewertung und Begründung von Teilaspekten der Abweichungen in der Beurteilung.
<input type="checkbox"/>	Deutliche Mängel	Es gibt wesentliche Abweichungen in der versicherungsmedizinischen Beurteilung des Leistungsvermögens gegenüber Vorberichten. Diese werden im Gutachten aufgezeigt, es unterbleibt jedoch eine Bewertung und Begründung dieser Abweichungen in der Beurteilung .
<input type="checkbox"/>	Gravierende Mängel	Es werden Vorberichte zitiert, aber ohne Aussage zum darin angegebenen Leistungsvermögen. ODER Es gibt wesentliche Abweichungen in der versicherungsmedizinischen Beurteilung des Leistungsvermögens gegenüber Vorberichten, die in der Beurteilung nicht aufgezeigt werden .

Validierung des Instruments

- Projektdurchführung von Mai bis Oktober 2023
- Schulung von 8 unabhängigen Experten (2 Psychiater, 2 Allgemeinmediziner, 2 Neurologen und 2 Rheumatologen)
- 10 anonymisierte Stichproben von Expertenberichten mit 3-5 Disziplinen
- Doppelblindes setting (online tool)
- Bewertung der Wichtigkeit der 24 Prüffragen für die Qualitätsentscheidung
 - Reduzierung der Anzahl der Prüffragen
 - Validierung des Instruments

Fazit

Das schweizerische Peer Review Verfahren (PRV-CH) bietet einen Weg zu einer besseren medizinischen Bewertung in der Sozialversicherung, die zu besseren und gerechteren Ergebnissen für Menschen führt, die auf diese Leistungen angewiesen sind.

Literatur:

- ¹ Schleifer, R., & Liebrez, M. (2023). Introducing the Federal Commission for Quality Assurance in Medical Assessment. *Therapeutische Umschau. Revue thérapeutique*, 80(2), 85-88.
- ² Deutsche Rentenversicherung (2021). Qualitätssicherung der Begutachtung - Manual zum Peer Review [Quality Assurance of Review - Manual for Peer Review]. https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Experten/infos_fuer_aerzte/begutachtung/manual_peer_review.html [accessed 15th September, 2023].
- ³ Strahl A, Gerlich C, Alpers GW, 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 Med Educ*. 18(1):135.

Inhaltsverzeichnis

1. Ausgangslage.....	4
2. Ziele der Untersuchung	6
3. Studiendesign.....	6
4. Auswertungen der Daten und Darstellung der Ergebnisse.....	9
4.1 Schweregrad der Mängel	9
4.2 Einschätzung Wichtigkeit der einzelnen Prüffragen.....	11
4.3 Zeitaufwand	12
4.4 Unterschiede der 23 Prüffragen nach Fachdisziplinen	14
4.5 Faktorenanalyse der 23 Prüffragen	16
4.6 Regressionsanalyse der 23 Prüffragen als Prädiktoren der Gesamtbeurteilung	16
5. Diskussion der Ergebnisse	17
5.1 Akquise von qualifizierten Reviewern (Machbarkeit)	17
5.2 Entwicklung und Durchführung einer ersten Schulung für die Reviewer.....	17
5.3 Erprobung der elektronischen Erfassungsform für die PRV-Vollversion	17
5.4 Genauere Erfassung des Zeitbedarfs der PRV-Vollversion.....	18
5.5 Identifizierung relevanter Prüffragen in der PRV-Kurzversion	19
5.6 Identifizierung der Prüffragen nach Schweregrad der Mängel.....	20
5.7 Entscheidung, ob und wie die PRV-Vollversion gekürzt werden soll.	20
5.8 Einschränkungen aus der PRV-Pilotstudie.....	21
6. Schlussfolgerungen und Ausblick	22

Tabellenverzeichnis

Tabelle 1: Übersicht der verwendeten Prüffragen der PRV-Vollversion und -Kurzversion	5
Tabelle 2: Studiendesign der Pilotstudie zur PRV-Vollversion	7
Tabelle 3: Häufigkeiten der Mängel pro Prüffrage der Konsensbeurteilung	9
Tabelle 4: Häufigkeiten der Mängel pro Prüffrage der Teilgutachten	10
Tabelle 5: Einschätzung der Wichtigkeit der Prüffragen der Konsensbeurteilung	11
Tabelle 6: Einschätzung der Wichtigkeit der Prüffragen der Teilgutachten	11
Tabelle 7: Zeitaufwand in Abhängigkeit der Gesamtbeurteilung	13
Tabelle 8: Zeitaufwand in Minuten nach Fachdisziplin	13
Tabelle 9: Unterschiede der Mängelbeurteilung nach Fachdisziplinen.....	14
Tabelle 10: Unterschiede der Wichtigkeitseinschätzung nach Fachdisziplinen	15
Tabelle 11: Übersicht der verwendeten Prüffragen der PRV-Vollversion und -Kurzversion	23

1. Ausgangslage

Eine Expertengruppe im Auftrag der Eidgenössischen Invalidenversicherung erarbeitete 2020 einen Fragenkatalog zum Peer-Review-Verfahren (PRV) unter der Leitung von Frau Dr. med. Hermelink ausgehend vom Prüfmanual der Deutschen Rentenversicherung (Fassung von 2018). Die Eidgenössische Kommission für Qualitätssicherung in der medizinischen Begutachtung (EKQMB) hat diese Vorarbeiten seit Beginn ihrer Tätigkeit 2022 übernommen, eingehend diskutiert und beschlossen, den Fragenkatalog dieses Verfahrens in einer Pilotstudie 2023 zu validieren, um eine Entscheidungsgrundlage für den Einsatz einer PRV-Kurzversion im Vergleich zur PRV-Vollversion zu erhalten.

Die EKQMB beabsichtigt die PRV-Vollversion für eine repräsentative Zufallsstichprobe von mindestens 100 aus allen 11'000 jährlich in der Schweiz erstellten Gutachten in einem Doppelblind-Verfahren anonymisiert mit unabhängigen, erfahrenen und im PRV geschulten Reviewern ab 2024 zu verwenden, um den Stand der Qualität der medizinischen Gutachten zu messen (IST-Zustand-Messungen)¹ und die Qualitätsentwicklung in den nächsten Jahren zu verfolgen, sowie um allfällige Qualitätsmängel bei Audits von Gutachterstellen zu eruieren und in einem Qualitätsdialog mit den Betroffenen zu einer Verbesserung beizutragen. Das von der EKQMB beabsichtigte PRV soll nicht nur fallunabhängig sein, sondern auch unabhängig vom Auftraggeber (Versicherung) und vom BSV (Vertragspartner der polydisziplinären Gutachterstellen) sein und sich neben formalen Anforderungen, auch stark auf die fachlichen Aspekte der Gutachterqualität konzentrieren.

Andererseits ist geplant, dass die Prüffragen der PRV-Kurzversion im Rahmen der fallbezogenen Qualitätssicherung aller IV-Gutachten von geschulten RAD-Ärzten (Regionale Ärztliche Dienste) nicht anonymisiert eingesetzt werden und so als einheitliches Verfahren der Qualitätsprüfung die zum Teil unterschiedliche Praxis minimieren und so die Rechtsgleichheit sicherstellen. Die PRV-Ergebnisse werden fallbezogen in Versichertendossiers abgelegt, an die jeweiligen Gutachter im Sinne eines Feedbacks gesandt sowie in einer Datenbank² abgespeichert. Diese Datenbank soll zukünftig der Eidgenössischen Kommission für Qualitätssicherung in der medizinischen Begutachtung (EKQMB) zur Verfügung gestellt werden. Die PRV-Kurzversion konzentriert sich auf die formalen Aspekte der Gutachten und deren Verwertbarkeit im Rahmen des IV-Verfahrens.

Während die PRV-Kurzversion im Rahmen der systematischen Kontrollen der RAD-Ärzte für alle Fachdisziplinen vom fallführenden Arzt (i.S. eines versicherungsmedizinischen «Allrounders») durchgeführt wird, sollen die Gutachten im Rahmen der PRV-Vollversion jeweils von

¹ Zunächst geht es hier, um Gutachten der Invalidenversicherung. Für allfällige Anwendungen im UVG-Bereich wären Anpassungen des Verfahrens notwendig.

² Diese Datenbank wird sich auf der Plattform «SuisseMed@p» basieren. Damit werden neben PRV-Ergebnisse auch zusätzliche Daten zur Verfügung stehen.

einem entsprechenden Facharzt beurteilt werden und die Konsensbeurteilungen bei polydisziplinären Gutachten von einem Allgemeinarzt der Inneren Medizin vorgenommen werden, um eine möglichst gute fachliche Überprüfung zu gewährleisten.

Tabelle 1: Übersicht der verwendeten Prüffragen der PRV-Vollversion und -Kurzversion

Kriterium	Frage-Nr.	PRV- Vollversion			PRV- Kurzversion		
		monodisz. Fach-GA	polydisz. Teil-GA	interdisz. Konsens	monodisz. Fach-GA	polydisz. Teil-GA	interdisz. Konsens
Formale Gestaltung	A1	A1	A1	A1			
	A2	A2		A2	A2		A2
Verständlichkeit	B1	B1	B1	B1			
	B2	B2	B2	B2			
	B3	B3	B3	B3			
Transparenz	C1	C1	C1		C1	C1	
	C2	C2	C2				
	C3	C3	C3				
	C4	C4	C4				
	C5	C5	C5				
Vollständigkeit	D1	D1	D1				
	D2	D2	D2		D2	D2	
	D3	D3	D3	D3	D3	D3	D3
	D4	D4	D4	D4	D4	D4	D4
	D5			D5			D5
Fachliche Kompetenzen	E1	E1	E1				
	E2	E2	E2				
	E3	E3	E3		E3	E3	
	E4	E4	E4		E4	E4	
	E5	E5	E5		E5	E5	
	E6	E6	E6	E6	E6	E6	E6
	E7	E7	E7	E7			
Wirtschaftlichkeit	F1	F1	F1				
	F2	F2	F2				
Nachvollziehbarkeit							

Bemerkung: Die Orange markierten Prüffragen wurden von CH-Experten neu entwickelt. Die nicht eingefärbten Prüffragen wurden aus der Version der Deutschen Renten Versicherung (DRV) übernommen.

Die Durchführung der PRV-Vollversion ist im Vergleich zur PRV-Kurzversion etwas umfangreicher (23 vs. 10 Prüffragen bei der Version für monodisziplinäre Gutachten, ferner auch 11 vs. 6 Prüffragen bei Konsensbeurteilung)³. Es muss, relativ gesehen, von einem höheren Personalaufwand ausgegangen werden.

³ Wir schätzen den zusätzlichen Zeitbedarf der PRV-Vollversion gegenüber der PRV- Kurzversion auf 10-15 Min pro Teilgutachten

2. Ziele der Untersuchung

Ziel dieser Pilotstudie bestand darin, die Realisierbarkeit des Vorhabens praktisch zu überprüfen. Insbesondere sollte das Schulungskonzept der Reviewer entwickelt und erprobt werden. Weiter sollte die Online-Version der PRV-Vollversion bereitgestellt und von den Anwendern praktisch erprobt werden. Beabsichtigt war zudem, die allfälligen Vorteile (Mehrwert) der Vollversion (falls diese bestehen) gegenüber der Kurzversion zu untersuchen. Die Ergebnisse der Untersuchung sollten beispielsweise dazu genutzt werden, die Vollversion zu kürzen oder sich für die Kurzversion zu entscheiden⁴.

Ziele der Studie:

- 1) Akquise von qualifizierten Reviewern (Machbarkeit).
- 2) Entwicklung und Durchführung einer ersten Schulung für die Reviewer.
- 3) Erprobung des online Erhebungstool für die PRV-Vollversion (siehe Anhang IV).
- 4) Genauere Erfassung des Zeitbedarfs der PRV-Vollversion zur Berechnung der Unterschiede (Der Zeitbedarf für der PRV-Kurzversion-Version aus RAD-Pilot ist bekannt).
- 5) Gibt es relevante Prüffragen in den 13 weggelassenen Prüffragen zwischen der PRV-Vollversion (24) und der PRV-Kurzversion (10) bzw. in den 5 weggelassenen Prüffragen zwischen der Konsensbeurteilung der PRV-Vollversion (11) und der Konsensbeurteilung der PRV-Kurzversion (6) der bi- und polydisziplinären GA?
- 6) Welche Prüffragen sind wie häufig mangelhaft und wie gravierend⁵? Gibt es möglicherweise «entscheidende» Prüffragen? Gibt es aus Sicht der Reviewer Prüffragen, welche besonders wichtig für die Qualitätsbeurteilung sind? Sind diese Prüffragen in der PRV-Kurzversion enthalten?
- 7) Entscheidung, ob und wie die PRV-Vollversion gekürzt werden soll.

Auf die Bestimmung der Interrater-Reliabilität wurde an dieser Stelle bewusst verzichtet, da 10 Gutachten eine relativ geringe Fallzahl in dieser Pilotstudie darstellt. Insbesondere bei einer niedrigen Übereinstimmung hätte zum Studienzeitpunkt nicht beurteilt werden können, ob diese Ergebnisse auf eine mangelnde Schulung, eine zu geringe Erfahrung der Reviewer mit dem Instrument oder auf etwaige tatsächliche Mängel des Verfahrens zurückzuführen sind.

3. Studiendesign

Es wurde ein Pool aus 10 polydisziplinären (3 bis 5 Disziplinen) Gutachten erstellt. Untersucht wurden dabei die Fachgutachten sowie die Konsensgutachten. Die häufigsten Fachrichtungen der polydisziplinären Gutachten in der Schweiz beinhalten Psychiatrie (24%), Rheumatologie

⁴ Gemäss mündlicher Besprechung mit einem der Entwickler der PRV-Vollversion Herrn Prof. R. Stieglitz am 9.11.2022 wäre eine Kürzung der Vollversion durchaus machbar. Wichtig wäre allerdings die Vergleichbarkeit der Ergebnisse der Kurz- und der Vollversion zu gewährleisten.

⁵ Ausprägung der Mängel (keine leicht, deutlich, gravierend).

(13%), Neurologie (14%) und Allgemeinmedizin (26%). Zusätzlich wurde bei der Gutachtenauswahl auch die Neuropsychologie (6%) berücksichtigt, wobei die neuropsychologischen Gutachten im Rahmen der PRV-Pilotstudie nicht explizit evaluiert wurden⁶.

Es wurde für diese Pilotstudie bewusst keine repräsentative Gutachtenstichprobe verwendet, sondern es wurde eine Vorselektion von mangelhaften Gutachten aus dem Jahr 2022⁷ vorgenommen, um eine genügende Streuung für die Antwortmöglichkeiten auf die Prüffragen zu erhalten. Wenn die Gutachten keine oder nur wenige Mängel aufweisen würden, könnte anhand der Untersuchung nichts über das Instrument der PRV-Vollversion oder der PRV-Kurzversion ausgesagt werden. Bei der Vorauswahl der «hochwertigen» und «mangelhaften» Gutachten wurde auf die Apriori-Beurteilung der Fachstelle EKQMB und einigen Kommissionsmitgliedern abgestellt.

Die Gutachten (siehe Tabelle 2) wurden von je 2 Psychiatern, 2 Rheumatologen, 2 Neurologen und 2 Allgemeinmedizinern evaluiert, wobei die Konsensbeurteilung von den 2 Allgemeinmedizinern vorgenommen wurden. Die Beurteilung erfolgte anhand von anonymisierten Gutachten⁸.

Tabelle 2: Studiendesign der Pilotstudie zur PRV-Vollversion

Gutachten	Disziplinen	Fachrichtung des Reviews und Konsens				
		A (23)	P (23)	R (23)	N (23)	K (11)
PGA-23002	A, P, R, N, NP	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23003	A, P, R	A1, A2	P1, P2	R1, R2		A1, A2
PGA-23010	A, P, R, N	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23011	A, P, R, N	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23012	A, P, R, N	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23017	A, P, R, N	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23018	A, P, R, N	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23019	A, P, N, NP	A1, A2	P1, P2		N1, N2	A1, A2
PGA-23020	A, P, N, ortho, kardio	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2
PGA-23023	A, P, N, ortho	A1, A2	P1, P2	R1, R2	N1, N2	A1, A2

PGA: Polydisziplinäres Gutachten, A: Allgemeinmedizin, P: Psychiatrie, N: Neurologie, R: Rheumatologie; NP: Neuropsychologie, K: Konsensbeurteilung (Die Zahl in Klammern entspricht der Zahl der Prüffragen).

Das orthopädische Teilgutachten wurde von den beiden Rheumatologen beurteilt.

⁶ Die neuropsychologischen Teilgutachten müssen unbedingt mitberücksichtigt werden, da eine PRV-Frage sich explizit auf die Verwendung/ Berücksichtigung von Zusatzuntersuchungen bezieht und ansonsten nicht bewertet werden kann.

⁷ Aktuelle formale Vorgaben des BSV sollen berücksichtigt werden.

⁸ Die Anonymisierung wurde von der Fachstelle EKQMB elektronisch mit Adobe Pro vorgenommen.

Neben der Beurteilung mit der PRV-Vollversion wurden die Gutachten von den Reviewern nach Beendigung der Bearbeitung des Prüfbogens bewertet (mangelhaft, weder noch, gut). Anschliessend wurden die Einzelantworten nochmals durchgegangen und die Bedeutung der einzelnen Prüffragen für die Gesamtbeurteilung mit einer 3-Likert-Skala (wichtig, weder noch, unwichtig) beurteilt.

Mit diesem Design wurde untersucht, ob mit der verkürzten PRV-Kurzversion von nur 10 Prüffragen (Teilmenge der Vollversion) die wichtigen Aspekte bereits erfasst sind bzw. welche Inhalte mit den anderen 14 Prüffragen der Vollversion hervortreten. Bei der Konsensbeurteilung K11 wurden für die PRV-Kurzversion nur 6 Prüffragen verwendet. Auch hier wurde die Frage gestellt, ob die 5 zusätzlichen Prüffragen der Vollversion einen Mehrwert bieten.

Anhand der deskriptiven Ergebnisse (Trends) wurde von den Kommissionsmitgliedern beurteilt, welche Art von Rückmeldungen sie für die Weiterentwicklung der Qualität der medizinischen Gutachten als besonders nützlich und welche sie als unnötig einschätzen.

Für diese Pilotstudie wurden im Februar 2023 die Verträge mit 8 Fachärzten (2 Psychiater, 2 Rheumatologen, 2 Allgemeinmediziner und 2 Neurologen) abgeschlossen, die von April bis Mai 2023 bezüglich aller Aspekte der PRV-Vollversion und bezüglich der Benutzung der Online-Plattform anhand von zwei anonymisierten Gutachten geschult wurden.

Von Anfang Juni bis Ende September 2023 wurden die Beurteilungen der 10 anonymisierten polydisziplinären Gutachten durch die 8 Reviewer vorgenommen.

Die Datensätze der Beurteilung der 10 Gutachten durch die 8 Reviewer konnten dann aus der Online-Plattform exportiert und dann als Excel-Datei zur Verfügung gestellt werden. Dies ermöglichte eine anschliessende Auswertung mit der Statistiksoftware SAS (Version 8.3 Deutsch SAS Enterprise Guide) .

4. Auswertungen der Daten und Darstellung der Ergebnisse

Zuerst wurden deskriptive Methoden der Statistik verwendet, um die Datenlage verständlich zu beschreiben: Häufigkeiten und prozentuale Anteile der Mängel der einzelnen Prüffragen, Median und Mittelwerte des Schweregrades der Mängel und der Wichtigkeit der einzelnen Prüffragen für die Gesamtbeurteilung der Gutachten wurden ermittelt.

Als zweites wurden inferenzanalytische Methoden der Statistik verwendet, um signifikante Unterschiede der Mittelwerte des Schweregrades der Mängel und der Wichtigkeit nach den Kriterien oder Fachdisziplinen (mittels Varianzanalysen) und um signifikante Zusammenhänge (mittels Regressionsanalysen, Faktorenanalysen) zu eruieren und nützliche Informationen zu liefern, die dazu dienen, allfällige Prüffragen zu eliminieren bzw. nicht zu eliminieren.

4.1 Schweregrad der Mängel

Zuerst wurden die prozentualen Anteile der Mängel für jede einzelne Prüffrage bestimmt und aufgelistet: So wurde schnell ersichtlich, bei welchen Prüffragen die meisten gravierenden Mängel festgestellt wurden bzw. bei welchen Prüffragen mehrheitlich kein Mangel beobachtet wurde. Zudem konnte der Schweregrad der Mängel auch mit dem Median oder dem Mittelwert und der Standardabweichung berechnet und für die weiteren differentiellen Analysen der Prüffragen verwendet werden. Die Formulierung der Prüffragen ist im Anhang III dargestellt.

Tabelle 3: Häufigkeiten der Mängel pro Prüffrage der Konsensbeurteilung

	Anzahl der Beurteilungen				Prozentuale Anteile				Median	Mittel-wert
	Kein Mangel =1	Leichte Mängel =2	Deutliche Mängel =3	Gravierende Mängel =4	Kein Mangel =1	Leichte Mängel =2	Deutliche Mängel =3	Gravierende Mängel =4		
A1	13	1	1	5	65%	5%	5%	25%	1	1.9
A2	10	0	2	8	50%	0%	10%	40%	1	2.4
B1	12	6	2	0	60%	30%	10%	0%	1	1.5
B2	6	2	5	7	30%	10%	25%	35%	3	2.7
B3	17	3	0	0	85%	15%	0%	0%	1	1.2
D3	13	2	2	3	65%	10%	10%	15%	1	1.8
D4	10	2	3	5	50%	10%	15%	25%	1	2.2
D5	7	2	0	11	35%	10%	0%	55%	4	2.8
E6	9	2	6	3	45%	10%	30%	15%	2	2.2
E7	12	2	2	4	60%	10%	10%	20%	1	1.9
	grün	gelb	rot		grün	gelb	rot		Median	Mittel-wert
G1	3	8	9		15%	40%	45%		2	2.3

Hellblau eingefärbt sind die Prüffragen der Version für die PRV-Kurzversion

Die Prüffrage *B2* «Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt?» des Kriteriums «Verständlichkeit» ist für das PRV Screening nicht vorgesehen, zeigte jedoch in dieser Pilotstudie bei der Konsensbeurteilung einen überdurchschnittlich hohen Anteil von 25% deutliche und 35% gravierende Mängel.

Die Prüffrage E6 «Werden Konsistenz und Plausibilität beurteilt?» des Kriteriums «Fachliche Kompetenz» zeigte einen hohen Anteil von 30% deutliche und 15% gravierende Mängel bei der Konsensbeurteilung und könnte für die PRV-Kurzversion empfohlen werden.

Tabelle 4: Häufigkeiten der Mängel pro Prüffrage der Teilgutachten

	Anzahl der Beurteilungen				Prozentuale Anteile				Median	Mittelwert
	Kein Mangel =1	Leichte Mängel =2	Deutliche Mängel =3	Gravierende Mängel =4	Kein Mangel =1	Leichte Mängel =2	Deutliche Mängel =3	Gravierende Mängel =4		
A1	64	2	4	8	82%	3%	5%	10%	1	1.4
B1	64	8	5	1	82%	10%	5%	1%	1	1.3
B2	57	9	9	3	73%	12%	12%	4%	1	1.5
B3	68	6	2	2	87%	6%	2%	2%	1	1.2
C1	73	5			94%	6%			1	1.1
C2	44	9	12	13	56%	12%	15%	17%	1	1.9
C3	53	12	10	3	68%	15%	13%	4%	1	1.5
C4	60	6	8	4	77%	8%	10%	5%	1	1.4
C5	38	21	8	11	49%	27%	10%	14%	1	1.9
D1	37	8	15	18	47%	10%	19%	23%	2	2.2
D2	50	14	2	12	64%	18%	3%	15%	1	1.7
D3	53	6	6	13	68%	8%	8%	17%	1	1.7
D4	49	8	5	16	63%	10%	6%	21%	1	1.8
E1	53	3	13	9	68%	4%	17%	12%	1	1.7
E2	69	6	2	1	88%	8%	3%	1%	1	1.2
E3	13	7	3	20	30%	16%	7%	47%	3	2.7
E4	31	14	19	14	40%	18%	24%	18%	2	2.2
E5	40	10	22	6	51%	13%	28%	8%	1	1.9
E6	35	14	18	11	45%	18%	23%	14%	2	2.1
E7	38	12	15	13	49%	15%	19%	17%	2	2.0
F1	71	5	0	2	91%	6%	0%	3%	1	1.1
F2	47	14	9	8	60%	18%	12%	10%	1	1.7
	grün	gelb	rot		grün	gelb	rot		Median	Mittelwert
G1	24	24	30		31%	31%	38%		2	2.1

Hellblau eingefärbt sind die Prüffragen der Version für das PRV-Kurzversion

Die Prüffrage D1 «Inwieweit wird die Befragung vollständig in Bezug auf die Fragestellung dargestellt?» des Kriteriums «Vollständigkeit» zeigte in der Pilotstudie einen hohen Anteil von 19% deutliche und 23% gravierende Mängel bei den Teilgutachten und konnte für die PRV-Kurzversion verwendet werden.

Die Prüffrage E1 «Inwieweit wurde der allgemein anerkannte Stand wissenschaftlicher medizinischer Erkenntnisse gemäss gültigen Leitlinien beachtet?» des Kriteriums «Fachliche Kompetenz» zeigte bei den Teilgutachten einen hohen Anteil von 17% deutliche und 12% gravierende Mängel. Die Ergebnisse dieser Prüffrage sprachen für die hohe fachliche Kompetenz der Reviewer und es stellte sich die Frage, ob diese Prüffrage auch für die PRV-Kurzversion der «Allrounder» der RAD empfohlen werden könnte, da diese die gültigen Leitlinien der verschiedenen Fachgesellschaften bereits jetzt berücksichtigen müssen.

Die Prüffrage E7 «Inwieweit erfolgt eine Darstellung der bisherigen Therapie und eine Stellungnahme zu weiteren therapeutischen Optionen mit Auswirkung auf das arbeitsrelevante Funktionsniveau?» des Kriteriums «Fachliche Kompetenz» zeigte bei den Teilgutachten einen

hohen Anteil von 19% deutliche und 17% gravierende Mängel und sollte für die PRV-Kurzversion ebenfalls empfohlen werden.

Hingegen zeigte die Prüffrage E2 «Macht ein(e) Facharzt/ Fachärztin Beurteilungen/fachliche Wertungen in fremdem Fachgebiet?» nur geringe Mängel. Somit bot es sich an diese Frage für die PRV-Kurzversion nicht zu verwenden.

4.2 Einschätzung Wichtigkeit der einzelnen Prüffragen

Als zweites wurde die Wichtigkeit der einzelnen Prüffragen für die Gesamtbeurteilung der Gutachten beschrieben (1: wichtig, 2: weder noch, 3: unwichtig), sowie die Gesamtbeurteilung (1: mangelhaft, 2: weder noch, 3: gut), welche als prozentuale Anteile dargestellt, respektive als Median und Mittelwerte berechnet und für die weiteren differentiellen Analysen der Prüffragen verwendet wurden.

Tabelle 5: Einschätzung der Wichtigkeit der Prüffragen der Konsensbeurteilung

	Anzahl der Beurteilungen			Prozentuale Anteile			Median	Mittelwert
	wichtig =1	weder noch =2	unwichtig =3	wichtig =1	weder noch =2	unwichtig =3		
A1	18	2	0	90%	10%	0%	1	1.1
A2	12	8	0	60%	40%	0%	1	1.4
B1	18	2	0	90%	0%	0%	1	1.1
B2	16	4	0	80%	20%	0%	1	1.2
B3	17	3	0	85%	15%	0%	1	1.2
D3	20	0	0	100%	0%	0%	1	1.0
D4	20	0	0	100%	0%	0%	1	1.0
D5	20	0	0	100%	0%	0%	1	1.0
E6	20	0	0	100%	0%	0%	1	1.0
E7	19	1	0	95%	5%	0%	1	1.1
G1	20	0	0	100%	0%	0%	1	1.0

Hellblau eingefärbt sind die Prüffragen der Version für die PRV-Kurzversion

E6 und E7 werden auch von allen Reviewern als wichtig erachtet und könnten für die PRV-Kurzversion in der Konsensbeurteilung empfohlen werden.

Tabelle 6: Einschätzung der Wichtigkeit der Prüffragen der Teilgutachten

	Anzahl der Beurteilungen			Prozentuale Anteile			Median	Mittelwert
	wichtig =1	weder noch =2	unwichtig =3	wichtig =1	weder noch =2	unwichtig =3		
A1	36	15	25	47%	20%	33%	2	1.9
B1	55	12	9	72%	16%	12%	1	1.4
B2	43	22	11	57%	29%	14%	2	1.6
B3	50	15	11	66%	20%	14%	1	1.5
C1	35	15	26	46%	20%	34%	2	1.9
C2	61	11	4	80%	14%	5%	1	1.3
C3	55	13	8	72%	17%	11%	1	1.4
C4	60	9	7	79%	12%	9%	1	1.3

C5	67	4	5	88%	5%	7%	1	1.2
D1	64	9	3	84%	12%	4%	1	1.2
D2	33	18	25	43%	24%	33%	2	1.9
D3	63	9	4	83%	12%	5%	1	1.2
D4	61	12	3	80%	16%	4%	1	1.2
E1	63	10	3	83%	13%	4%	1	1.2
E2	49	16	11	64%	21%	14%	1	1.5
E3	63	6	7	83%	8%	9%	1	1.3
E4	68	6	2	89%	8%	3%	1	1.1
E5	59	14	3	78%	18%	4%	1	1.3
E6	66	8	2	87%	11%	3%	1	1.2
E7	63	11	2	83%	14%	3%	1	1.2
F1	52	18	6	68%	24%	8%	1	1.4
F2	60	10	6	79%	13%	8%	1	1.3
G1	75	1	0	99%	1%	0%	1	1.0

Hellblau eingefärbt sind die Prüffragen der Version für das PRV-Kurzversion

Die Prüffragen C2 bis C5 des Kriteriums «Transparenz» werden von den Reviewern als wichtiger beurteilt als die Prüffrage C1 und könnten eher für die PRV-Kurzversion empfohlen werden.

Die Prüffrage D1 des Kriteriums «Vollständigkeit» wird von den Reviewern als wichtiger beurteilt als die Prüffrage D2 und könnte eher für die PRV-Kurzversion empfohlen werden.

E1 und E7 des Kriteriums «Fachliche Kompetenz» werden von den Reviewern als wichtiger beurteilt als E2 und könnten eher für die PRV-Kurzversion empfohlen werden.

4.3 Zeitaufwand

Als drittes wurden die Mittelwerte der Zeit berechnet, welche insgesamt für die Beurteilung des Gutachtens verwendet wurden. Hier gibt es zwei verschiedene Zeitangaben: einerseits für die Fachdisziplin mit 23 Prüffragen und andererseits für die Konsensbeurteilung mit 11 Prüffragen. Es war zu erwarten, dass der Aufwand bei sehr mangelhaften Gutachten viel höher ist als bei sehr guten Gutachten. Deshalb sollte der Zeitaufwand für die drei Gesamtbeurteilungen (1: mangelhaft, 2: weder noch, 3: gut) berechnet werden.

Der arithmetische Mittelwert für den Zeitaufwand aller Teilgutachten betrug 94.9 Minuten und für die Konsensbeurteilungen 94.5 Minuten. Obwohl die Konsensbeurteilungen viel kürzer waren, war es häufig notwendig, die einzelnen Teilgutachten zu konsultieren, was den relativ hohen Zeitaufwand erklärte.

Die Reviewer wurden nach Abschluss der PRV-Beurteilung eines Gutachtens nicht nur zum Zeitaufwand, sondern auch zur Gesamtbeurteilung des Gutachtens als «mangelhaft» «gut» oder «weder noch» befragt. Der Zeitaufwand wurde nun in Bezug auf diese Gesamtbeurteilung gesetzt und zeigte, dass der Zeitaufwand für die Beurteilung von mangelhaften Gutachten deutlich höher war, wie in der Tabelle 7 dargestellt.

Tabelle 7: Zeitaufwand in Abhängigkeit der Gesamtbeurteilung

	Anzahl Gutachten	Mittelwert Zeitaufwand in Min.	Standardabweichung in Min.
«Mangelhafte» Teilgutachten	43	107.4	44.0
«weder mangelhafte noch gute» Teilgutachten	18	76.0	26.0
«gute» Teilgutachten	16	82.8	25.8
«Mangelhafte» Konsensbeurteilungen	13	100.7	35.0
«weder mangelhafte noch gute» Konsensbeurteilungen	6	83	18.6
«gute» Konsensbeurteilungen	1	80	-

Die Varianzanalyse zeigte der Zeitaufwand für die Beurteilung der Teilgutachten nach den drei Gesamtbeurteilungen (1: mangelhaft, 2: weder noch, 3: gut) einen signifikanten Unterschied ($F=5.71$, $p<0.005$). Bei den wenigen Konsensbeurteilungen zeigte die Varianzanalyse keinen signifikanten Unterschied ($F=0.76$, $p=0.48$).

Der Zeitaufwand unterschied sich signifikant ($F=2.67$, $p<0.05$) nach Fachdisziplin, wie die Tabelle 8 zeigt.

Tabelle 8: Zeitaufwand in Minuten nach Fachdisziplin

Fachdisziplin	Anzahl	Mittelwert	SD	Minimum	Maximum
Allgemeine / Innere Medizin	20	95	29	35	150
Neurologie	18	75	28	35	120
Psychiatrie	20	95	33	45	150
Rheumatologie	16	113	59	60	265
Orthopädie	2	135	21	120	150
Total	76	95	40	35	265

4.4 Unterschiede der 23 Prüffragen nach Fachdisziplinen

Die Varianzanalyse zeigte, dass sich die Mittelwerte der Fachdisziplinen in Bezug auf den Schweregrad der Mängel der Prüffragen A1, D1, D2, D3, E1, E7 und F1 signifikant unterscheiden. Es gab keine eindeutige Zuordnung der Mängel nach Fachrichtung. Da nur 10 Gutachten in dieser Pilotstudie beurteilt wurden, konnten die Gründe für diese signifikanten Unterschiede nicht eindeutig den Fachdisziplinen oder den Fachärzten zugeschrieben werden.

Tabelle 9: Unterschiede der Mängelbeurteilung nach Fachdisziplinen

Prüf- frage	Allgemeine / Innere Medi- zin		Neurologie		Psychiatrie		Rheumatolo- gie		F-Wert	p
	M	SD	M	SD	M	SD	M	SD		
A1	2.2	1.4	1.3	0.8	1.1	0.2	1.0	0.0	6.44***	0.0002
B1	1.1	0.3	1.7	0.5	1.4	0.8	1.5	1.0	1.68	0.1647
B2	1.7	1.0	1.7	0.8	1.4	0.8	1.0	0.0	2.18	0.0801
B3	1.2	0.3	1.2	0.5	1.4	0.9	1.0	0.0	1.15	0.3424
C1	1.0	0.0	1.0	0.0	1.2	0.4	1.3	0.3	1.53	0.2023
C2	2.2	1.4	2.0	1.1	1.8	1.0	1.8	1.2	0.44	0.7818
C3	1.3	0.7	1.5	0.7	2.0	0.9	1.5	1.1	1.77	0.1452
C4	1.4	0.9	1.2	0.5	1.9	1.1	1.2	0.8	1.96	0.1096
C5	1.8	1.1	1.8	1.0	2.0	1.2	2.1	1.2	0.23	0.9199
D1	3.0	1.2	1.8	1.1	1.9	1.1	1.9	1.2	3.83**	0.0071
D2	1.6	0.9	1.8	1.2	1.5	0.5	1.9	1.3	4.64**	0.0022
D3	1.2	0.7	1.5	1.0	2.5	1.4	1.6	1.0	4.06**	0.0051
D4	1.7	1.2	1.6	1.2	2.1	1.3	2.1	1.2	0.74	0.5646
E1	1.6	1.2	1.6	1.1	2.8	1.0	1.0	0.0	9.47***	<.0001
E2	1.2	0.5	1.3	0.8	1.2	0.5	1.1	0.3	0.42	0.7953
E3_a	1.5	0.5	1.5	0.5	1.6	0.5	1.4	0.5	0.59	0.6707
E3	2.6	1.4	3.4	0.9	3.1	1.2	1.8	1.3		
E4	1.8	1.2	2.1	1.1	2.6	1.1	2.5	1.1	1.57	0.1924
E5	2.0	1.2	1.7	1.0	2.2	1.0	1.9	1.1	0.51	0.7279
E6	1.7	1.0	2.2	1.3	2.5	1.0	2.1	1.2	1.55	0.1985
E7	1.3	0.7	2.0	1.1	2.4	1.1	2.5	1.3	3.73**	0.0082
F1	1.0	0.0	1.5	1.0	1.0	0.0	1.1	0.3	3.43*	0.0127
F2	1.8	1.2	1.7	1.1	1.7	0.9	1.8	1.1	0.05	0.9946
G	2.0	0.8	2.0	0.9	2.4	0.8	1.9	0.9	1.04	0.3943

*** Signifikanzniveau mit einer Fehlerwahrscheinlichkeit von $p < 0.001$

** Signifikanzniveau mit einer Fehlerwahrscheinlichkeit von $p < 0.01$

* Signifikanzniveau mit einer Fehlerwahrscheinlichkeit von $p < 0.05$

Die Varianzanalyse zeigte, dass sich die Mittelwerte der Fachdisziplinen in Bezug auf die Einschätzung der Wichtigkeit der Prüffragen A1, B3, C1, C2, C3, D1, D2, E1, E2, E5 und E7 signifikant unterscheiden. Insbesondere die Allgemeinmediziner fanden alle Prüffragen wichtig.

Tabelle 10: Unterschiede der Wichtigkeitseinschätzung nach Fachdisziplinen

Prüf- frage	Allgemeine / Innere Medi- zin		Neurologie		Psychiatrie		Rheumatolo- gie		F-Wert	p
	M	SD	M	SD	M	SD	M	SD		
A1	1.2	0.4	2.3	0.8	2.0	0.9	2.1	1.0	6.26***	0.0002
B1	1.1	0.2	1.5	0.6	1.7	0.9	1.5	0.8	2.42	0.0566
B2	1.3	0.5	1.9	0.7	1.8	0.9	1.5	0.8	2.38	0.0599
B3	1.1	0.2	1.8	0.7	1.8	0.9	1.4	0.8	4.31**	0.0035
C1	1.2	0.4	2.5	0.5	1.7	0.8	2.3	1.0	11.3***	<.0001
C2	1.0	0.0	1.4	0.5	1.6	0.8	1.1	0.3	3.52*	0.0112
C3	1.0	0.0	1.5	0.6	1.7	0.9	1.4	0.7	3.21*	0.0177
C4	1.0	0.0	1.4	0.6	1.5	0.8	1.4	0.7	1.87	0.1260
C5	1.0	0.0	1.8	1.1	1.3	0.6	1.4	0.8	1.78	0.1417
D1	1.0	0.0	1.8	1.4	1.0	0.0	1.4	0.7	4.04**	0.0052
D2	1.2	0.4	2.5	0.5	1.7	0.8	2.3	0.9	14.2***	<.0001
D3	1.0	0.0	1.3	0.6	1.2	0.5	1.5	0.7	2.25	0.0720
D4	1.0	0.0	1.4	0.6	1.3	0.6	1.3	0.6	2.13	0.0862
E1	1.0	0.0	1.4	0.5	1.1	0.2	1.5	0.8	3.97**	0.0058
E2	1.0	0.0	1.7	0.7	1.7	0.8	1.8	1.0	4.57**	0.0024
E3_a	1.0	0.0	1.3	0.7	1.4	0.8	1.5	0.8	0.59	0.6707
E3	1.0	0.0	1.3	0.7	1.2	0.6	1.5	0.8	1.75	0.1497
E4	1.0	0.0	1.2	0.4	1.2	0.4	1.3	0.3	0.92	0.4590
E5	1.0	0.0	1.5	0.6	1.2	0.4	1.3	0.6	3.05*	0.0224
E6	1.0	0.0	1.1	0.3	1.4	0.6	1.3	0.6	1.93	0.1150
E7	1.0	0.0	1.6	0.6	1.1	0.2	1.3	0.6	5.22***	0.0010
F1	1.2	0.4	1.6	0.6	1.5	0.7	1.5	0.8	1.50	0.2123
F2	1.1	0.2	1.2	0.4	1.5	0.7	1.5	0.9	1.86	0.1270
G	1.0	0.0	1.0	0.2	1.0	0.0	1.0	0.0	0.80	0.5312

*** Signifikanzniveau mit einer Fehlerwahrscheinlichkeit von $p < 0.001$

** Signifikanzniveau mit einer Fehlerwahrscheinlichkeit von $p < 0.01$

* Signifikanzniveau mit einer Fehlerwahrscheinlichkeit von $p < 0.05$

4.5 Faktorenanalyse der 23 Prüffragen

Es wurden vier Faktorenanalysen gerechnet, für die Prüffragen der Teilgutachten und für die Prüffragen der Konsensbeurteilungen, je für die Mängelbeurteilungen als auch für die Wichtigkeitseinschätzungen. Alle vier Faktorenanalysen der Hauptkomponentenanalysen ergaben, dass alle Prüffragen einen einzigen Faktor bilden und keine Prüffrage «ausreißt» und eliminiert werden müsste, da diese ein deutlich anderes Konstrukt misst. Auch die sehr formale Prüffrage *A1* lud noch relativ hoch auf dem 1. Faktor, was darauf hinweist, dass keine Prüffragen einen eigenen Faktor bilden und etwas ganz anderes messen.

4.6 Regressionsanalyse der 23 Prüffragen als Prädiktoren der Gesamtbeurteilung

Die Bewertung des Schweregrades der Mängel mit den 23 Prüffragen der Teilgutachten und 10 Prüffragen der Konsensbeurteilungen hing mehr oder weniger stark mit den 97 Gesamtbeurteilungen der Gutachten zusammen: Die Regressionsanalyse berechnete die Stärke dieser Zusammenhänge und ermöglichte, zu sehen, welche Prüffragen unbedingt als Prädiktoren der Gesamtbeurteilung belassen werden müssen: Das war nur bei der Prüffrage *G* der Nachvollziehbarkeit der Argumentationskette hoch signifikant der Fall. Leicht signifikant mit einer Fehlerwahrscheinlichkeit von weniger als 5% wurde die Prüffrage *E3* «Würdigt der/die zuständige medizinische Sachverständige beurteilungsrelevante Zusatzdiagnostik integrativ?» bei den Teilgutachten und die Prüffrage *D4* «Inwieweit werden die vom Auftraggeber gestellten versicherungsmedizinischen Fragen inkl. evtl. Zusatzfragen vollständig beantwortet?» bei der Konsensbeurteilung.

Alle anderen Prüffragen wurden in der Regressionsanalyse nicht signifikant und erklären nicht viel mehr Varianz als diese bei den Prüffragen. Das übergeordnete Kriterium „Versicherungsmedizinische Nachvollziehbarkeit des Gutachtens“ war ein Qualitätskriterium, dessen Erfüllung als entscheidend und unbedingt erforderlich angesehen wurde. Im übergeordneten Qualitätskriterium wurde das gesamte Gutachten im Zusammenhang betrachtet und die Nachvollziehbarkeit des Gutachtens anhand der Argumentationskette bewertet. Das zeigte sich nun auch empirisch.

5. Diskussion der Ergebnisse

Die Ziele der Pilotstudie wurden in Kapitel 2 formuliert und werden an dieser Stelle bezüglich der Ergebnisse aus Kapitel 4 nun diskutiert:

5.1 Akquise von qualifizierten Reviewern (Machbarkeit)

Für diese Pilotstudie konnten gemäss den Anforderungen der Ausschreibung (Siehe Anhang I) 8 erfahrene Fachärzte gefunden werden, die auch bereit sind, als unabhängige Reviewer weiterzuarbeiten. Die Pilotstudie wurde nur auf Deutsch durchgeführt. Im Hinblick auf die Durchführung der PRV-Vollversion im Jahre 2024 bei einer repräsentativen Stichprobe aller im Jahre 2023 durchgeführten Gutachten, müssten auch Französisch sprachige oder zweisprachige Reviewer rekrutiert werden. Ca 5% der Gutachtenstichprobe wird italienisch sein. Fast alle Tessiner Fachärzte haben an einer deutschsprachigen oder französischsprachigen Universität studiert und sind meistens perfekt zweisprachig.

5.2 Entwicklung und Durchführung einer ersten Schulung für die Reviewer

Die Schulung der 8 Fachärzte wurde gemäss Plan in Anhang II zwischen dem 4. April mit einer Online-Einführung und Ende Mai mit der Schlussbesprechung durchgeführt. Sehr nützlich erwies sich die Präsenzs Schulung, in der die PRV-Beurteilung eines zuvor anonymisierten und eingereichten Gutachtens konkret behandelt wurde. Dabei konnte das Verständnis einzelner Prüffragen und deren Abgrenzung von anderen Prüffragen sowie die jeweiligen Ankerdefinitionen des Schweregrades der Mängelbeurteilung geklärt werden. Nach der online durchgeführten PRV-Evaluation eines zweiten anonymisierten Gutachtens wurden mit den Reviewern die Differenzen in der Beurteilung einzelner Prüffragen geklärt. Die inhaltlichen Beschreibungen sowohl der Kriterien als auch der Ankerdefinitionen erwiesen sich als sehr nützlich und genügend präzise. Von den Reviewern wurden keine Umformulierungen oder Ergänzungen der Kriterien oder der Ankerdefinitionen verlangt.

5.3 Erprobung der elektronischen Erfassungsform für die PRV-Vollversion

Bereits während der Online-Einführung der Schulung vom 4. April wurde das Online-Tool zur elektronischen Erfassung der PRV-Vollversion (siehe Anhang IV) vorgestellt und danach per E-Mail der Zugang mit einem persönlichen Passwort ermöglicht. Lediglich eine der acht Personen hatte Schwierigkeiten beim Online-Zugriff, welche jedoch mit dem Administrator geklärt werden konnten.

Die Online-Plattform zur elektronischen Erfassung der PRV-Vollversion funktionierte während der ganzen Pilotstudie einwandfrei und zur vollen Zufriedenheit aller Beteiligten.

Die zu beurteilenden Gutachten wurden von der Fachstelle EKQMB mit Adobe Pro anonymisiert und über die gesicherte Verbindung des Bundesamtes für Informatik via Filetransfer Service übermittelt. Auch dieser Prozess funktionierte einwandfrei.

5.4 Genauere Erfassung des Zeitbedarfs der PRV-Vollversion

Im Rahmen der im Jahr 2021 durchgeführten RAD-Pilotstudie von Dr. Hermelink betrug der durchschnittliche Zeitaufwand für die PRV-Kurzversion mit zehn Prüffragen 91 Minuten pro Fach- oder Teilgutachten, wobei allein die Lektüre des Gutachtens durchschnittlich 48 Minuten in Anspruch nahm. Gemäss Kapitel 4.3 lag der arithmetische Mittelwert für den Zeitaufwand aller Teilgutachten in dieser Pilotstudie bei 94.9 Minuten und war somit mit 24 Prüffragen nicht signifikant höher als bei der Screening-Kurzform mit 10 Prüffragen.

Obwohl die Erfahrung mit dem PRV-Review eine potenzielle Reduzierung des Zeitaufwands ermöglichen könnte, bleibt der Aufwand, insbesondere bei mangelhaften Gutachten, die eine ausführliche und verständliche Evaluation der Mängel erfordern, beträchtlich. Die Pilotstudie zeigte, dass der durchschnittliche Zeitaufwand für das PRV-Review bei mangelhaften Teilgutachten 107 Minuten und bei Konsensbeurteilungen 100 Minuten betrug, im Vergleich zu 80 Minuten bei qualitativ guten Gutachten.

Für ein polydisziplinäres Gutachten mit vier Disziplinen ist ein durchschnittlicher Zeitaufwand von etwa 450 Minuten, also 7,5 Stunden, anzusetzen, was einem ganzen Arbeitstag entspricht, wenn die inhaltliche Qualität seriös und professionell analysiert und den Gutachtern qualifizierte Rückmeldungen gegeben werden sollen.

Es erscheint plausibel, dass die Rahmenbedingungen einen signifikanten Einfluss auf die Durchführung der PRV-Reviews haben könnten, möglicherweise sogar einen grösseren als die Anzahl der Prüffragen. Wenn den Ärzten ausreichend Zeit zur Verfügung gestellt wird und die Stellenprozentage bei der landesweiten Einführung der PRV-Kurzversion zur einheitlichen Beurteilung aller Gutachten angemessen angepasst werden, könnte dies eine gründliche und umfassende inhaltliche Prüfung der externen Gutachten fördern.

Um alle 11'000 Gutachten pro Jahr, darunter etwa 5'000 multidisziplinäre, 2'000 bidisziplinäre und 4'000 monodisziplinäre Gutachten zu bewerten, müsste die Arbeitszeit von etwa 34 Vollzeitärzten pro Jahr veranschlagt werden. Da die RAD-Ärzte die eingegangenen Gutachten bereits lesen und bewerten müssen, wird geschätzt, dass die systematische Einführung von PEP in allen RAD der IV-Stellen in der Schweiz eine durchschnittliche Personalaufstockung

um etwa 8 Vollzeitstellen landesweit erfordern wird, um die Durchführung der PEP-Bewertungen unter guten Bedingungen zu gewährleisten.

5.5 Identifizierung relevanter Prüffragen in der PRV-Kurzversion

In Kapitel 4.1 wurden die Beurteilungen des Schweregrades der Mängel für jede Prüffrage dargestellt (siehe Tabellen 3 und 4), und Kapitel 4.2 widmet sich der von den Reviewern eingeschätzten Wichtigkeit jeder einzelnen Prüffrage für die Gesamtbeurteilung der Qualität eines Gutachtens (siehe Tabellen 5 und 6). In diesen vier Tabellen sind die Prüffragen der PRV-Kurzversion hellblau hervorgehoben, um schnell zu identifizieren, ob es relevante Prüffragen in den 14 weggelassenen Prüffragen zwischen der PRV-Vollversion (24) und der PRV-Kurzversion (10) sowie in den 5 weggelassenen Prüffragen zwischen der Konsensbeurteilung der PRV-Vollversion (11) und der Konsensbeurteilung der PRV-Kurzversion (6) gibt.

Es hat sich in dieser Pilotstudie gezeigt, dass mit den bisher in der PRV-Kurzversion weggelassenen Prüffragen D1, E1 und E7 mehr gravierende Mängel identifiziert wurden als mit den bisherigen Screening-Prüffragen D2 und E2 der Teilgutachten. Die genaue Formulierung der Prüffragen ist in Anhang III zu finden.

Bei der Konsensbeurteilung wurden mit der Prüffrage B2 deutlich häufiger gravierende Mängel (Mittelwert von 2.7) festgestellt als bei den Teilgutachten der Allgemeinmedizin durch dieselben Reviewer (Mittelwert von 1.7), was darauf hinweist, dass die Anforderungen an die Verständlichkeit bei der Konsensbeurteilung viel höher sind als bei den Teilgutachten.

Auch bezüglich der Einschätzung der Wichtigkeit werden die Prüffragen D1, E1 und E7 als wichtiger eingestuft als die Prüffrage E2. Ferner werden die Prüffragen C2 bis C5 des Kriteriums Transparenz von den Reviewern als wichtiger bewertet als die Prüffrage C1.

In ihrer Sitzung vom 25. Oktober 2023 hat die Kommission diesen Bericht besprochen. Bezüglich der Prüffrage E1 wurde angemerkt, dass die Ärzte der RAD oft nicht alle Fachdisziplinen abdecken, sondern häufig Allgemeinmediziner sind, weshalb E1 nicht in die PRV-Kurzversion aufgenommen werden soll. Die Kommission hat sich für die Beibehaltung der Prüffrage D2 in der PRV-Kurzversion ausgesprochen, da sie die Kodierung der Diagnosen nach der ICD oder dem DSM als wichtiges Qualitätsmerkmal erachtet. Schliesslich hat die Kommission beschlossen, für die PRV-Kurzversion die Fragen D1 und E7 hinzuzufügen und E2 zu streichen.

5.6 Identifizierung der Prüffragen nach Schweregrad der Mängel

Zu Beginn der Pilotstudie wurde zudem die Frage aufgeworfen: „Existieren möglicherweise entscheidende Prüffragen? Und welche dieser Fragen werden von den Reviewern als besonders wichtig für die Qualitätsbeurteilung angesehen?“ Zudem wurde untersucht, ob diese Prüffragen in der PRV-Kurzversion vertreten sind. Wie bereits erwähnt, werden die Mängel aller Prüffragen in den Tabellen 3 und 4 detailliert dargestellt. Besonders auffällig ist die Prüffrage D5 „Gibt es bei interdisziplinären Gutachten eine echte integrative Beurteilung?“, bei der in der Konsensbeurteilung 55% gravierende Mängel festgestellt wurden; diese Frage ist bereits in der aktuellen PRV-Kurzversion berücksichtigt. Ebenfalls signifikante Mängel zeigt die Prüffrage E3 „Würdigt der/die zuständige medizinische Sachverständige beurteilungsrelevante Zusatzdiagnostik integrativ?“, die in den Teilgutachten mit 47% gravierenden Mängeln auffällt und ebenfalls in der PRV-Kurzversion enthalten ist. Laut den Reviewern sind nahezu alle Prüffragen von Bedeutung, allerdings erscheinen die formalen Fragen aus den Kriterien A, B und C als etwas weniger wichtig.

5.7 Entscheidung, ob und wie die PRV-Vollversion gekürzt werden soll.

Aus Sicht der Reviewer gab es keine Prüffrage, die für die Qualitätsbeurteilung der medizinischen Gutachten als unwichtig angesehen wurde.

Es gab jedoch drei eher formale Prüffragen A1, C1 und D2, welche nur zu 30% als «wichtig», respektive zu 30% als unwichtig erachtet wurden (Siehe Tabelle 6):

- A1 «Inwieweit werden die Gliederungsvorgaben inkl. allfälliger Vorgaben zur interdisziplinären Gesamtbeurteilung (Konsensbeurteilung) gemäss den Anforderungen eingehalten?» des Kriteriums Formale Gestaltung zu 36% als «wichtig» und zu 33% als «unwichtig» betrachtet.
- C1 «Ist die Untersuchungszeit (Datum, Uhrzeit von – bis) dokumentiert?» des Kriteriums Transparenz zu 35% als «wichtig» und zu 34% als «unwichtig» betrachtet.
- D2 «Inwieweit sind die Diagnosen nach einem anerkannten Diagnosesystem verschlüsselt?» des Kriteriums Vollständigkeit zu 43% als «wichtig» und zu 33% als «unwichtig» betrachtet.

Als Argument für die Beibehaltung dieser drei weniger wichtigen Prüffragen wurde in den Diskussionen der Kommission im Jahr 2022 angeführt, dass diese formale Beurteilung sehr schnell und einfach durchgeführt werden kann und keine signifikante Reduktion des Aufwands bedeutet. Da für diese formale Beurteilung kein Spezialwissen erforderlich ist, wurde auch vorgeschlagen, dass die Bewertung dieser formalen Kriterien an eine geschulte Fachperson delegiert werden kann.

Bezüglich des Beschlusses der Kommission vom Juni 2023, den Qualitätsindikator „Ist die Untersuchungszeit angemessen?“ einzuführen, ist es natürlich erforderlich, dass die Untersuchungsdauer angegeben ist und vom Reviewer zur Kenntnis genommen wurde. In diesem

Kontext kann an der Prüffrage C1 festgehalten werden. Die Beurteilung der Angemessenheit der Untersuchungszeit erfordert jedoch fachärztliche Kompetenz und sollte unter dem Kriterium E neu formuliert werden. Andererseits deutet eine unangemessen kurze Untersuchungszeit auf Mängel in der Vollständigkeit der Anamnese und der Abklärung der Befunde hin.

An Ihrer Sitzung vom 25. Oktober 2023 hat die EKQMB beschlossen, für die PRV-Vollversion alle Prüffragen zu verwenden und den Fragenkatalog nicht zu kürzen.

5.8 Einschränkungen aus der PRV-Pilotstudie

In dieser Pilotstudie wurden zehn Gutachten evaluiert, was bedeutet, dass die ermittelten signifikanten Unterschiede in der Mängelbeurteilung und der Wichtigkeitseinschätzung nicht eindeutig den Teilgutachten oder den Fachärzten zugeordnet werden können. Zudem wurde bewusst auf die Berechnung der Interrater-Reliabilität verzichtet, um den initialen Fokus auf die erwähnten qualitativen Aspekte zu legen.

Trotz dieser Einschränkungen bietet die Studie wertvolle Erkenntnisse: Der Einsatz der PRV-Vollversion und der PRV-Kurzversion hat gezeigt, dass das Instrument grundsätzlich funktioniert, auch wenn der zeitliche und finanzielle Aufwand für dessen Durchführung erheblich ist. Durch breit angelegte Schulungen kann jedoch mit einer Steigerung der Effizienz in zukünftigen Anwendungen gerechnet werden.

Ein weiterer Aspekt ist die Identifikation von Engpässen bei den Reviewern (hier auch in Hinblick auf die unterschiedlichen Sprachregionen), insbesondere bei Psychiatern, die aufgrund ihrer hohen Arbeitslast zu Verzögerungen in der Bearbeitung des PRV führen können, wodurch der Prozess (insgesamt) verlangsamt wird.

6. Schlussfolgerungen und Ausblick

Die Pilotstudie wurde erfolgreich abgeschlossen, wobei die Schulung effektiv implementiert und erfolgreich durchgeführt wurde. Die elektronische Online-Plattform erwies sich als robust und funktionierte ohne nennenswerte technische Schwierigkeiten. Die Reviewer fanden die Ankerdefinitionen der Prüffragen klar und hilfreich, mit nur minimalen Anpassungen während der Schulung notwendig.







Hinsichtlich der Anpassung des Fragebogenkatalogs, obwohl nur wenige formale Prüffragen identifiziert wurden, die keine bedeutende Reduzierung der Bearbeitungszeit ermöglichten, entschied die Kommission in ihrer Sitzung vom 25. Oktober 2023, alle Prüffragen beizubehalten. Ferner wird die PRV-Kurzversion durch die Hinzufügung der Prüffragen D1 und E7 bereichert und durch die Streichung der Prüffrage D2 gestrafft. Bezüglich des zukünftigen Einsatzes des PRV-Verfahrens: Der Zeitaufwand für die Beurteilung eines einzelnen polydisziplinären Gutachtens bleibt beträchtlich. Für die umfassende Bewertung einer Zufallsstichprobe von 100 PGAs sind Gesamtkosten von etwa 200.000 CHF zu veranschlagen. Des Weiteren belaufen sich die Kosten für die Durchführung einer PRV-Beurteilung von 50 PGAs im Rahmen von Audits auf ungefähr 100.000 CHF.

Angesichts der Entscheidung der Kommission, sechs Qualitätsindikatoren zu erheben, ist es geplant, die entsprechenden Prüffragen A2, C1, D3, E5 und E6 des PRV dahingehend anzupassen, dass sie die Formulierung der Ankerdefinitionen des Schweregrades der Mängel umfassen. Diese fünf Prüffragen sollen zudem in der PRV-Kurzversion erhoben werden.

Weiterhin hat die Kommission folgende Schritte vorgesehen:

- Die Publikation des Pilotberichts sowie des Fragenkatalogs der PRV-Vollversion, um Transparenz und Zugänglichkeit der Ergebnisse zu gewährleisten.
- Die Empfehlung zur Nutzung der PRV-Vollversion zur internen Qualitätssicherung der Gutachterstellen und zu Ausbildungszwecken für neue Gutachter, um die Kompetenz in diesem Bereich kontinuierlich zu fördern.
- Die Empfehlung der Nutzung der PRV-Kurzversion an das Bundesamt für Sozialversicherungen (BSV) und die IV Stellen, um eine standardisierte und effiziente Beurteilung zu ermöglichen.
- Die Anstellung von zusätzlichen Reviewern durch die EKQMB, um auch die Gutachten in Französisch und Italienisch adäquat beurteilen zu können, was die sprachliche Abdeckung erhöht.
- Die Beurteilung einer repräsentativen Zufallsstichprobe von jährlich 100 Gutachten, um eine kontinuierliche Qualitätskontrolle und -verbesserung zu gewährleisten.
- Die Durchführung von Audits bei Zulassungsverfahren oder bei Hinweisen auf gravierende Mängel, um die Integrität der Gutachterprozesse zu sichern.

Tabelle 11: Übersicht der verwendeten Prüffragen der PRV-Vollversion und -Kurzversion

Kriterium	Frage-Nr.	PRV- Vollversion			PRV- Kurzversion		
		monodisz. Fach-GA	polydisz. Teil-GA	interdisz. Konsens	monodisz. Fach-GA	polydisz. Teil-GA	interdisz. Konsens
Formale Gestaltung	A1	A1	A1	A1			
	A2	A2		A2	A2		A2
Verständlichkeit	B1	B1	B1	B1			
	B2	B2	B2	B2			
	B3	B3	B3	B3			
Transparenz	C1	C1	C1		C1	C1	
	C2	C2	C2				
	C3	C3	C3				
	C4	C4	C4				
	C5	C5	C5				
Vollständigkeit	D1	D1	D1		D1	D1	
	D2	D2	D2				
	D3	D3	D3	D3	D3	D3	D3
	D4	D4	D4	D4	D4	D4	D4
	D5			D5			D5
Fachliche Kompetenzen	E1	E1	E1				
	E2	E2	E2				
	E3	E3	E3		E3	E3	
	E4	E4	E4		E4	E4	
	E5	E5	E5		E5	E5	
	E6	E6	E6	E6	E6	E6	E6
	E7	E7	E7	E7			
Wirtschaftlichkeit	F1	F1	F1				
	F2	F2	F2				
Nachvollziehbarkeit							

Bemerkung: Die Orange markierten Prüffragen wurden von CH-Experten neu entwickelt. Die nicht eingefärbten Prüffragen wurden aus der Version der Deutschen Renten Versicherung (DRV) übernommen.

Anhang 1 Anforderungsprofil der Reviewer

Reviewer I (Allgemeine Innere Medizin), PRV Verfahren

- Anforderungen gemäss Art. 7m ATSV, Abs. 1, Buchstaben a, b, c:
 - a) Weiterbildungstitel nach Artikel 2 Absatz 1 Buchstaben b und c (Medizinalberufeverordnung vom 27. Juni 2007);
 - b) Eintragung im Register MedReg nach Artikel 51 Absatz 1 (Medizinalberufegesetz vom 23. Juni 2006);
 - c) Eine gültige Berufsausübungsbewilligung in der Schweiz.
- Mindestens 10 Jahre versicherungsmedizinische und/oder gutachterliche Erfahrungen
- Fachärztinnen und Fachärzte der allgemeinen inneren Medizin bzw. vergleichbare Ausbildung.
- Zertifikat des Vereins Versicherungsmedizin Schweiz (Swiss Insurance Medicine, SIM) oder Chefärztinnen und Chefärzte sowie leitende Ärztinnen und Ärzte in Universitätskliniken und A-Spitälern.
- Sehr gute Kenntnisse von mindestens zwei Landessprachen.
- Bereitschaft zu einer mindestens dreijährigen Mitarbeit (wenn möglich bis zu 12 Jahren) als Peer-Reviewer in Auftrag der EKQMB.
- Aufwand 30 bis 60 Stunden pro Jahr, flexible Zeiteinteilung mit marktüblichen finanziellen Entschädigungen.

Reviewer II (Fachreviewer), PRV Verfahren

- Anforderungen gemäss Art. 7m ATSV, Abs. 1, Buchstaben a, b, c:
 - d) Weiterbildungstitel nach Artikel 2 Absatz 1 Buchstaben b und c (Medizinalberufeverordnung vom 27. Juni 2007);
 - e) Eintragung im Register nach Artikel 51 Absatz 1 (Medizinalberufegesetz vom 23. Juni 2006);
 - f) Eine gültige Berufsausübungsbewilligung in der Schweiz.
- Mindestens 10 Jahre versicherungsmedizinische und/oder gutachterliche Erfahrungen
- Fachärztinnen und Fachärzte für Psychiatrie und Psychotherapie, Neurologie, Rheumatologie.
- Zertifikat des Vereins Versicherungsmedizin Schweiz (Swiss Insurance Medicine, SIM) oder Chefärztinnen und Chefärzte sowie leitende Ärztinnen und Ärzte in Universitätskliniken und A-Spitälern.
- Sehr gute Kenntnisse von mindestens zwei Landessprachen.
- Bereitschaft zu einer mindestens dreijährigen Mitarbeit (wenn möglich bis zu 12 Jahren) als Peer-Reviewer in Auftrag der EKQMB.
- Aufwand 30 bis 60 Stunden pro Jahr, flexible Zeiteinteilung mit marktüblichen finanziellen Entschädigungen.

Anhang II: Durchführung der Schulung der Peers

Teilnehmer: 8 / Kursleiter: 1 (+1 Assistenz)

- 1) Zusendung der Unterlagen (nach Erhalt des unterschriebenen Vertrages): Studium des PRV-Manuals und der Prüffragebögen (ca. 2 Stunden).
- 2) Online Schulung (4. April 2023) zum PRV-Konzept, zum Manual, zu den PRV-Prüffragen, sowie zum Online-Erhebungstool (ca. 3 Stunden).
- 3) Zusendung eines anonymisierten polydisziplinären Gutachtens per Inca-Mail: vorgängiges Studium des Gutachtens (ca. 2 Stunden).
- 4) Vor Ort (17. April 2023 in Bern* und 22. Mai in Zürich): Besprechung der Mängel des Gutachtens und ausfüllen der Prüffragen in 2er-Teams und Besprechung von einzelnen Prüffragen anhand der festgestellten Mängel mit Kursleiter (ca. 3 Stunden).
- 5) Zusendung eines anonymisierten polydisziplinären Gutachtens per Inca-Mail, Erstellung des Online-Zugriffs und Durchführung der PRV-Evaluation des polydisziplinären Gutachtens via Online-Tool (ca. 4 Stunden).
- 6) Schriftliches Feedback (Excel-Liste) der 8 Ratings mit der Auflistung aller festgestellten Mängel: Online Besprechung der Diskrepanzen von einzelnen Prüffragen pro Teildisziplin und der Konsensbeurteilung durch die Allgemeinärzte (ca. 2 Stunden): Praktisch alle Differenzen konnten geklärt und bereinigt werden. Die klärenden Gespräche führten zu einem besseren Verständnis der Inhalte der Prüffragen und der Ankerdefinitionen. Es wurden von den Fachärzten dabei keine Umformulierungen oder zusätzliche Präzisierungen gewünscht.

Zeitbedarf ca.16 Stunden

Kostenschätzung: 28'160 CHF

Vorgesehen ist für das Pilotprojekt ein Kurs auf Deutsch mit 8 Teilnehmer (4 Disziplinen)

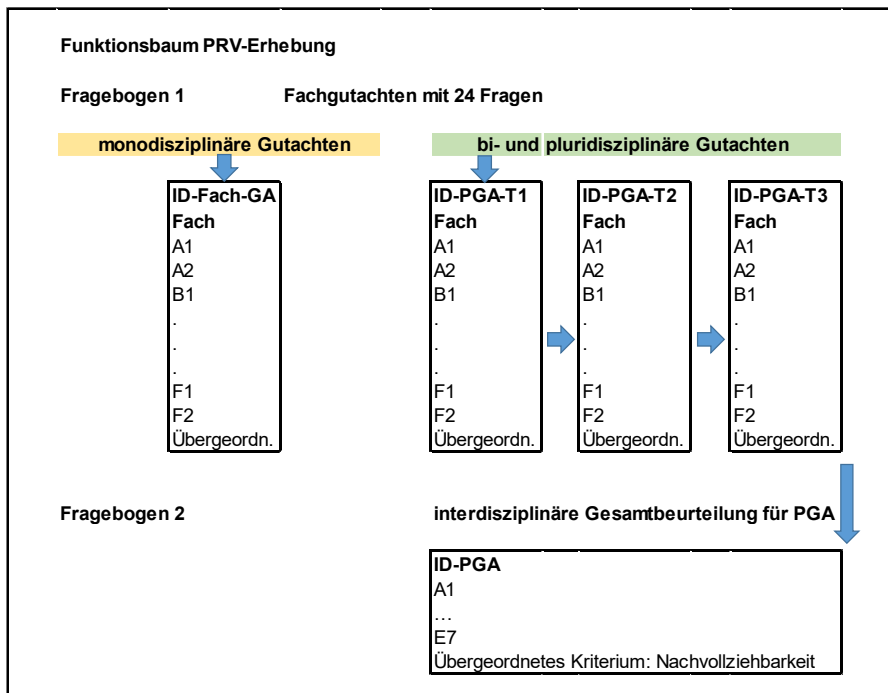
Anhang III: PRV-Fragekatalog der Pilotstudie 2023

Formale Gestaltung	A1	Inwieweit werden die Gliederungsvorgaben inkl. allfälliger Vorgaben zur interdisziplinären Gesamtbeurteilung (Konsensbeurteilung) gemäss den Anforderungen eingehalten?
	A2	Sind die Fristen zwischen Untersuchungszeitpunkt und Gutachteneingang bei der IV angemessen?
Verständlichkeit	B1	Ist das Gutachten sprachlich korrekt und unmissverständlich verfasst?
	B2	Inwieweit sind Fachausdrücke und Abkürzungen, die für das Verstehen des Gutachtens wesentlich sind, allgemeinverständlich erklärt?
	B3	Inwieweit wird die versicherungsmedizinische Terminologie richtig angewendet?
Transparenz	C1	Ist die Untersuchungszeit (Datum, Uhrzeit von – bis) dokumentiert?
	C2	Inwieweit sind Herkunft, zeitliche Einordnung und Kernaussagen von medizinischen Informationen erkennbar?
	C3	Inwieweit wird benannt, mit welchen Untersuchungsverfahren die Befunde erhoben werden?
	C4	Inwieweit werden die Informationsquellen bei der eigenen Befunderhebung unterschieden?
	C5	Inwieweit wird dargestellt, auf welcher Grundlage die Beurteilung der Untersuchungsbefunde erfolgte?
Vollständigkeit	D1	Inwieweit werden die Befragung und die Befunderhebung vollständig in Bezug auf die Komplexität der Problematik dargestellt?
	D2	Inwieweit sind die Diagnosen nach einem anerkannten Diagnosesystem verschlüsselt?
	D3	Inwieweit werden wesentliche Abweichungen in der versicherungsmedizinischen Beurteilung der Arbeitsfähigkeit gegenüber relevanten Vorberichten beschrieben und bewertet?
	D4	Inwieweit werden die vom Auftraggeber gestellten versicherungsmedizinischen Fragen inkl. evtl. Zusatzfragen vollständig beantwortet?
	D5	Gibt es bei interdisziplinären Gutachten eine echte integrative Beurteilung?
Fachliche Kompetenz	E1	Inwieweit wird der allgemein anerkannte Stand wissenschaftlicher medizinischer Erkenntnisse gemäss gültigen Leitlinien beachtet?
	E2	Macht ein(e) Facharzt/ Fachärztin Beurteilungen/fachliche Wertungen in fremdem Fachgebiet?
	E3	Würdigt der/die zuständige medizinische Sachverständige beurteilungsrelevante Zusatzdiagnostik integrativ?
	E4	Werden Funktionsdefizite und dokumentierte Befunde in Bezug zu den gestellten Diagnosen gesetzt?
	E5	Sind personenbezogene und psychosoziale Faktoren berücksichtigt?
	E6	Wird die gutachterliche Beurteilung der Konsistenz und Plausibilität nachvollziehbar begründet?
	E7	Inwieweit erfolgt eine Darstellung der bisherigen Therapie und eine Stellungnahme zu weiteren therapeutischen Optionen mit Auswirkung auf das arbeitsrelevante Funktionsniveau?
Wirtschaftlichkeit	F1	Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der versicherungsmedizinischen Fragestellung geeignet und notwendig?
	F2	Inwieweit ist die eingesetzte Diagnostik für die Beantwortung der versicherungsmedizinischen Fragestellung ausreichend?
Nachvollziehbarkeit	G1	Bewerten Sie die Nachvollziehbarkeit des Gutachtens anhand der Argumentationsführung

Anhang IV: Systemanforderungen: Online Erhebungs-Tool für PRV

1. Grobanforderungen

Der Prüfbogen «Fachgutachten» muss von allen Gutachtern mit der Angabe der betreffenden Fachdisziplin angegeben werden. Der Prüfbogen «interdisziplinäre Gesamtbeurteilung» muss nur vom beauftragten Generalisten der Allgemeinen Inneren Medizin polydisziplinären Gutachten ausgefüllt werden. Bei bidisziplinären GA, ist es einer der beiden Fachärzte, welcher die Gesamtbeurteilung vornimmt.



Es gibt eine Liste von ausgebildeten Reviewern, die für spezifische bi- oder polydisziplinäre Gutachten zuständig sind. Die Fachärzte der Allgemeinen Inneren Medizin können sowohl ihr Fachgutachten als auch die «interdisziplinäre Gesamtbeurteilung» bei bi- und polydisziplinären Gutachten als Generalisten ausfüllen.

2. Funktionale Kurzbeschreibung

Login mit Passwort pro Reviewer (CH-Login).

Das online Erhebungs-Tool dient den externen Reviewern zur sicheren Beurteilung der anonymisierten poly-, bi- und monodisziplinären Gutachten. Ein monodisziplinäres Gutachten wird in einer Fachdisziplin erstellt. Poly- und bidisziplinäre Gutachten bestehen aus mehreren Teilgutachten verschiedener Disziplinen und aus einer interdisziplinären Konsensbeurteilung.

Es gibt zwei Prüfbogen-Typen: einen Prüfbogen für monodisziplinäre Gutachten mit 24 Prüffragen bzw. Teilgutachten mit 23 Prüffragen und einen Prüfbogenbogen für die interdisziplinäre Konsens-Beurteilung mit 11 Fragen.

Die Fragen bestehen aus einer umfangreichen Anleitung (ca. 2'000 Zeichen) und Antworten auf einer 4-stufigen Skala (1 = Kein Mangel, 2 = Leichte Mängel, 3 = Deutliche Mängel und 4 = Gravierende Mängel) und aus einer Volltext-Eingabe bei den Stufen 2 bis 4 pro Frage (Beschreibung der Mängel).

Die Antworten werden sukzessive in eine Datenbank abgespeichert, auf die der betreffende Reviewer jederzeit zugreifen kann (read and write bezüglich seiner Beurteilungen).

Die Fachstelle EKQMB hat Zugriff auf die Datenbank und kann die Beurteilungen der Reviewer extrahieren und auswerten.

Das Tool wurde zuerst in deutscher Sprache erstellt und wird auch auf Französisch und Italienisch für die entsprechenden Gutachten und für die Benutzer entwickelt, um alle Landessprachen abzudecken.

3. Verarbeitungsschritte / -regeln

1. Erfassung und Zugangskontrolle eines externen Reviewers.
 2. Der Reviewer darf nur seine Beurteilungen sehen und ändern (read and write), aber nicht diejenigen der anderen Reviewer. Ein Reviewer kann mehrere Gutachten beurteilen.
 3. Jedes Gutachten hat eine ID-Nummer und muss als Erstes erfasst und abgespeichert werden. Jede Antwort muss sukzessive in einer DB gespeichert werden und es sollte dem Reviewer jederzeit möglich sein, seine früheren Antworten zu lesen und zu ändern.
 4. In der Regel wird ein Gutachten, ob polydisziplinär oder monodisziplinär, jeweils von einem Reviewer beurteilt. In einigen Fällen können die einzelnen fachspezifischen Teilgutachten von einem Reviewer des entsprechenden Faches beurteilt werden.
 5. Die Beurteilung eines Gutachtens kann erst abgeschlossen werden, wenn alle Fragen auf einer 4-stufigen Skala (1. Kein Mangel, 2. Leichte Mängel, 3. Deutliche Mängel und 4. Gravierende Mängel) und mit einer minimalen Volltext-Eingabe (Beschreibung der Mängel) bei den Stufen 2 bis 4 pro Frage erfasst wurden: letzter Kontrollschritt.
 6. Das LOGIN-System basiert in einer ersten Phase auf dem Benutzernamen der E-Mail-Adresse und einem Passwort. In einer zweiten Phase soll der LOGIN-Prozess dann sehr sicher über 2-Faktoren (E-Mail und SMS) via CH-LOGIN erfolgen können, sobald das BIT der Firma Equalizer die Schnittstelle erstellt hat.
 7. Bei den polydisziplinären Gutachten gibt es eine unterschiedliche Anzahl Teilgutachten und eine interdisziplinäre Konsensbeurteilung. Deshalb ist es am einfachsten, wenn die Anzahl Teilgutachten vom Reviewer als Benutzer gesteuert wird: Der Reviewer weiss, wie viele Teilgutachten als Fachgutachten zuerst reviewed werden sollen, bevor die interdisziplinäre Konsensbeurteilung ausgefüllt wird: Am Anfang entscheidet der Reviewer stets, ob er den fachspezifischen Prüfbogen oder die interdisziplinäre Konsensbeurteilung ausfüllen will.
- **Input:** Aus den online Beurteilungen werden die Antworten in eine DB geschrieben und pro Reviewer und pro Gutachten gespeichert.
 - **Output:** Datenbank mit den PRV-Beurteilungen der Reviewer.
 - **Abnahmekriterien:** Zuverlässigkeit, Datensicherheit, Benutzerfreundlichkeit.

Anhang V: Screenshots Online-Erhebungstool PRV

https://www.portal.bsv.admin.ch/#

Prüffrage B1) Ist das Gutachten sprachlich korrekt und unmissverständlich verfasst?

Geltungsbereich: Fachgutachten

Gefordert ist eine knappe, prägnante Ausdrucksweise. Unnötige Längen, z. B. durch reine Wiederholung von bereits beschriebenen Sachverhalten oder ausführliche Diskussionen irrelevanter Befunde, sollen vermieden werden. Zu Wiederholungen zählt nicht, wenn in einem Gliederungspunkt explizit nochmals gleichsinnige Stellungnahmen erfragt werden. Zentral ist die Verständlichkeit der Aussagen.

- Kein Mangel**
- Leichte Mängel**
 - Das Gutachten enthält unnötige inhaltliche Wiederholungen.*
 - 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.*

Prüffragenkatalog Peer-Review Fachgutachten: ID 1

Zurück Weiter