

RELY-Studien Methodik und Ergebnisse

(Deutsche Übersetzung, angelehnt an die Publikation Kunz et al BMC Psychiatry 2019; 19: 205)

RELY 1-Studie

Zielsetzung und Design

Ziel der RELY 1-Studie war zu überprüfen, ob sich mit der funktionsorientierten Begutachtung unter den psychiatrischen Gutachtern eine akzeptable Reproduzierbarkeit bei der Einschätzung der Arbeitsfähigkeit erreichen lässt. Die Methodik zu RELY 1 ist in einem bereits veröffentlichten Protokollpaper im Detail beschrieben [Bachmann et al., 2016]. RELY 1 ist eine multizentrische Reproduzierbarkeitsstudie, in der jeweils vier Psychiater unabhängig voneinander die Arbeitsfähigkeit desselben Antragstellers einschätzten. Konsekutive Antragsteller wurden über die IV-Stelle Zürich und Suva rekrutiert, welche in einem der vier unabhängigen Begutachtungszentren (asim Basel, ZMB Basel, MEDAS Zentralschweiz und MEDAS Interlaken) untersucht wurden.

Rekrutierung

Psychiater: Aufgenommen wurden Psychiater, die zum Zeitpunkt der Studie Begutachtungen im Auftrag der Invalidenversicherung oder der Suva durchführten. Um ein reales Szenario zu gewährleisten, wurden keine weiteren Einschlusskriterien festgelegt. Die Psychiater wurden in den vier teilnehmenden Begutachtungszentren rekrutiert.

Antragsteller: Ärzte und Sachbearbeiter der IV-Stelle Zürich und der Suva kontaktierten Antragsteller, die zum ersten Mal für eine polydisziplinäre Begutachtung (einschliesslich psychiatrischer Begutachtung) vorgesehen waren. Ein weiteres Einschlusskriterium war die Verständigung auf Deutsch ohne Dolmetscher. Nach ersten Informationen zur Studie durch die IV-Mitarbeiter und Einwilligung zur Weitergabe der Kontaktangaben an unser Studienteam erhielten die Antragsteller vom Studienteam ausführliche Informationen zur Studie. Nach schriftlicher Einwilligung zur Teilnahme gemäss Ethikantrag wurden sie definitiv in die Studie aufgenommen.

Durchführung

Training funktionsorientierte Begutachtung

Alle Psychiater absolvierten ein Training in funktionsorientierter Begutachtung (funktionsorientiertes Interview und strukturierte Dokumentation und Einschätzung der Arbeitsfähigkeit mittels dem Instrument zum funktionellen Assessment in der Psychiatrie, IFAP). Während des Trainings (3x3h) führten wir Gruppendiskussionen durch, stellten Fallbeispiele vor und besprachen die Aufgaben, welche die Psychiater zuhause bearbeitet hatten, bzw. bereiteten die Aufgaben vor, welche die Psychiater bis zur nächsten Session zu bearbeiten hatten.

Allokation, Begutachtungsgespräch und Videoring

Antragsteller wurden via MED@P-Plattform nach dem Zufallsprinzip den beteiligten Zentren zugeordnet, dort rekrutiert und dem nächst verfügbaren Psychiater zugeteilt. Dieser führte das funktionsorientierte Begutachtungsgespräch und die anschließende Beurteilung der Leistungs- und Arbeitsfähigkeit mittels IFAP durch. Jedes Gespräch wurde auf Video aufgenommen, drei weitere Studienpsychiater betrachteten unabhängig voneinander die Videos, studierten mitgelieferte Patientenakten und Arbeitsplatzbeschreibung und nahmen mit dem IFAP eine Beurteilung der Leistungs- und der Arbeitsfähigkeit vor.

Datenerhebung und statistische Auswertung

Merkmale der teilnehmenden Psychiater und Antragsteller

Psychiater: Angaben zu Alter, Geschlecht, Berufserfahrung als Facharzt und Gutachter, Anzahl Gutachten im vergangenen Jahr und nebenläufige Tätigkeit in der Patientenversorgung wurden zu Beginn der Studie mittels Fragebogen gesammelt.

Antragsteller: Angaben zu Alter, Geschlecht, Zivilstatus, Nationalität und Herkunftsland wurden im Vorfeld der Begutachtung mittels Fragebogen erfasst. Diagnosen mit Auswirkungen auf die Arbeitsfähigkeit (ICD-10, Kapitel V), Schweregrad der Erkrankung auf einer Skala von 0 (=nicht beeinträchtigt) bis 10 (=schwer beeinträchtigt) und die Ähnlichkeit zum ‚typischen Antragsteller‘ (Repräsentativität der Stichprobe) wurden mittels IFAP durch die Studiengutachter beurteilt.

Gutachterliche Einschätzung zur Arbeitsfähigkeit

Primärer Endpunkt der Studie war die von den Psychiatern eingeschätzte Arbeitsfähigkeit der Patienten bezogen auf eine alternative Tätigkeit (‘eine dem Leiden angepasste Tätigkeit’). An dieser

orientiert sich der Sachbearbeiter der Versicherung bei der Berechnung der IV-Rente. Sekundäre Endpunkte waren a) die von den Psychiatern geschätzte Arbeitsfähigkeit der Antragsteller bezogen auf die letzte Arbeit, b) die subjektive Sicherheit der Psychiater in die eigene Einschätzung der Arbeitsfähigkeit (letzte Arbeit, alternative Tätigkeit) auf einer Skala von 0 (=sehr unsicher) bis 10 (=sehr sicher), c) die von den Patienten wahrgenommene Fairness mit dem Ablauf der Begutachtung, einschliesslich ihrer Gesamtzufriedenheit, auf einer Skala von 0 (=sehr unzufrieden) bis 10 (=sehr zufrieden), und d) das Feedback der Psychiater zur funktionsorientierten Begutachtung.

Wir berechneten die Reliabilität (=Unterscheidbarkeit von Personen mit erhaltenen, eingeschränkten und unzureichenden Ressourcen) anhand des ICC (Intraclass Correlation Coefficient, 1=perfekte Reliabilität und 0=keine Reliabilität). Die Interpretation des ICCs war wie folgt: geringe (ICC < 0.40), mässige (0.40 bis 0.59), gute (0.60 bis 0.74) und exzellente (0.75 bis 1.00) Reliabilität. Aufgrund der Ergebnisse unseres systematischen Reviews zur Reproduzierbarkeit [Barth et al, 2017] erwarteten wir eine Reliabilität von ICC=0.4 in der konventionellen Begutachtung und erachteten für die funktionsorientierte Begutachtung einen ICC von 0.6 (Übergang von 'mässig' nach 'gut') als akzeptabel.

Die Übereinstimmung unter Psychiatern wurde anhand zweier Parameter berechnet: 1) Standardmessfehler als Ausmass für die Streuung der Ergebnisse (SEM, Standard Error of Measurement, 'niedriger ist besser') und 2) Anteil an Vergleichen zwischen zwei Gutachtern, der bei der Beurteilung der Arbeitsfähigkeit das Kriterium 'maximal akzeptabler Unterschied' erfüllt [Schandelmaier et al., 2015].

Für die Interpretation vom SEM und dem Anteil an Vergleichen zwischen zwei Gutachtern, der das Kriterium 'maximal akzeptabler Unterschied' erfüllt, verwendeten wir als 'maximal akzeptablen Unterschied', wenn die Beurteilung der Arbeitsfähigkeit durch zwei Gutachter weniger oder gleich 25 Prozentpunkte auseinander lag. Dieser Wert entsprach den Angaben von Psychiatern in der Versorgung und psychiatrischen Gutachtern in der schweizweiten Umfrage unter mehr als 600 Akteuren [Schandelmaier et al., 2015]. Versicherer, Anwälte und Richter erachteten einen kleineren Unterscheid (20 Prozentpunkte) als maximal akzeptabel.

Ergebnisse in grösserer Detailtiefe,

(angelehnt an die Publikation Kunz et al BMC Psychiatry 2019; 19: 205)

Merkmale der teilnehmenden Psychiater und Antragsteller

Neunzehn Psychiater und 30 Antragsteller nahmen an der RELY 1-Studie teil. Zum Zeitpunkt der Studie waren die Psychiater im Schnitt 15.6 Jahre als Facharzt und 13.8 Jahre als Gutachter tätig. Die Antragsteller litten am häufigsten an affektiven Störungen (46.4%), gefolgt von neurotischen, Belastungs- und somatoformen Störungen (21.4%). Der mittlere Schweregrad der Störungen lag bei 5.3 (0=überhaupt keine Störung, 10=sehr schwere Störung).

Reproduzierbarkeit der gutachterlichen Einschätzung zur Arbeitsfähigkeit

Jeder Antragsteller wurde von 4 Psychiatern begutachtet (30 Patienten x 4 Psychiater = 120 Ratings). Die mittlere Arbeitsfähigkeit der 120 Ratings betrug 43.6% für die letzte Arbeit und 55.0% für eine alternative Tätigkeit. Abbildung 1 zeigt die gutachterlichen Einschätzungen der Arbeitsfähigkeit für alle 30 Antragsteller bezogen auf die letzte Arbeit und eine alternative Tätigkeit.

Reliabilität: Der ICC-Wert bezogen auf die letzte Arbeit betrug 0.38 und zeigte somit geringe Reliabilität, d.h. eine geringe Fähigkeit der Gutachter, Antragsteller mit erhaltener, eingeschränkter und unzureichender Arbeitsfähigkeit zu unterscheiden. Bezogen auf eine alternative Tätigkeit war die Reliabilität mit einem ICC-Wert von 0.43 (gerade noch) mässig.

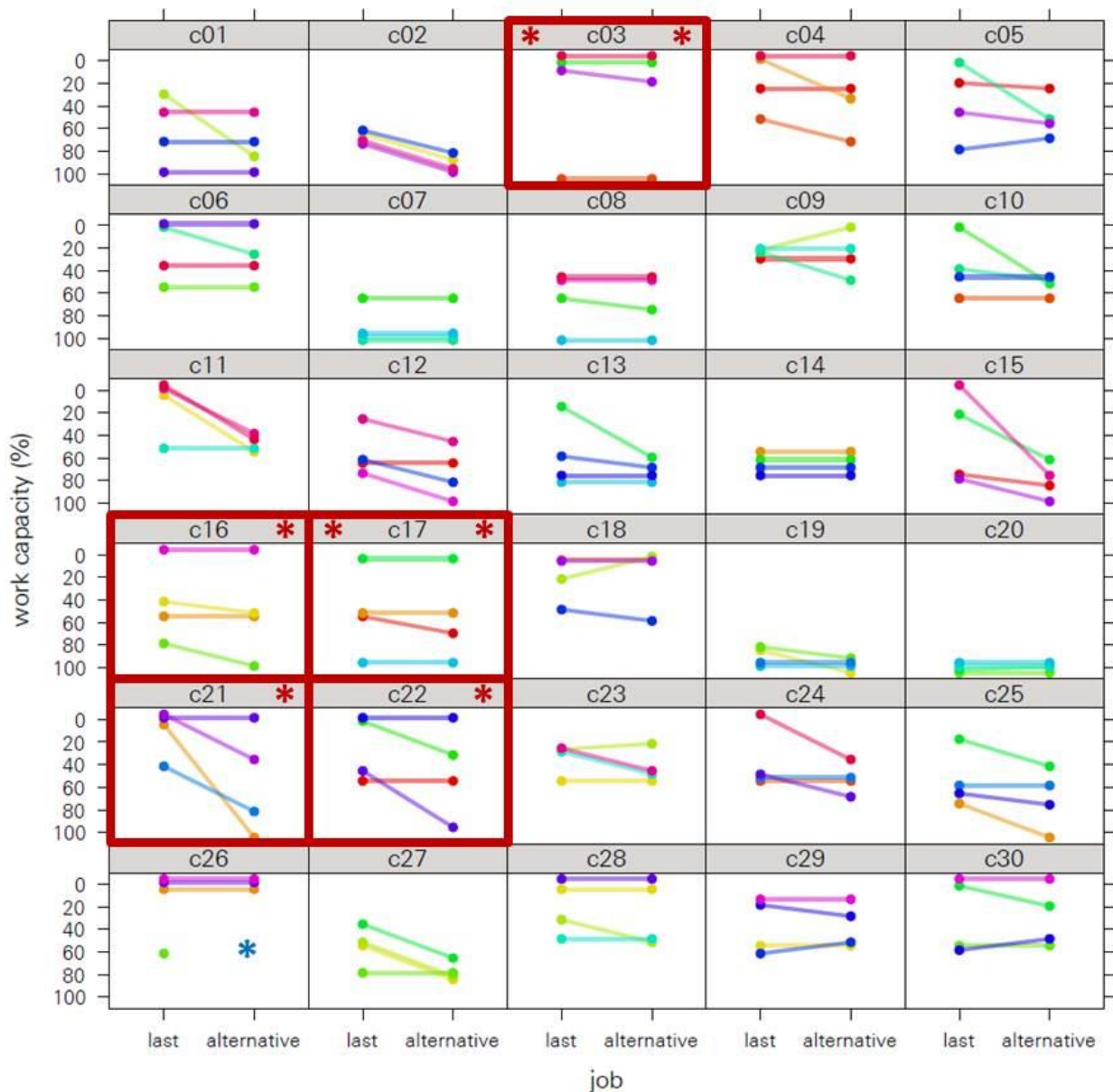
Übereinstimmung: Die Übereinstimmung (gemessen als SEM, Standard Error of Measurement) betrug 26.0 Prozentpunkte Arbeitsfähigkeit für die letzte Arbeit und 24.6 Prozentpunkte Arbeitsfähigkeit für eine alternative Tätigkeit. Die in RELY 1 beobachteten SEM-Werte lagen weit hinter den Erwartungen der Akteure (SEM-Wert <9.0 Prozentpunkte Arbeitsfähigkeit) aus der Umfrage [Schandelmaier et al., 20015].

Auch der Anteil an Vergleichen zwischen zwei Gutachtern, der bei der Beurteilung der Arbeitsfähigkeit das Kriterium 'maximal akzeptabler Unterschied' erfüllte, blieb deutlich hinter den Erwartungen zurück. Den 'maximal akzeptablen Unterschied' von 25 Prozentpunkten Arbeitsfähigkeit (erwartet von Psychiatern und Gutachtern) erfüllten 52.2% aller Vergleiche für die letzte Arbeit und 61.6% aller Vergleiche für eine alternative Tätigkeit.

Auf die Frage nach ihrer subjektiven Sicherheit, dass die eigene Einschätzung der Arbeitsfähigkeit der tatsächlichen Arbeitsfähigkeit des Antragstellers entspricht, gaben die Psychiater

eine mittlere Sicherheit von 7.37 für die letzte Arbeit und 7.17 für eine alternative Tätigkeit an (Skala von 0=total unsicher bis 10=absolut sicher).

Abbildung 1: Grafische Darstellung der Einschätzungen zur Arbeitsfähigkeit in RELY 1.



Erklärung: Jeder der 30 Antragsteller (c01-c30) wurde von vier Gutachtern bezüglich Arbeitsfähigkeit für die letzte Arbeit und für eine alternative Tätigkeit beurteilt. Jedes Kästchen repräsentiert einen Antragsteller (Beispiel: c16=claimant 16). Auf der vertikalen Achse ist die Arbeitsfähigkeit (*work capacity*, %) von 100% bis 0% dargestellt. Die beiden Markierungen auf der horizontalen Achse (*job*) stellen die Bezugsgröße letzte Arbeit (*last*) und alternative Tätigkeit (*alternative*) dar. Die farbigen Punkte kennzeichnen die Schätzungen der vier psychiatrischen Gutachter (jeder Gutachter mit eigener Farbe). Die Verbindungslinie zwischen zwei Punkten verdeutlicht die Änderung der geschätzten Arbeitsfähigkeit von letzter Arbeit zu einer alternativen Tätigkeit. Die roten Umrahmungen markieren Ratings mit einer Extremabweichung (100 Prozentpunkte Arbeitsfähigkeit). Die roten Sternchen (*) markieren, ob sich die Extremabweichungen auf die letzte Arbeit (linker Stern) oder eine alternative Tätigkeit (rechter Stern) beziehen. Das blaue Sternchen (*) markiert ein aufgrund von Regelbruch ausgeschlossenes Rating.

Perspektive Antragsteller und Psychiater

Die Gesamtzufriedenheit der Antragsteller und die subjektiv wahrgenommene Fairness mit dem Ablauf der Begutachtung betrug im Mittel 8.00 (Skala 0=sehr unzufrieden bis 10=sehr zufrieden). Die Psychiater bewerteten die funktionsorientierte Begutachtung positiv. Sie hoben die Strukturiertheit des Vorgehens („man vergisst nichts“), die funktionelle Ausrichtung, und die positive Resonanz der Antragsteller (‘Arbeit zu Beginn des Gesprächs ansprechen’) hervor. Die meisten Psychiater gaben an, die funktionsorientierte Begutachtung nun auch in ihrer eigenen Praxis anzuwenden. Psychiater würden ein intensiveres Training begrüßen und machten konkrete Vorschläge zur Verbesserung von Manual und IFAP-Instrument.

Bewertung der RELY 1-Ergebnisse

Mit dem in der RELY 1-Studie gefundenen Reliabilität für die Einschätzung der Arbeitsfähigkeit in einer alternativen Tätigkeit (ICC=0.43) wurde das gesteckte Ziel nicht erreicht. Die Übereinstimmung unter den Psychiatern lag weit hinter den Erwartungen der Akteure des Schweizer Begutachtungswesens. Administrative Neuerungen der Invalidenversicherung im Jahr 2012 (MEDAP-Plattform; 130-Tage-Regelung zur maximalen Dauer für die Erstellung eines polydisziplinären Gutachtens), die zeitgleich mit dem Start der Rekrutierung der RELY 1-Studie stattfanden und die sich erheblich auf die Prozesse in den Begutachtungsstellen auswirkten, interagierten substantiell mit der Rekrutierung der 30 Antragsteller in den Begutachtungsstellen. Dies führte zu einer erheblichen Verzögerung der Rekrutierung. Durch diese Phase der Inaktivität ging der RELY 1-Studie ein ganzes Jahr der geplanten Durchführungsdauer verloren, mit weitreichenden Konsequenzen für die Umsetzung der Studie. Dies führte zu einer Verzögerung der Datenerhebung (Ratings) und einem mittleren Training-Rating-Intervall von 404 Tagen.

Gestützt auf die nahezu einstimmige Rückmeldung der Studiengutachter in den begleitenden Interviews, dass ihnen die Herangehensweise der funktionsorientierten Begutachtung substantiell helfen würde, die Frage nach der Arbeitsfähigkeit zu beantworten und die hohe Akzeptanz der Antragsteller in Bezug auf Zufriedenheit und empfundener Fairness des Begutachtungsprozesses gingen wir davon aus, dass die funktionsorientierte Begutachtung die Reproduzierbarkeit der gutachterlichen Einschätzungen zur Arbeitsfähigkeit verbessern könnte und dass primär die nicht planmässige Studiendurchführung mit dem mehr als einjährigen Intervall zwischen Training und Rating in der Studie eine höhere Reproduzierbarkeit verhindert haben.

RELY 2-Studie

Zielsetzung und Design

Gemäss SNF-Antrag (<http://p3.snf.ch/Project-144200>) war für RELY 2 ursprünglich eine randomisierte Studie vorgesehen. Aufgrund der geringen Reproduzierbarkeit der gutachterlichen Einschätzungen zur Arbeitsfähigkeit in RELY 1 erübrigte sich der geplante randomisierte Vergleich 'funktionsorientierte versus konventionelle Begutachtung'. Stattdessen wurde die funktionsorientierte Begutachtung weiterentwickelt mit dem Ziel, eine höhere Reliabilität und höhere Übereinstimmung zu erreichen: Instrument und Manual wurden weiter präzisiert, die Trainingsdauer verdoppelt und die Psychiater intensiver an den Regeln des Manuals kalibriert. Des Weiteren wurde sichergestellt, dass das Rating der Videos zeitlich näher zum Training der Psychiater stattfand. Dies bedeutete de facto eine Wiederholung der einarmigen Reliabilitätsstudie RELY 1, bei der die weiterentwickelte funktionsorientierte Begutachtung auf den Prüfstand kam und die Hauptfrage beantworten sollte: „Wie hoch ist die Reproduzierbarkeit der Einschätzungen von Leistungs- und Arbeitsfähigkeit, wenn die funktionsorientierte psychiatrische Begutachtung gemäss Studienplan umgesetzt wird?“

Die Zuordnung der Antragsteller zu den Psychiatern und die Verblindung der Rater gegenüber dem gesprächsführenden Gutachter und untereinander waren vergleichbar zu RELY 1. Die Durchführung der Videoaufnahmen und das Rating der Begutachtungsvideos erfolgten identisch zu RELY 1. Neben 25 neuen Begutachtungsvideos fanden 15 Videos aus RELY 1 Eingang in die RELY 2-Studie, um den Zeitverlust von RELY 1 zu kompensieren.

Rekrutierung

Psychiater: Elf RELY 1-Psychiater, deren Begutachtungsgespräche die Minimalkriterien für ein funktionsorientiertes Interview erfüllt hatten [von Allmen et al., in preparation], wurden für die Durchführung der gefilmten Begutachtungsgespräche von RELY 2 rekrutiert. 24 neue Psychiater meldeten sich auf die Einladung zur Studienteilnahme über die Gutachtermittgliederliste der Swiss Insurance Medicine.

Antragsteller: Die Rekrutierung von 25 neuen Antragstellern der IV-Stelle Zürich erfolgte nach denselben Kriterien wie in RELY 1.

Datenerhebung und statistische Auswertung

In RELY 2 wurden dieselben Endpunkte in derselben Art und Weise wie in RELY 1 erhoben. Primärer Endpunkt der Studie war die von den Psychiatern eingeschätzte Arbeitsfähigkeit der Patienten bezogen auf eine alternative Tätigkeit. Die statistischen Analysen erfolgten identisch zu RELY 1. Die Reliabilität wurde anhand der ICC-Statistik berechnet, die Übereinstimmung anhand von SEM und dem Anteil an Vergleichen zwischen zwei Gutachtern, der bei der Beurteilung der Arbeitsfähigkeit das Kriterium für einen 'maximal akzeptablen Unterschied' (von 10, 15, bis 50 Prozentpunkten) erfüllte. Analog zum Telefoninterview in RELY 1 erfassten wir am Ende von RELY 2 die Einschätzung der teilnehmenden Psychiater zur funktionsorientierten Begutachtung mittels Onlinefragebogen.

Ergebnisse

Merkmale der teilnehmenden Psychiater und Antragsteller

Fünfunddreissig Psychiater und 40 Antragsteller (davon 15 Videos mit Begutachtungsgesprächen aus RELY 1) nahmen an der Studie teil. Die meisten Psychiater waren männlich (82.9%) und gehörten der Alterskategorie 41-50 Jahre an (40.0%). Sie waren im Mittel 14.8 Jahre als Facharzt und 12.4 Jahre als Gutachter tätig. Die Antragsteller (52.5% männlich) hatten ein durchschnittliches Alter von 48.6 Jahren. Am häufigsten litten die Antragsteller an affektiven Störungen (55.9%), gefolgt von neurotischen, Belastungs- und somatoformen Störungen (17.6%). Der mittlere Schweregrad der Störungen lag bei 5.0.

Reproduzierbarkeit der gutachterlichen Einschätzung zur Arbeitsfähigkeit

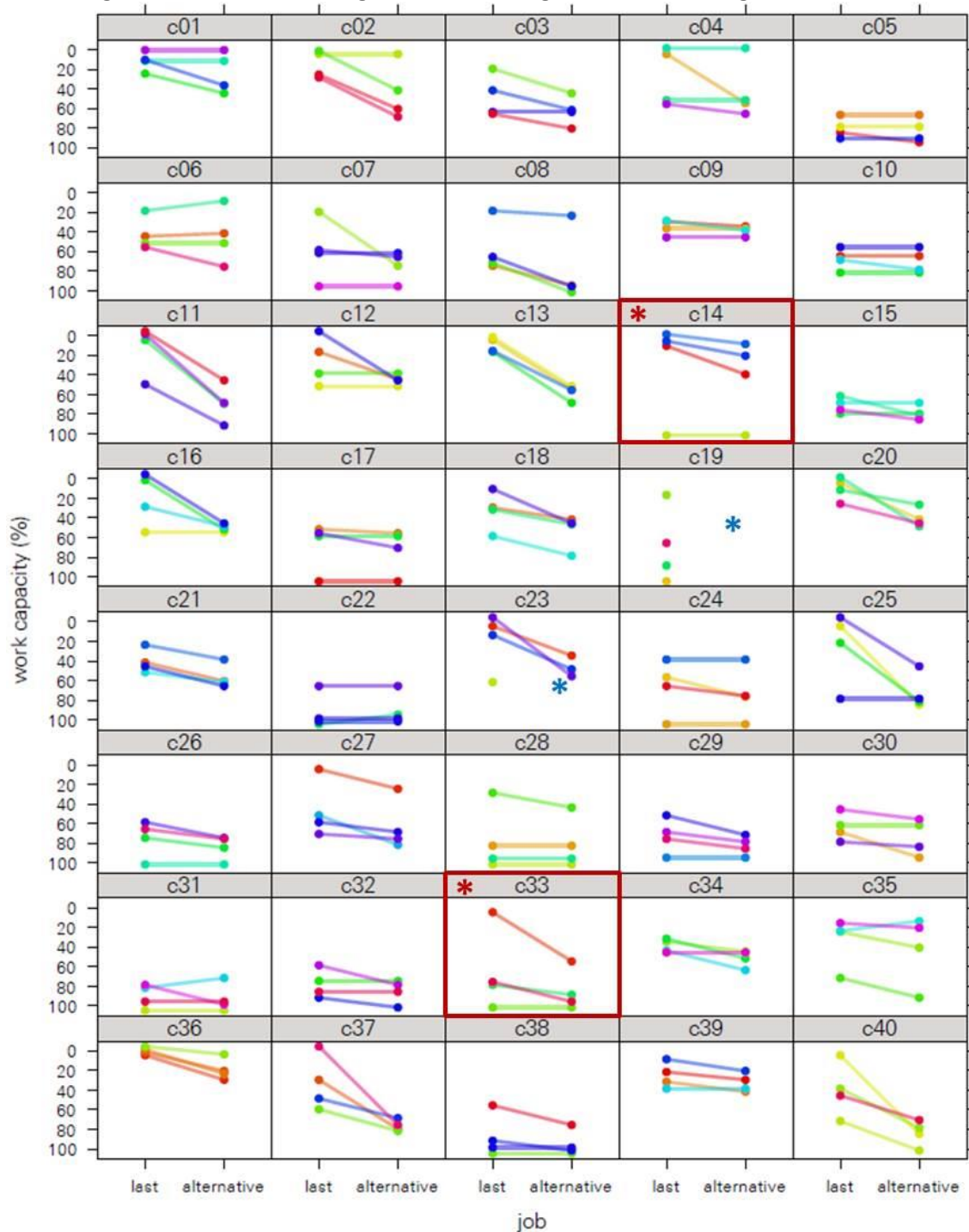
Die mittlere Arbeitsfähigkeit der insgesamt 160 Ratings betrug 46.3% für die letzte Arbeit und 62.9% für eine alternative Tätigkeit. Abbildung 2 zeigt die gutachterlichen Einschätzungen der Arbeitsfähigkeit aller 40 Antragsteller für die letzte Arbeit und einer alternativen Tätigkeit.

Reliabilität: Die ICC-Werte in RELY 2 waren mässig und betrugen 0.47 bezogen auf die letzte Arbeit und 0.44 bezogen auf eine alternative Tätigkeit.

Übereinstimmung: Die SEM-Werte betrugen 23.9 Prozentpunkte Arbeitsfähigkeit für die letzte Arbeit und 19.4 Prozentpunkte Arbeitsfähigkeit für eine alternative Tätigkeit und überstiegen deutlich den 'maximal akzeptablen Unterschied' aus der Umfrage unter den Akteuren [Schandelmaier et al., 2015]. Einen 'maximal akzeptablen Unterschied' von 25 Prozentpunkten Arbeitsfähigkeit (Psychiater und Gutachter) erfüllten 61.6% aller Vergleiche für die letzte Arbeit und 73.6% aller Vergleiche für eine alternative Tätigkeit.

Auf die Frage nach ihrer subjektiven Sicherheit in die eigene Einschätzung der Arbeitsfähigkeit gaben die Psychiater eine mittlere Sicherheit von 7.68 für die letzte Arbeit und 7.44 für eine alternative Tätigkeit an.

Abbildung 2: Grafische Darstellung der Einschätzungen zur Arbeitsfähigkeit in RELY 2.



Erklärung: Alle 40 Antragsteller (c01 bis c40) wurden von vier Gutachtern bezüglich Arbeitsfähigkeit für die letzte Arbeit und eine alternative Tätigkeit beurteilt. Jedes Kästchen repräsentiert einen Antragsteller (Beispiel: c16 = claimant 16). Die vertikale Achse zeigt die Arbeitsfähigkeit (*work capacity*, %) von 100% bis 0%. Die Markierungen auf der horizontalen Achse (*job*) stellen die Bezugsgrösse ‚letzte Arbeit‘ (*last*) und ‚alternative Tätigkeit‘ (*alternative*) dar. Die farbigen Punkte zeigen die Schätzungen der vier Gutachter (jeder Gutachter mit eigener Farbe). Die Linie zwischen zwei Punkten verdeutlicht

die Änderung der geschätzten Arbeitsfähigkeit von letzter Arbeit zu einer alternativen Tätigkeit. Die roten Umrahmungen markieren Ratings mit einer Extremabweichung. Die roten Sternchen (*) markieren, ob sich die Abweichungen auf die letzte Arbeit (linker Stern) oder eine alternative Tätigkeit (rechter Stern) beziehen. Die blauen Sternchen (*) markieren Ratings, die aufgrund von Regelbrüchen ausgeschlossen wurden.

Perspektive Antragsteller und Psychiater

Die Gesamtzufriedenheit der Antragsteller und die subjektiv wahrgenommene Fairness mit dem Ablauf der Begutachtung betrug im Mittel 9.42 (Skala 0=sehr unzufrieden bis 10=sehr zufrieden). Nach Abschluss von RELY 2 nahmen 24 der 25 für RELY 2 neu rekrutierten Studienpsychiater an der Onlinebefragung zum Training der funktionsorientierten Begutachtung teil. Die Psychiater bewerteten die funktionsorientierte Begutachtung als sehr positiv, fanden die Herangehensweise der funktionsorientierten Begutachtung substantiell als hilfreich, um die Arbeitsfähigkeit der Antragsteller einzuschätzen. Die Mehrheit plädierte für den Aufbau solcher Angebote gleichermaßen an Gutachter mit weniger wie mit langjähriger Erfahrung. Ein Drittel empfahl ein verpflichtendes Trainingsangebot für die Psychiater.

Bewertung der RELY 2-Ergebnisse

Mit dem intensiveren Training in funktionsorientierter Begutachtung und dem zeitnahen Einsatz in der Studie hatten wir in RELY 2 bei der Einschätzung der Arbeitsfähigkeit eine Reliabilität von etwa 0.6 erwartet, also eine mässig bis gute Fähigkeit der Gutachter, Antragsteller mit erhaltener, eingeschränkter und unzureichender Arbeitsfähigkeit zu unterscheiden. Erzielt wurden ICC-Werte von 0.44, bezogen auf eine alternative Tätigkeit, d.h. eine (gerade noch) mässige Fähigkeit der Gutachter, Antragsteller mit unterschiedlicher Arbeitsfähigkeit zuverlässig zu unterscheiden. Auch die Übereinstimmung unter den Psychiatern blieb deutlich hinter den Erwartungen der Akteure des schweizerischen Begutachtungswesens. Auch in RELY 2 gaben die Studiengutachter positive Rückmeldung für die Rolle der funktionsorientierten Begutachtung bei der Einschätzung der Arbeitsfähigkeit. Die Mehrheit plädierte für den Auf- und Ausbau solcher Angebote, ggf. sogar als verpflichtendes Bildungsangebot für Psychiater. Die Antragsteller empfanden den Ablauf der funktionsorientierten Begutachtung in hohem Masse als ein faires Verfahren und äusserten sehr grosse Zufriedenheit mit den Begutachtungsgesprächen.

Vergleich RELY 1 mit RELY 2

Zielsetzung und Design

Weil sich der beim SNF beantragte Vergleich als randomisiert kontrollierte Studie zwischen konventioneller und funktionsorientierter Begutachtung aufgrund der unerwartet niedrigen Reproduzierbarkeit in RELY 1 nicht realisieren liess, führten wir einen post-hoc Vergleich zwischen RELY 1 (weniger intensives Training, verzögerter Einsatz) und RELY 2 (intensiveres Training, zeitnaher Einsatz) durch. Auch wenn die Zuordnung zu den Trainingsformaten und den Umsetzungsrealitäten nicht mehr randomisiert und die Durchführung sequentiell und nicht zeitlich parallel erfolgte, waren die sonstigen Rahmenbedingungen vergleichbar: Patientenpool, Gutachterpool, Methodik, Outcomes, Analyse der Daten. Wie der geplante randomisierte Vergleich postulierte auch der nicht-randomisierte Vergleich, dass ein intensiveres Training der Psychiater mit zeitnahe Einsatz in RELY 2 die Reproduzierbarkeit der Arbeitsfähigkeit verbessern würde, verglichen mit dem weniger intensiven Training und der zeitlich verzögerten Umsetzung in RELY 1.

Statistische Auswertung

In RELY 1 und RELY 2 berechneten wir die Reliabilität (ICC) und die Übereinstimmung unter den Psychiatern (SEM und Anteil Vergleiche zwischen zwei Gutachtern, deren Einschätzungen dem Kriterium 'maximal akzeptabler Unterschied' genügten, wenn diese die Arbeitsfähigkeit desselben Patienten einschätzten). Für jede der drei Parameter wurde der Vergleich RELY 1 versus RELY 2 durchgeführt und zwar bezogen auf den primären Endpunkt, die alternative Tätigkeit.

Das Analyseverfahren hat gewisse Einschränkungen. Es geht davon aus, dass keine Abhängigkeiten zwischen den vergleichenden Studien bestehen, was bei RELY 1 und 2 nicht der Fall ist: So gibt es einige wenige Psychiater und Antragsteller, welche in RELY 1 wie auch in RELY 2 partizipierten. Da es zu den vorgenommenen Vergleichen allerdings kein etabliertes methodisches Verfahren gibt, wurde das hier verwendete Analyseverfahren trotz seiner Limitationen eingesetzt.

Ergebnisse

Der Vergleich der ICC-Werte (Reliabilität) für die Arbeitsfähigkeit in einer alternativen Tätigkeit zeigte keinen substantiellen Unterschied zwischen den beiden Studien (RELY 1 vs. 2, ICC: 0.43 vs. 0.44, Tabelle 1). Der SEM-Wert (Übereinstimmung) für die Arbeitsfähigkeit in einer alternativen Tätigkeit wurde in RELY 1 auf 24.6 Prozentpunkte Arbeitsfähigkeit und in RELY 2 auf 19.4 Prozentpunkte Arbeitsfähigkeit geschätzt. Diese Veränderung im SEM um -5.2 Prozentpunkte Arbeitsfähigkeit (d.h., Verbesserung der Übereinstimmung) war signifikant (95% Konfidenzintervall -9.7 bis -0.6). Die

Zunahme der Vergleiche, die das Kriterium 'maximal akzeptabler Unterschied' von 25 Prozentpunkten Arbeitsfähigkeit (RELY 1 vs. 2: 63.3% vs. 73.6%) war ebenfalls signifikant ($p=0.008$).

Tabelle 1: Reliabilität und Übereinstimmung in RELY 1 und RELY 2.

Geschätzte Arbeitsfähigkeit, Reliabilität und Übereinstimmung für die letzte Arbeit und eine alternative Tätigkeit.

Bezugsgrösse Arbeitsfähigkeit		Arbeitsfähigkeit Mittelwert	Reliabilität ICC	Übereinstimmung		
				Anteil Vergleiche zwischen zwei Psychiatern, deren Ratings für Arbeitsfähigkeit sich ≥ 25 Prozentpunkte unterscheiden	SEM	Maximal akzeptierter Unterschied
					Prozentpunkte Arbeitsfähigkeit	Prozentpunkte Arbeitsfähigkeit
Letzte Arbeit	RELY 1 N=120	43.6%	0.38	52.2%	26.0%	72.2%
	RELY 2 N=160	46.3%	0.47	61.7%	23.9%	66.1%
Alternative Tätigkeit	RELY 1 N=119	55.0%	0.43	63.3%	24.6%	68.1%
	RELY 2 N=155	62.9%	0.44	73.6%	19.4%	53.8%

Bewertung des Vergleichs RELY 1 mit RELY 2

Übereinstimmung: Der nicht-randomisierte Vergleich der beiden multi-zentrischen Reproduzierbarkeitsstudien – RELY 1 und 2 – über die gutachterliche Beurteilung der Arbeitsfähigkeit von Patienten mit psychischen Störungen zeigte, dass intensiveres Training in funktionsorientierter Begutachtung die Übereinstimmungen der Arbeitsfähigkeitsbeurteilungen zwischen Psychiatern statistisch signifikant und aus gutachterlicher Perspektive in einem relevanten Umfang verbesserte.

Angesichts der sehr grossen Unterschiede in den Ausgangswerten blieben die Ergebnisse trotz Verbesserung deutlich hinter den Erwartungen einer breiten Gruppe von Akteuren im Begutachtungswesen zurückblieb.

Reliabilität: Trotz Verringerung der Streuung ('Varianz') um 36% in RELY 2, hatte das Training keine Auswirkungen auf die eingeschränkte Fähigkeit der Gutachter, Patienten mit geringer Beeinträchtigung in der Arbeitsfähigkeit (bezogen auf die alternative Tätigkeit) von solchen mit ausgeprägter Beeinträchtigung zu unterscheiden. Dies erklärt sich dadurch, dass es durch die funktionsorientierte Begutachtung zu einer genaueren Einschätzung der Leistungsfähigkeit kam. Beim genauen Hinschauen entdeckten die Gutachter auch bei jenen gewisse Einschränkungen, denen sie früher vollständige Leistungsfähigkeit attestiert hätten, und sie entdeckten auch bei jenen eine gewisse Leistungsfähigkeit, denen sie früher vollständige Einschränkung attestiert hätten. Jeder hat gewisse Einschränkungen, aber auch eine gewisse Leistungsfähigkeit. Bezogen auf die Leistungsfähigkeit rücken die Antragsteller mehr in die Mitte und werden ähnlicher. Nur wenige sind völlig leistungsfähig bzw. leistungsunfähig. Damit bleibt die Unterscheidbarkeit schwierig: Diese Beobachtung reflektiert das wirkliche Leben. 'Es ist schwieriger, Personen, die relativ ähnlich sind, zu unterscheiden, verglichen mit Personen, die sehr unterschiedlich sind.' [Streiner et al. Health Measurement Scales. 2014]

Darüber hinaus nahmen die beiden wichtigsten Akteure – Antragsteller und Gutachter - die funktionsorientierte Begutachtung als fair und als eine nützliche Ergänzung zur konventionellen Begutachtung wahr, um die berufsbezogene Funktionalität des Antragsstellers, seine Einschränkungen und Möglichkeiten zu erheben und abzubilden. Die Versicherer betonten die grössere Klarheit und Strukturiertheit der Aussagen über die Arbeitsfähigkeit, die auch zu einer grösseren Transparenz der Gutachten beitragen sollte.

Zusammenfassung:

1. Ausgangspunkt: Psychiater erzielen nur niedrige Übereinstimmung, wenn sie die Arbeitsfähigkeit von Pat. mit psychischen Störungen beurteilen.
2. Die funktionsorientierte Begutachtung, ein neu entwickeltes Vorgehen mit Fokus auf Arbeit, findet grosse Anerkennung bei Psychiatern (hilfreich, da strukturierter und systematischer), bei Patienten (subjektiv als fair empfunden) und Versicherern (transparent und funktionsorientiert).
3. Intensiviertes Training in funktionsorientierter Begutachtung verbessert in signifikantem und gutachterlich relevantem Ausmass die Übereinstimmung, bleibt aber trotzdem deutlich hinter den Erwartungen von Akteuren in der Schweiz zurück [Swiss Med.Weekly 2015]
4. Verbesserte Transparenz (= bessere Nachvollziehbarkeit der gutachterlichen Überlegungen) bedeutet nicht verbesserte Übereinstimmung unter den Gutachtern in der Beurteilung der Arbeitsfähigkeit bei demselben Patienten
5. Diese Situation ist nach wie vor unbefriedigend. Braucht es eine gesellschaftliche Diskussion über das Ausmass an gewünschter Übereinstimmung unter Gutachtern? Wie lässt sich die gewünschte Übereinstimmung erreichen? Was muss dazu gemacht werden? Zu welchen Kosten?