

UNI  
BASEL



**Departement Klinische Forschung**

EbIM Forschung & Bildung

# Die RELY – Studien: Ergebnisse

**RELY 1 und 2 zu Reliabilität/ Übereinstimmung  
der Begutachtung**

**RELY 3: Transparenz und Nachvollziehbarkeit**



FONDS NATIONAL SUISSE  
SCHWEIZERISCHER NATIONALFONDS  
FONDO NAZIONALE SVIZZERO  
SWISS NATIONAL SCIENCE FOUNDATION



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Eidgenössisches Departement des Innern EDI  
Bundesamt für Sozialversicherungen BSV

**suva**

---

# Die RELY Studien

- WP5:** Überprüfung der Reliabilität der funktionsorientierte Begutachtung (WP3) bei Einschätzung der AUF
- WP6:** Randomisierter Vergleich der Reliabilität der Einschätzung der AUF: funktionsorientierte versus konventionelle Begutachtung.
- WP7:** *Informationsqualität* von funktionsorientierten Gutachten: Sicht der Nutzer

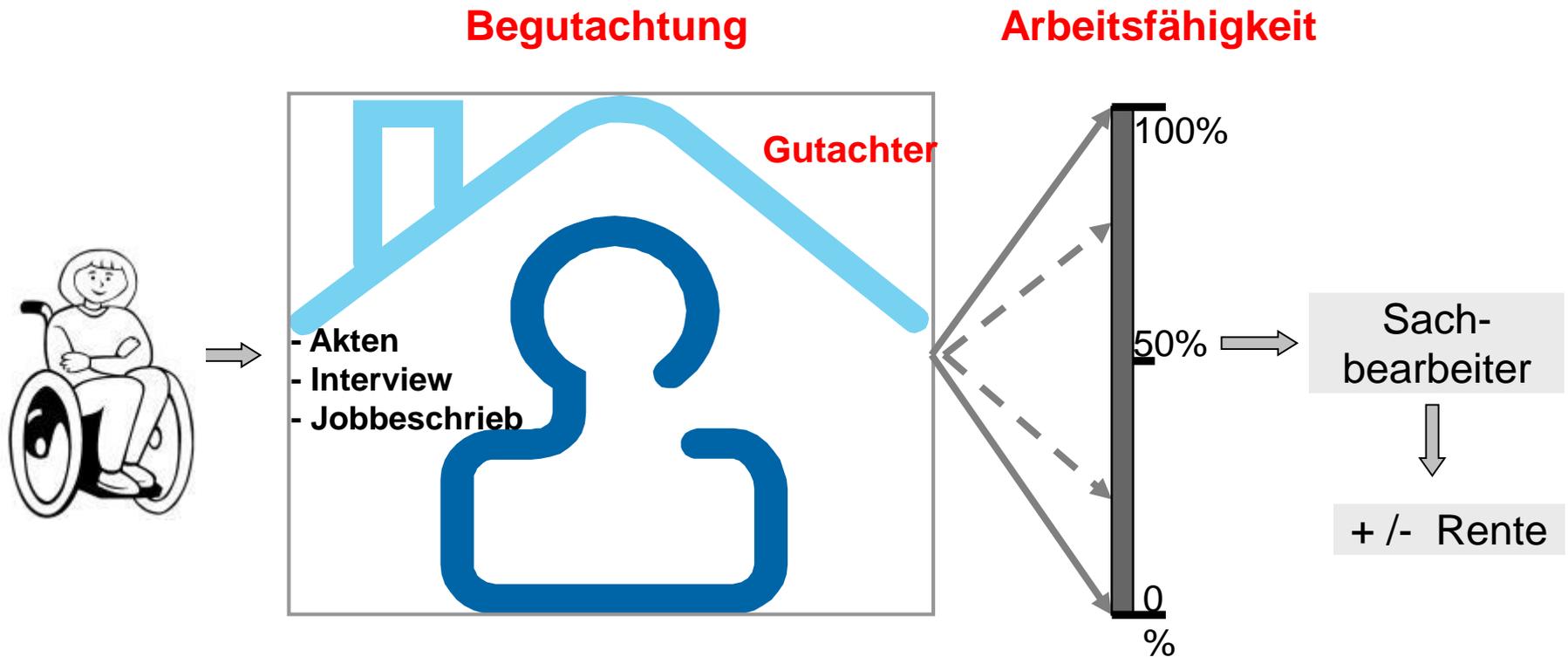
---

Abgeleitet aus allgemeiner Kritik und dokumentiert im Stakeholder Survey 2015 und im systematischen Review im BMJ 2017

## Unsere Ziele

- 1) Erhöhte Reproduzierbarkeit der Begutachtungsergebnisse
- 2) Bessere Transparenz und Nachvollziehbarkeit

# Prozess der Begutachtung als „Instrument“ (*Black Box*) zur Erfassung der Leistungsfähigkeit.



# Reproduzierbarkeit

## Interrater Reliabilität

(Unterscheidbarkeit)

Wie gut können 2 oder mehr GA Personen mit *intakter, noch erhaltener, eingeschränkter, fehlender* Arbeitsfähigkeit zuverlässig unterscheiden?

## Interrater Agreement

(Übereinstimmung)

Wie gut sind 2 oder mehr GA in der Lage, unter ähnlichen Begutachtungsbedingungen die gleiche Arbeitsfähigkeit einzuschätzen?

---

# Funktionsorientierte Begutachtung Training



**Funktionsorientiertes  
Interview**



**IFAP**  
Instrument Funktionelles  
Assessment in der Psychiatrie

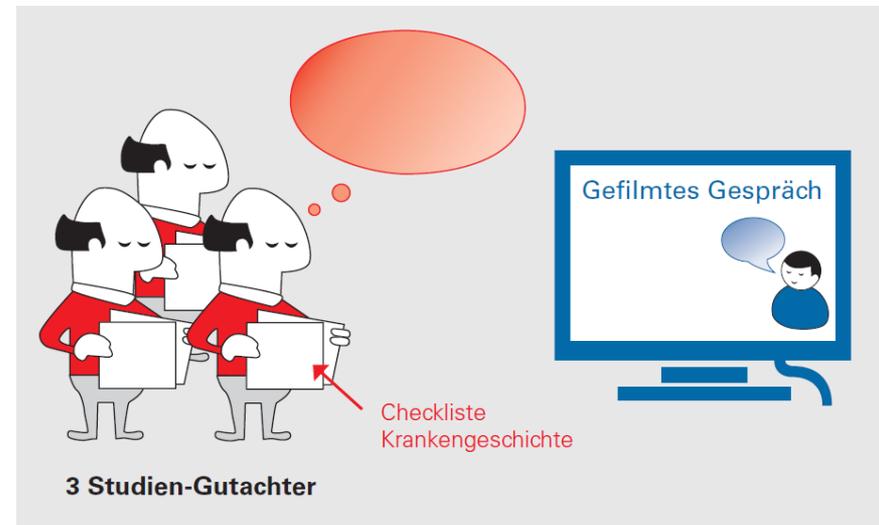
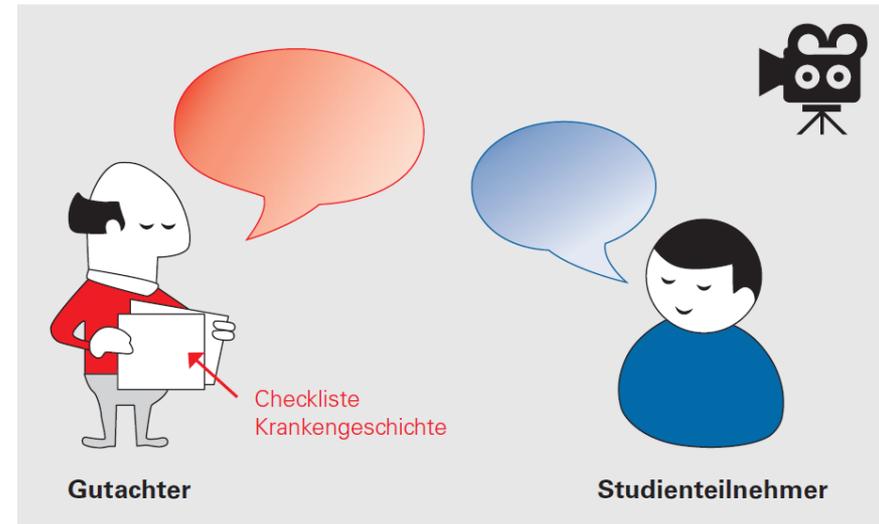
# Ablauf

## RELY 1

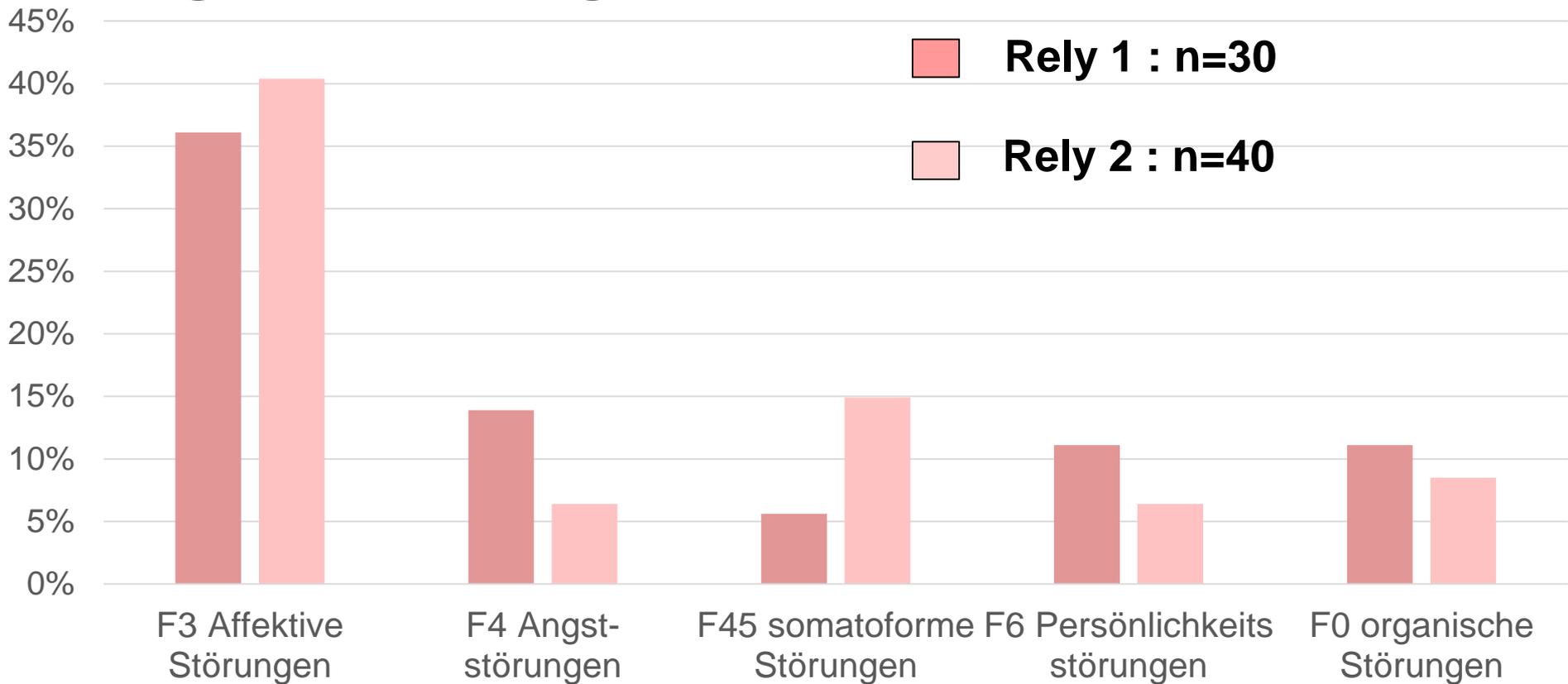
**Training**  
19 Psychiater

30 Antragsteller

Reliabilität / Übereinstimmung  
**%-Wert Arbeitsfähigkeit**  
in letzter Arbeit  
in alternativer Tätigkeit



# Diagnosen Antragsteller



**Schweregrad der mentalen  
Störung (Skala 0 - 10)**

**RELY 1**

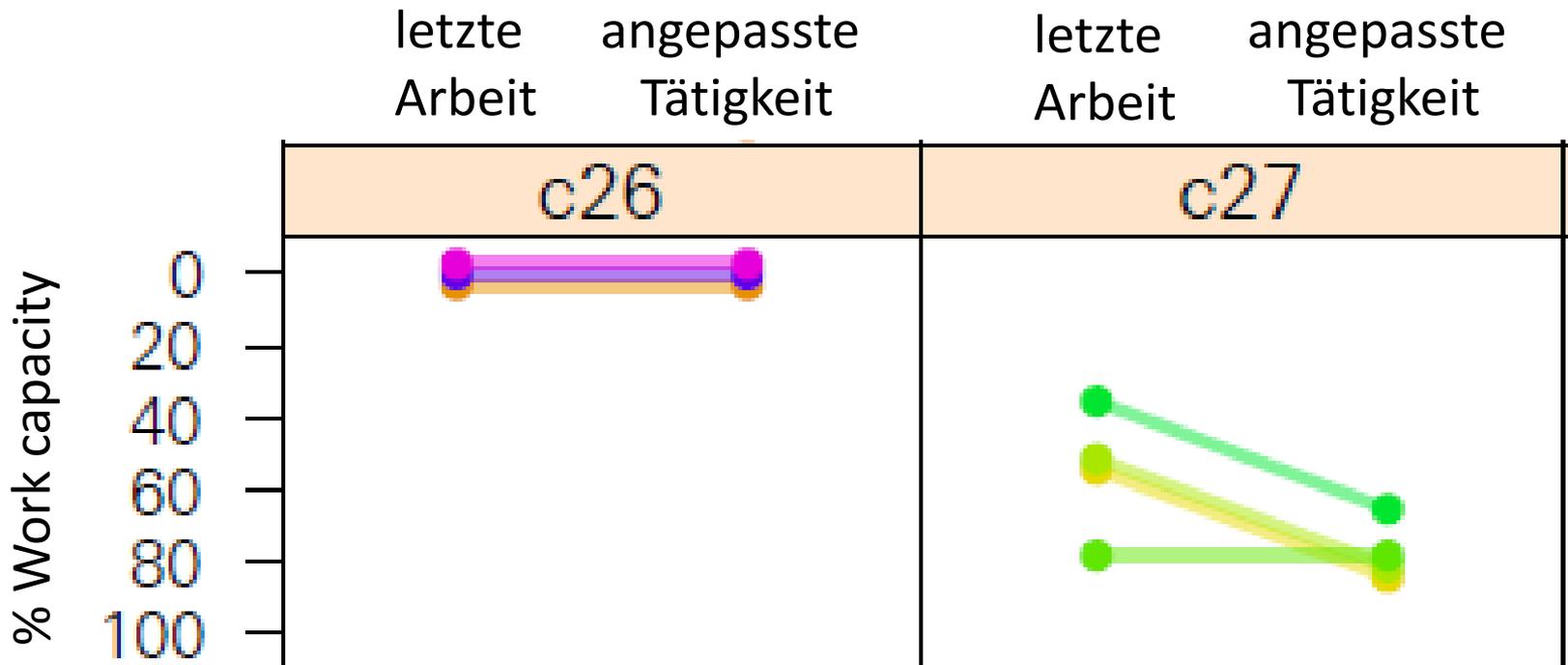
5.34

**RELY 2**

4.95

# Wie lese ich die Ergebnisse ?

## Beurteilung der %-Arbeitsfähigkeit



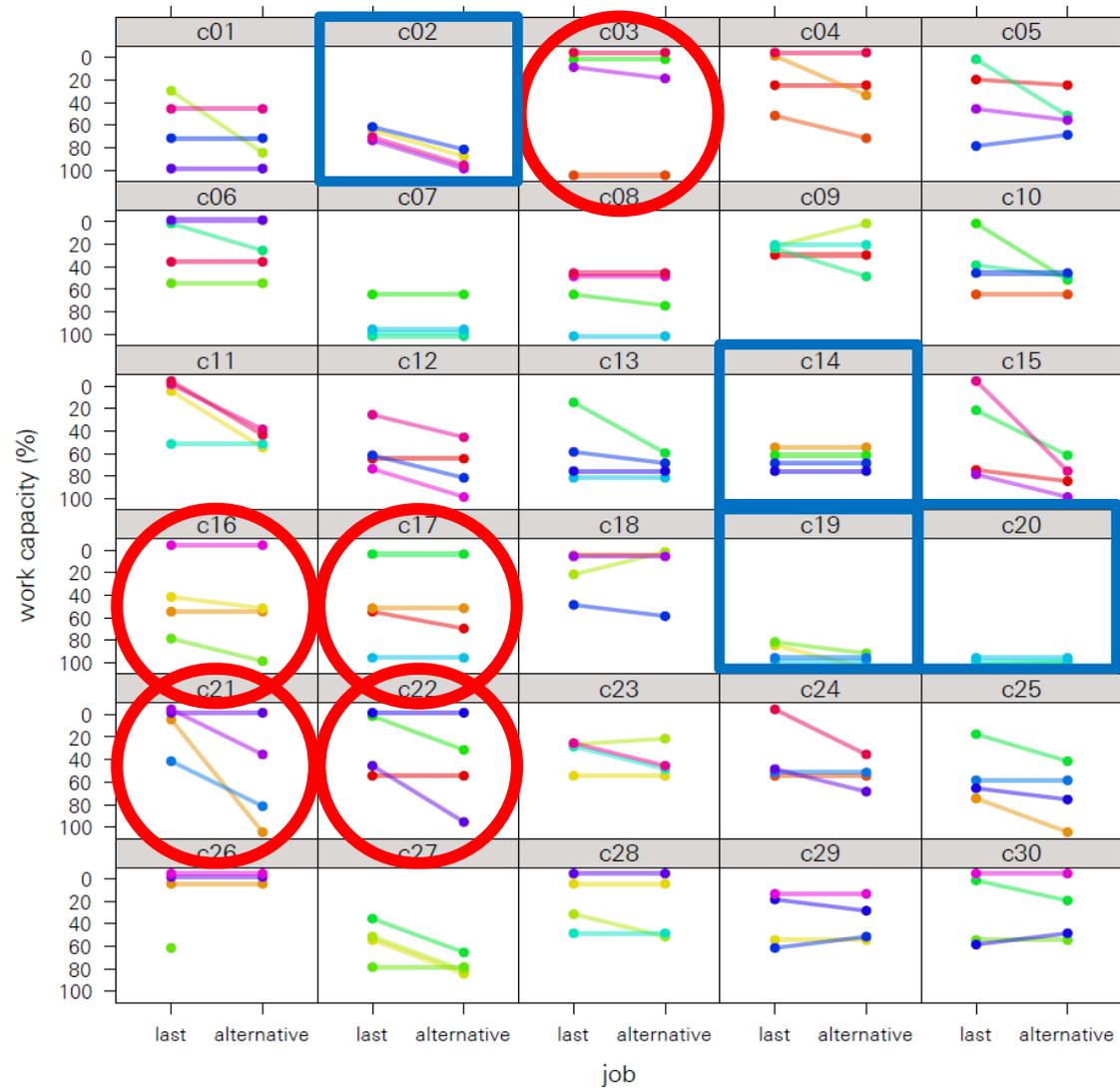
# RELY 1

## Beurteilungen der Gutachter

N = 30 Antragsteller

**Differenz 100%-Punkte:**

7/60 (12%) Beurteilungen



---

# Unsere Erklärung für RELY 1 - Ergebnisse

- grosse Zeitspanne Training-Rating
- 3x3h Training zu wenig intensiv

---

## Training RELY 2



**IFAP**

**Funktionsorientiertes  
Interview**

**Instrument Funktionelles  
Assessment in der Psychiatrie**

### **Intensiveres Training**

- Verdopplung Trainingszeit
- Weiterentwicklung Manual
- Vermehrte Kalibrierung an Regeln

### **Rating im Anschluss an das Training**

# Ablauf

## RELY 1

### Training

19 Psychiater

30 Antragsteller

Reliabilität/ Übereinstimmung  
%-Arbeitsfähigkeit

## RELY 2

### Training

35 Psychiater

40 Antragsteller

Reliabilität / Übereinstimmung  
%-Arbeitsfähigkeit

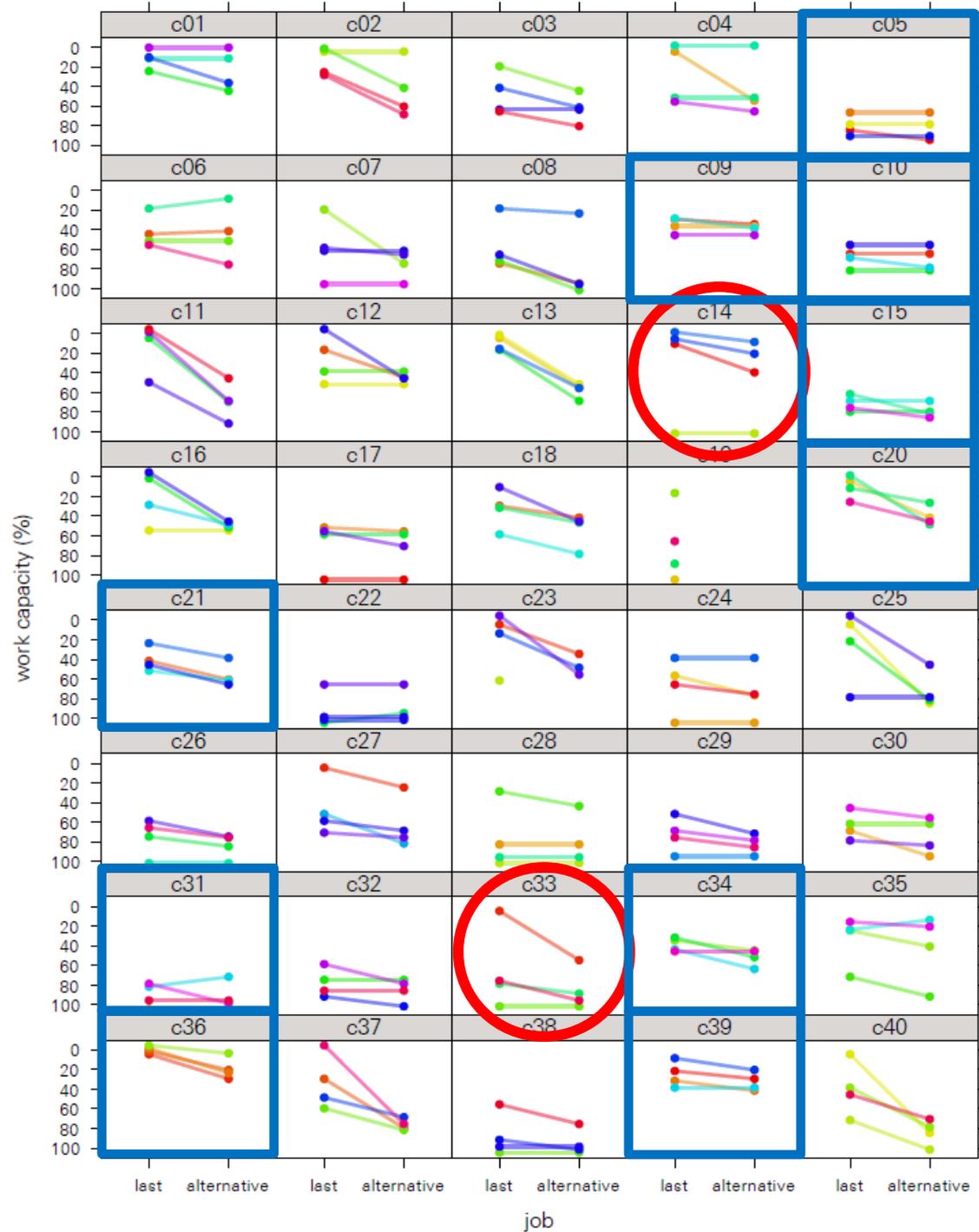
# RELY 2

## Beurteilungen der Gutachter

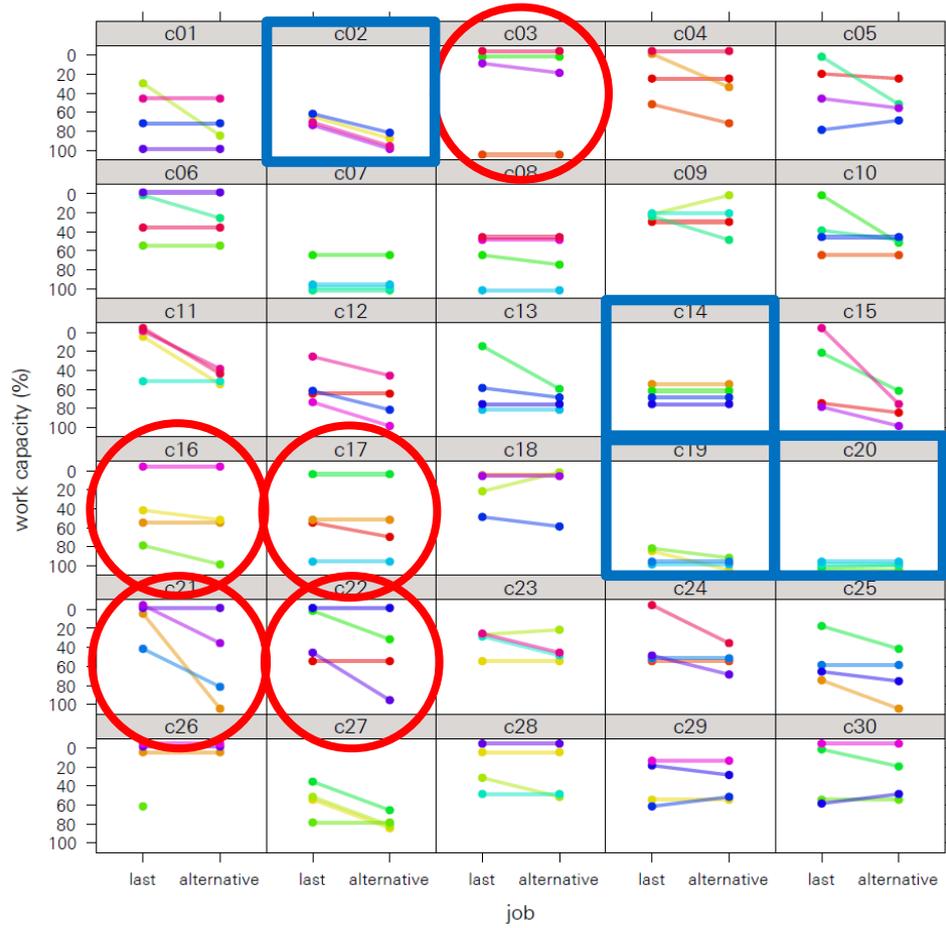
N = 40 Antragsteller

### Differenz 100%-Punkte:

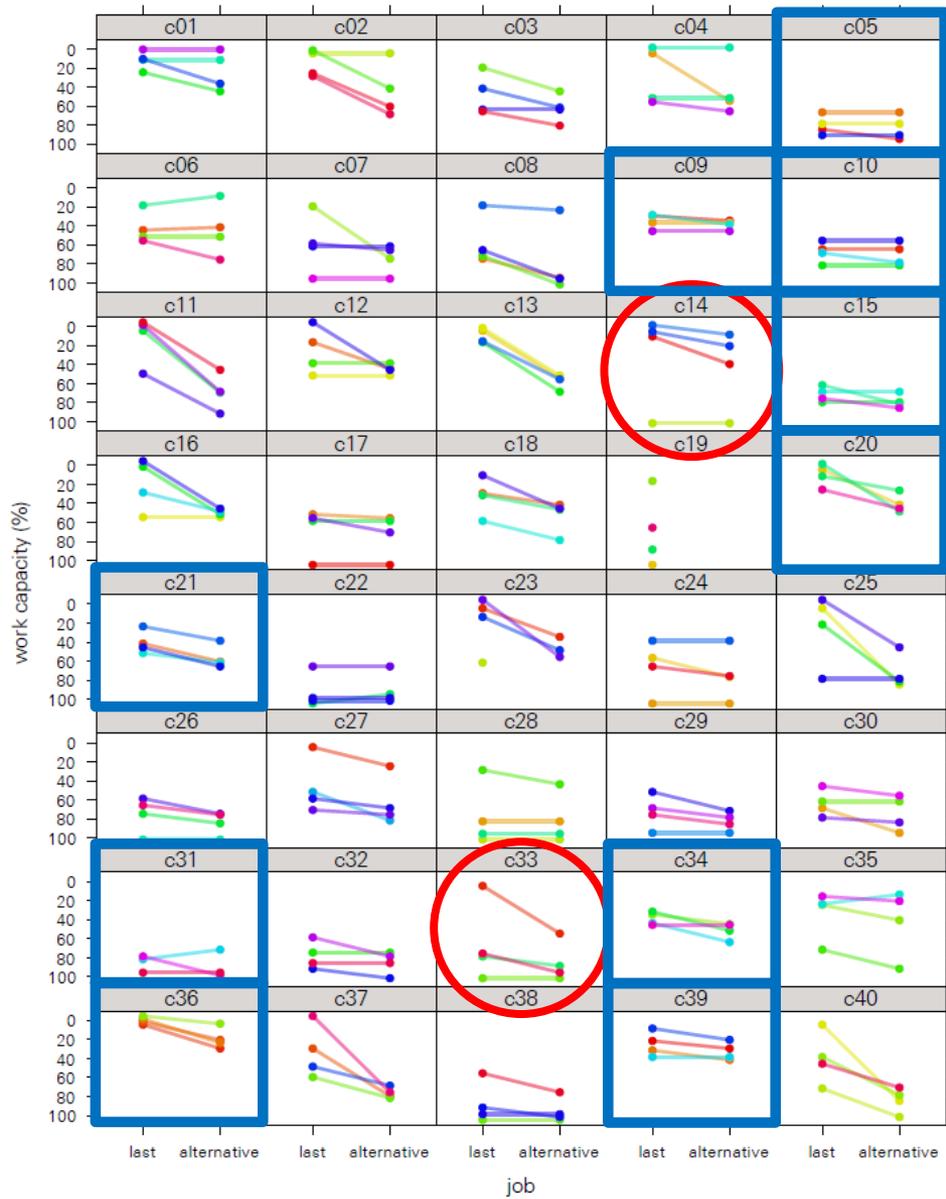
- letzte Arbeit N = 2
- alternative Arbeit N = 0



# RELY 1



# RELY 2



# Reproduzierbarkeit

3 Kennwerte

## Interrater Reliabilität

(Unterscheidbarkeit)

Wie gut können 2 oder mehr GA Personen mit *intakter, noch erhaltener, eingeschränkter, fehlender* Arbeitsfähigkeit zuverlässig unterscheiden?

**1) ICC**

**Intraclass correlation coefficient**

## Interrater Agreement

(Übereinstimmung)

Wie gut sind 2 oder mehr GA in der Lage, unter ähnlichen Begutachtungsbedingungen die gleiche Arbeitsfähigkeit einzuschätzen?

**2) Anteil Vergleiche zw. 2 GA, die das Maximal-Kriterium erfüllen**

**3) SEM (Standardmessfehler, Streuung)**

# Reproduzierbarkeit

## 1) Interrater Reliabilität

(Unterscheidbarkeit)

Wie gut können 2 oder mehr GA Personen mit *intakter, noch erhaltener, eingeschränkter, fehlender* Arbeitsfähigkeit zuverlässig unterscheiden?

**ICC**

**Intraclass correlation coefficient**

ICC-Wert	Interpretation
0.75 – 1	sehr gut
0.6 – 0.75	gut
0.4 – 0.59	mässig
0 – 0.39	unzureichend

# Ergebnisse: Reliabilität und Übereinstimmung

	Durchschnittliche Arbeitsfähigkeit Alternative Tätigkeit
<b>RELY 1</b> 120 Beurteilungen	55%
<b>RELY 2</b> 160 Beurteilungen	63%

# Ergebnisse: Reliabilität und Übereinstimmung

## 1. Varianzkomponenten und Reliabilität

für die alternative Tätigkeit

		Gesamt- Varianz Streuung um %-Wert AF	Anteile an Varianz			Reliabilität
			Psychia- ter	Antrag- steller	Rest	ICC
Letzte Arbeit	RELY 1	1092	263 (24%)	414 (38%)	415 (38%)	<b>0.38</b>
	RELY 2	1064	76 (7%)	495 (47%)	493 (46%)	<b>0.47</b>
Angepasste Tätigkeit	RELY 1	1060	88 (8%)	457 (43%)	515 (49%)	0.43 (0.22-0.60)
	RELY 2	669	50 (7%)	292 (44%)	328 (49%)	0.44 (0.25-0.59)

# Ergebnisse: Reliabilität und Übereinstimmung

## 1. Varianzkomponenten und Reliabilität

für die alternative Tätigkeit

		Gesamt- Varianz Streuung um %-Wert AF	Anteile an Varianz			Reliabilität
			Psychia- ter	Antrag- steller	Rest	ICC
Letzte Arbeit	RELY 1	1092	263 (24%)	414 (38%)	415 (38%)	<b>0.38</b> (0.19-0.55)
	RELY 2	1064	76 (7%)	495 (47%)	493 (46%)	<b>0.47</b> (0.29-0.61)
Alternative Tätigkeit	RELY 1	<b>1060</b>	88 (8%)	457 (43%)	515 (49%)	<b>0.43</b>
	RELY 2	<b>669</b>	50 (7%)	292 (44%)	328 (49%)	<b>0.44</b>

# Reproduzierbarkeit

## 1) Interrater Reliabilität

(Unterscheidbarkeit)

Wie gut können 2 oder mehr GA Personen mit *intakter, noch erhaltener, eingeschränkter, fehlender* Arbeitsfähigkeit zuverlässig unterscheiden?

**ICC**

**Intraclass correlation coefficient**

ICC-Wert	Interpretation
0.75 – 1	sehr gut
0.6 – 0.75	gut
0.4 – 0.59	mässig
0 – 0.39	unzureichend

# Varianz-Komponenten mit Einfluss auf den Prozentwert Arbeitsfähigkeit

Woher kommt die Streuung der Varianzkomponenten?

- **Varianz Psychiater**, z.B.:
  - (Un-)Strukturiertheit des Vorgehens
  - Erfahrung als Psychiater/Gutachter
  - Subjektive «Strenge/Milde»
  - Politische Einstellung
- **Varianz Antragsteller**, z.B.:
  - Sozio-demografische Merkmale
  - Diagnose, Schweregrad
  - Motivation, Selbstwahrnehmung zu arbeiten
- **Rest-Varianz**
  - Situative Faktoren; Interaktion Psychiater\*Antragsteller
  - Rahmenbedingungen im weitesten Sinne (z.B. gesellschafts-politisches Klima, Auswirkung verschiedener BGer-Urteile auf Begutachtung; Personalwechsel in der Studie;)

---

# Schlussfolgerung für die Reliabilität in RELY

→ Antragsteller werden ähnlicher, jeder hat gewisse Einschränkungen, nur wenige sind völlig leistungs(un-) fähig  
*(diese kommen i.a. nicht zur Begutachtung ... )*

→ Unterscheidbarkeit bleibt schwierig

**Es ist schwieriger,**

**Personen, die relativ ähnlich sind, zu unterscheiden,**

**als Personen, die sehr unterschiedlich sind**

(Streiner 2014)

---

# Ergebnisse: Reliabilität und Übereinstimmung

## 2. Übereinstimmung: Wie viele erfüllen das Kriterium?

**Maximal akzeptabler Unterschied in der Beurteilung der AF** (Skala von 100%-0%, Stakeholder Survey)

**→ < 25 Prozentpunkte Arbeitsfähigkeit**

Wenn 2 psychiatrische Gutachter unabhängig voneinander bei derselben vP die Arbeitsfähigkeit schätzen:

**«Wie gross darf aus Ihrer Sicht maximal die Differenz bezüglich Arbeitsfähigkeit sein?»**

	<b>Rechts- vertreter</b> (n=81)	<b>Psychiater</b> (Behandler) (n=242)	<b>Psychiater</b> (Gutachter) (n=114)	<b>Richter</b> (n=47)	<b>Administration</b> (Versicherer) (n=108)
<b>... im aktuellen Verfahren mit den bekannten Einschränkungen</b>	<b>15%</b> (10-20%)	<b>20%</b> (10- <b>25%</b> )	<b>20%</b> (10- <b>25%</b> )	<b>15%</b> (10-20%)	<b>10%</b> (10-20%)

*Schandelmaier et al. Stakeholder Survey Swiss Med Wkly 2015*

---

# Ergebnisse: Reliabilität und Übereinstimmung

## 2. Übereinstimmung: Wie viele erfüllen das Kriterium?

Maximal akzeptabler Unterschied in der Beurteilung der AF → ≤ 25 Prozentpunkte Arbeitsfähigkeit

### Beispiel

GA Amann beurteilt 30% AF

GA Bolzli beurteilt 50% AF → Unterschied 20%-Punkte AF ✓

GA Zapf beurteilt 70% AF ⇒ Unterschied 40%-Punkte AF ∅

# Ergebnisse: Reliabilität und Übereinstimmung

## 2. Übereinstimmung: Wie viele erfüllen das Kriterium?

	Übereinstimmung	
	2) Wie gut können 2 GA 'die gleiche' Arbeitsfähigkeit einschätzen?  'gleich' = $\leq 25$ Prozentpunkte AF	Streuungsmass 3) Standard Error of Measurement  Kleiner ist besser
<b>RELY 1</b> n=177 Vergleiche	<b>61.6% der Vergleiche</b> (109/177 Vergleiche)	<b>24.6 %-Punkte AF</b>
<b>RELY 2</b> n=231 Vergleiche	<b>73.6% der Vergleiche</b> (170/231 Vergleiche)	<b>19.4 %-Punkte AF</b>

## 2) Übereinstimmung: a) Anteil Vergleiche, die das Kriterium erfüllen



= 109/177 Vergleiche (61.6%)

### RELY 1

#### Kriterium:

Die Beurteilung AF von 2 GA liegt weniger als 25 %-Punkte auseinander

Anzahl Vergleiche

..

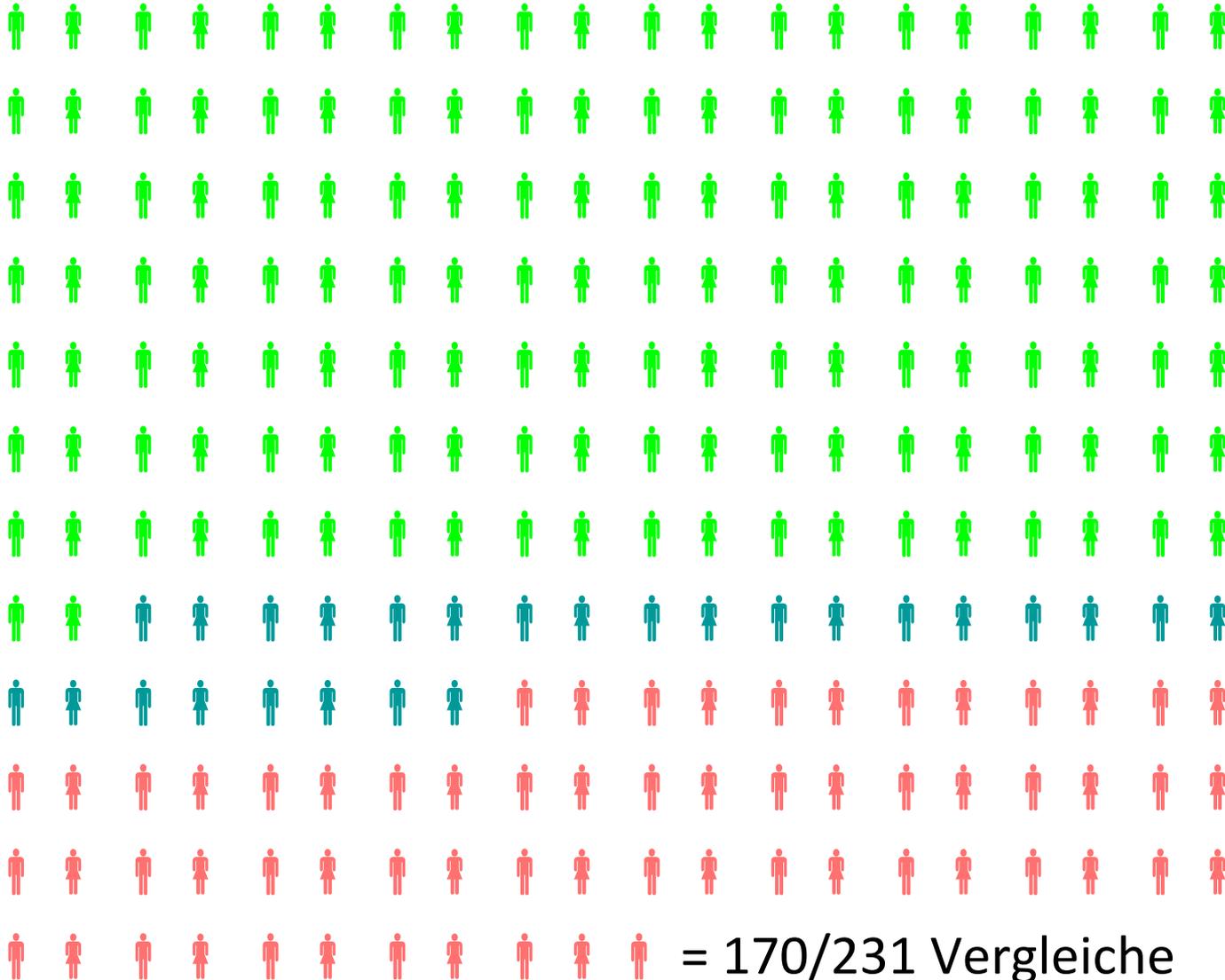
 : erfüllen das Kriterium

 : erfüllen das Kriterium nicht

### RELY 1:

Unterschied zwischen 2 Gutachtern war **bei 38.4% der Vergleiche grösser als 25 Prozentpunkte AF**

## 2) Übereinstimmung: a) Anteil Vergleiche, die das Kriterium erfüllen



### RELY 2

#### Kriterium:

Die Beurteilung  
AF von 2 GA  
liegt weniger als  
25 %-Punkte  
auseinander

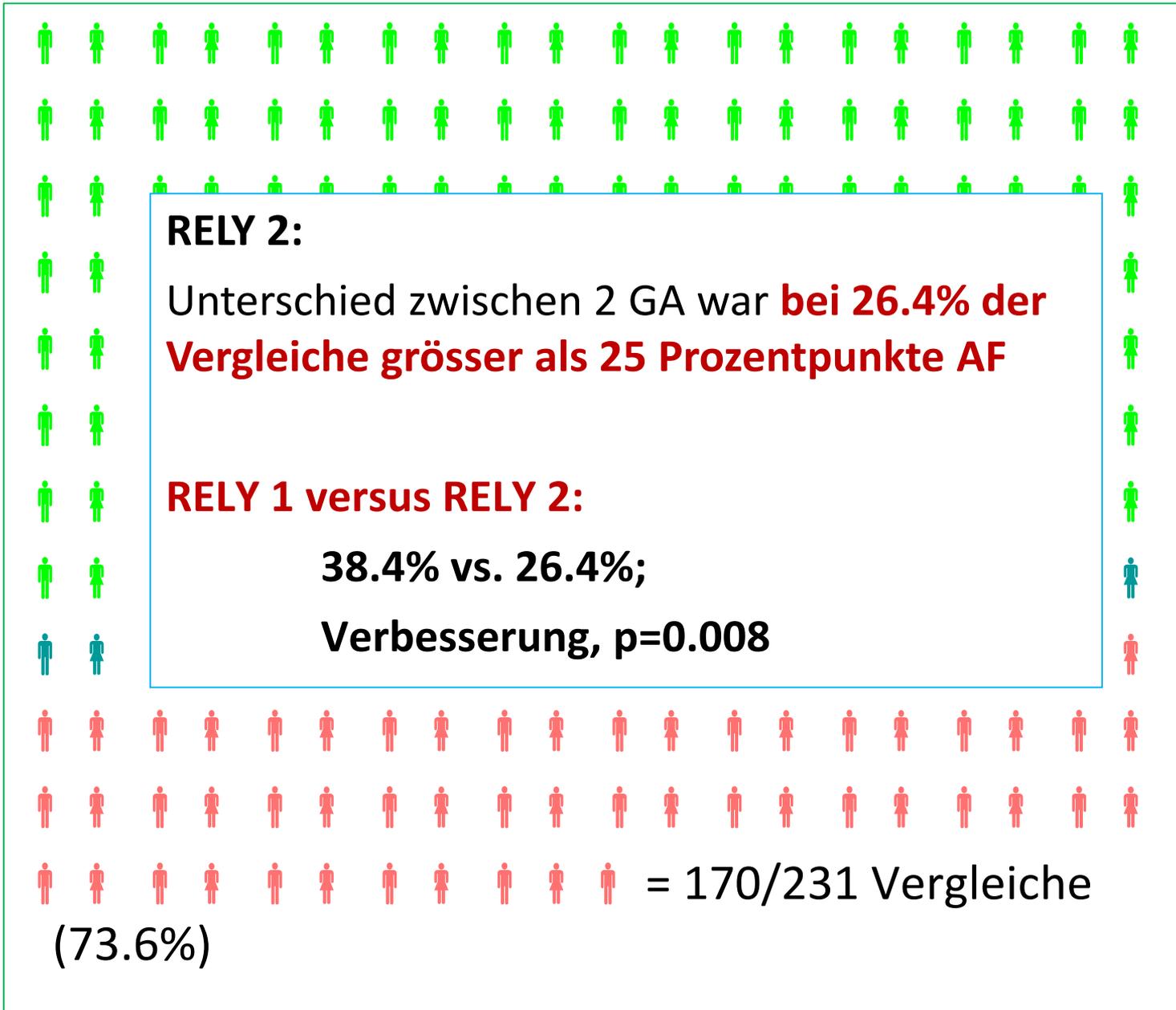
Anzahl Vergleiche

 : erfüllen das  
Kriterium

 : erfüllen das  
Kriterium nicht

 : erfüllen neu  
das Kriterium  
(=Verbesserung  
in RELY 2)

## 2) Übereinstimmung: a) Anteil Vergleiche, die das Kriterium erfüllen



## RELY 2

### Kriterium:

Die Beurteilung AF von 2 GA liegt weniger als 25 %-Punkte auseinander

# Ergebnisse: Reliabilität und Übereinstimmung

## 3. Übereinstimmung: Streuung

	Übereinstimmung	
	a) Wie gut können 2 GA 'die gleiche' Arbeitsfähigkeit einschätzen?  'gleich' = $\leq 25$ Prozentpunkte AF	Streuungsmass b) SEM, Standard Error of Measurement  Kleiner ist besser
<b>RELY 1</b> n=177 Vergleiche	<b>61.6% der Vergleiche</b> (109/177 Vergleiche)	<b>24.6 %-Punkte AF</b>
<b>RELY 2</b> n=231 Vergleiche	<b>73.6% der Vergleiche</b> (170/231 Vergleiche)	<b>19.4 %-Punkte AF</b>

Wenn 2 psychiatrische Gutachter unabhängig voneinander beim selben Patienten die Arbeitsfähigkeit schätzen:

**«Wie gross darf aus Ihrer Sicht maximal die Differenz bezüglich Arbeitsfähigkeit sein?»**

	<b>Rechts- vertreter</b> (n=81)	<b>Psychiater</b> (Behandler) (n=242)	<b>Psychiater</b> (Gutachter) (n=114)	<b>Richter</b> (n=47)	<b>Administration</b> (Versicherer) (n=108)
<b>... im aktuellen Verfahren mit den bekannten Einschränkungen</b>	<b>15%</b> (10-20%)	<b>20%</b> (10- <b>25%</b> )	<b>20%</b> (10- <b>25%</b> )	<b>15%</b> (10-20%)	<b>10%</b> (10-20%)

*Schandelmaier et al. Stakeholder Survey Swiss Med Wkly 2015*

# Berechnung SEM für maximal akzeptable Unterschiede

a) Erwartung der Stakeholder		b) Beobachtet in den RELY Studien			
Erwarteter 'Maximal akzeptabler Unterschied'*	Berechneter 'Standard error of measurement'			Beobachteter 'Standard error of measurement'	Berechneter 'Maximal akzeptabler Unterschied'
25% AF	9.0% AF	Angepasste Tätigkeit	RELY 1	24.6% AF	68.1% AF
			RELY 2	19.4% AF	53.9% AF

de Vet 2006

---

# Unsere Ergebnisse zur Reproduzierbarkeit (1)

Verglichen mit RELY 1 ist in RELY 2

**1) Gesamtvarianz + Psychiatervarianz (Streuung):** signifikant reduziert

**2) Keine Besserung in der Reliabilität**

Fähigkeit der GA, Personen mit leichter, mittlerer, schwerer Einschränkung der AF zu unterscheiden

**3) Messfehler bei GA verringert: Übereinstimmung** unter den Gutachtern in beiden Kenngrössen **signifikant verbessert:**

- Anteil GA mit akzeptabler Abweichung ist von 61.6% auf 73.4% (um 19%) angestiegen
- SEM (Streuungsmaß) zeigt zwischen GA ca. 20% weniger Streuung

**4) Trotzdem: Grosse Unterschiede zwischen Gutachtern bleiben deutlich hinter den Erwartungen der Akteure zurück**

---

# Transparenz und Nachvollziehbarkeit

## Feedback Psychiater Training

### RELY 1

#### Telefoninterview (N = 18)

- Vom Training profitiert (alle)
- Nutzung des funktionsorientierten Interview in ihrer Praxis (N = 11<sup>+</sup>)
- Mehr Training (N = 7)

### RELY 2

#### Online Fragebogen (N = 24)

- Vom Training profitiert (alle)
- Begutachtungen jetzt funktionsorientierter und strukturierter (N = 22)
- Künftiges Angebot sollte sein...
  - freiwillig mit Empfehlung SGVP/SGPP (N = 14)
  - verpflichtend durch FMH, SIM, SGVP/SGPP, IV, Suva (N = 10)

# Feedback Antragsteller Zufriedenheit

Frage	RELY 1 (N = 30)	RELY 2 (N = 25)
<b>Globale Zufriedenheit (Skala 0-10), M</b>	<b>8.0</b>	<b>9.4</b>
Frage	Antwort 'JA'	Antwort 'JA'
Fragt der Psychiater «Was möglich, was nicht»?	24 (80%)	22 (88%)
Fragt der Psychiater nach Arbeitseinschränkungen?	25 (83%)	24 (96%)
Konnten Sie alles Wichtige sagen?	29 (97%)	25 (100%)
Ging der Psychiater respektvoll mit Ihnen um?	28 (93%)	25 (100%)

---

# RELY 3 – Fokusgruppen mit Versicherern

## Akzeptanz

- Eindeutige und strukturierte Aussagen über die Arbeitsfähigkeit
- Idealerweise führt das zu weniger Rückfragen über den medizinischen Sachverhalt
- Trägt zur Transparenz der Gutachten bei

## Implementation in Praxis

- Sollte ohne Schwierigkeiten implementierbar sein
- Ausreichende Schulung notwendig
- Auf Kosten achten

---

## Ergebnisse (2)

### Bessere Transparenz und Nachvollziehbarkeit

Gutachter: 'Können unsere Arbeit besser machen'

Patienten: 'Fühlen uns ernst genommen'

Versicherer: 'Können gutachterliche Empfehlung  
besser nachvollziehen'

---

**Erst die Mitarbeit vieler Menschen hat die  
RELY-Studien möglich gemacht,  
allen ein grosses Dankeschön**

den Psychiatern

den Patienten

den MEDAS-Instituten

den Mitarbeitern der IV-Stelle Zürich

der Begleitgruppe

der FIP-Gruppe

den Behindertenverbänden

den Fachgesellschaften

den Anwälten und Richtern

den Versicherern

---

## **Ein grosses Dankeschön auch an meine Mitarbeitenden und KollegInnen**

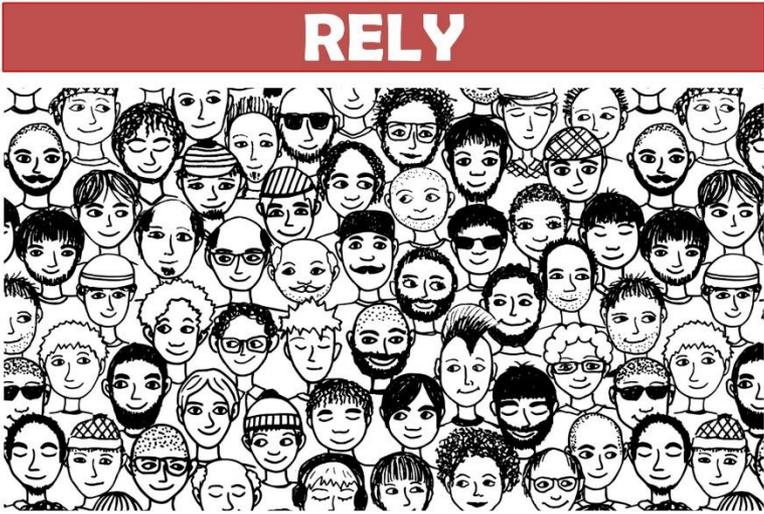
Wout de Boer, Katrin Fischer, David von Allmen, Monica Bachmann, Nicole Vogel, Jason W. Busse, Thomas Zumbrunn

## **den Mitgliedern der FIP-Gruppe**

Renato Marelli, Martin Eichhorn, Ulrike Hoffmann-Richter, Joerg Jeger, Ralph Mager, Etienne Colomb, Heinz J. Schaad

## **den Finanzierern**

Schweizerischer Nationalfonds,  
Bundesamt für Sozialversicherungen, Suva



Prof. Regina Kunz • Dr. David von Allmen • Prof. Katrin Fischer • Dr. med. Wout de Boer • Dr. med. Renato Marelli • Dr. med. Martin Eichhorn • Thomas Zumbunn • Dr. med. Jörg Jeger • Dr. med. Ulrike Hoffmann-Richter • Prof. Ralph Mager • Dr. med. Etienne Colomb • Dr. med. Heinz Schaad • Dr. Monica Bachmann • Nicole Vogel • Prof. Jason Busse • Dr. med. Oskar Bänziger • Brigitte Walter Meyer • Sacha Röschard • Dr. med. Stefan Schandelmaier • Prof. Gordon Guyatt • lic, iur. Yvonne Bollag • Regina Altermatt • Corinne Schraner • Daniel Hess • Andrea Leibold • Dr. med. Ronald Walshe • Heidrun Demirden • Silvia Joder • José Staff • Astrid Palca • Dr. med. Roderich Koesele • Sarah Kedzia • Raphaël Dettwiler • Prof. Wolf Langewitz • Helena Langewitz • Dr. med. Olaf Hentrich • Claudia Bretscher • Dr. Andreas Brunner • Dr. med. Walter Gekle • Martin Schilt • Prof. Ueli Kieser • Dr. Volker Pribnow • Martin Reinert • Dr. med. Fulvia Rota • Dr. med. Rita Schaumann-von Stosch • Michael Stiebel • Dr. Andreas Traub • Marc Gysin • Peter Eberhard • Dr. med. Marco Bachmann • Dr. med. Roman Fischer • Dr. med. Natalie Franke • Dr. med. Jan Felix Hoffmann • Dr. med. Andreas Moldovanyi • Dr. med. Konstantin Moskvitin • Dr. med. Joachim Nelles • Dr. med. Thomas Weber • Dr. med. Peter Keel • Dr. med. Martina Korthal Altes • Dr. med. Tim Niemeyer • Dr. med. Thomas Ihde • Dr. med. Heribert Pizala • Dr. med. Felix Schwarzenbach • Dr. med. Vreni Häller • Dr. med. Karen Fürstenau • Dr. med. Armin Walter • Dr. med. Andreas Linde • Dr. med. Thomas Fellmann • Dr. med. Andres Howald • Dr. med. Christoph Feinendegen • Dr. med. Gerhard Mohr • Dr. med. Arno Zormann • Dr. med. Julia Röseler • Dr. med. Markus Guzek • Dr. med. Bernard Minder • Dr. med. Ueli Blumer • Dr. med. Axel Wallossek • Dr. med. Monika Diethelm-Knoepfel • Dr. med. Lars Hermann • Dr. med. Elmar Meichtry • Dr. med. Andreas Kaldune • Dr. med. Christoph Lenk • Dr. med. Marita Manheim • Dr. med. Stefan Freidel • Dr. med. Beat Schaub • Dr. med. Daniel Thommen • Dr. med. Michael Huber • Dr. med. Helen Klieber • Dr. med. Thomas Cotar • Dr. med. Sabina Kenk Meisser • Frau Beate Martin • Nicole Bruni •