



UNIVERSITÄTS**medizin.**

MAINZ

Benutzung von Skalen für patientenberichtete Endpunkte: Herausforderungen bei der Entwicklung, Validierung, Anwendung und Interpretation

Susanne Singer

Abt. Epidemiologie und Versorgungsforschung

Institut für Biometrie, Epidemiologie und Informatik (IMBEI)

Gliederung

Herausforderungen bei der...

1. Entwicklung
2. Validierung
3. Anwendung
4. Interpretation

1. Entwicklung

Phasen der Entwicklung von PRO

Idee

Phase I: Issues



- **Herausforderungen:**

- die wirklich relevanten Punkte herausfiltern
- nicht zu wenig, nicht zu viel
- (alle) relevanten Patientengruppen einschließen

Beispiel: COVID19-QLQ

Phasen der Entwicklung von PRO

Idee

Phase I: Issues

Phase II: Items

- **Herausforderungen:**

- eindeutige Übersetzbarkeit
 - *schwierig*: „bothered by...“
 - *besser*: „problems“, „difficulties“
- Übertragbarkeit: Indikation

Übertragbarkeit Indikation

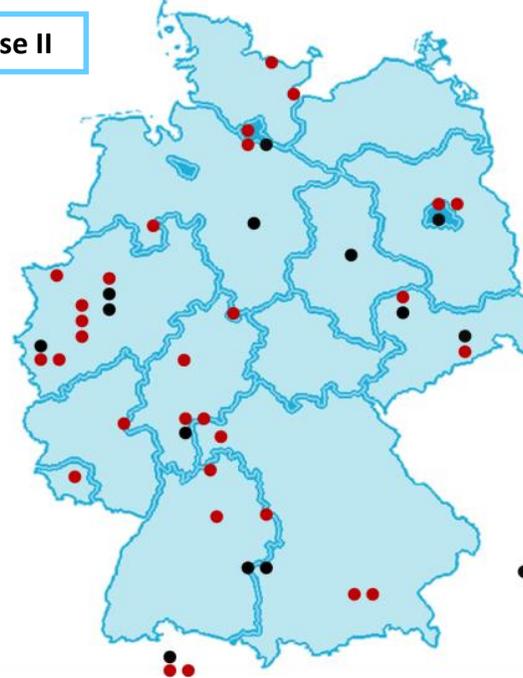
Beispiel: Krebs (EORTC QLQ-C30) → Aplastische Anämie/PNH

Phase I



(● Patienten, n=19 ; ● Ärzte, n=8)

Phase II



(● Patienten, n=30 ; ● Ärzte, n=14)

Neue (spezifische) Items

- Lebensqualität:
 - z. B. „Haben Sie versucht, Sachen zu vermeiden, die zu einem Infekt führen könnten (z.B. Nutzung von Bus & Bahn)?“
- Versorgung:
 - z. B. „Ich fühlte mich durch die Konfrontation mit Krebspatienten und dem Tod im Krankenhaus bzw. in der Praxis belastet.“
 - z. B. „Es störte mich, dass die Ärzte und das Pflegepersonal sich nicht mit meiner Erkrankung auskannten.“

Phasen der Entwicklung von PRO

Idee

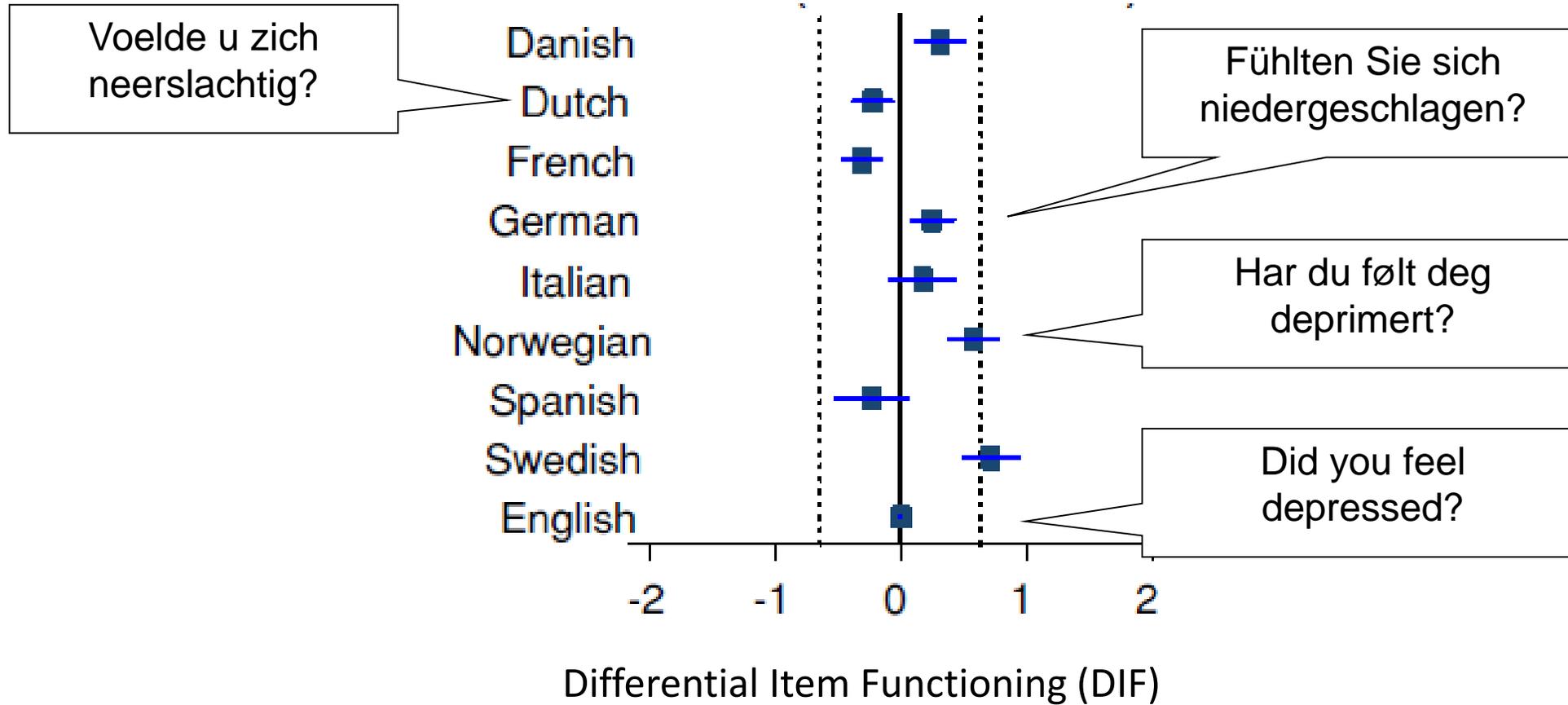
Phase I: Issues

Phase II: Items

- **Herausforderungen:**

- eindeutige Übersetzbarkeit
 - *schwierig*: „bothered by...“
 - *besser*: „problems“, „difficulties“
- Übertragbarkeit: klinisch
- Übertragbarkeit: linguistisch
 - „Koffer tragen können“ – benutzt man überall Koffer?
 - „weite Strecke gehen können“ – was heißt weit?
 - „2 km gehen“ – rechnen alle in km?

Übertragbarkeit Sprache / Kultur



Phasen der Entwicklung von PRO

Idee

Phase I: Issues

Phase II: Items

Phase III: Pilotierung

- **Herausforderungen:**

- Pilotierung des Wordings UND Exploration der Skalenstruktur
- Mindestmenge an Sprachen
- Mindestmenge an Patienten pro Sprache
- Management der verschiedenen Sprachversionen



EORTC QLQ - H&N35

Patienten berichten manchmal die nachfolgend beschriebenen Symptome oder Probleme. Bitte beschreiben Sie, wie stark Sie diese Symptome oder Probleme während der letzten Woche empfunden haben. Wählen Sie bitte die Antwort aus, die am besten auf Sie zutrifft.

Während der letzten Woche:	Überhaupt			
	nicht	Wenig	Mäßig	Sehr
31. Hatten Sie Schmerzen im Mund?	1	2	3	4
32. Hatten Sie Schmerzen im Kiefer?	1	2	3	4
33. Hatten Sie wunde Stellen im Mund?	1	2	3	4
34. Hatten Sie Halsschmerzen?	1	2	3	4
35. Hatten Sie Probleme, Flüssiges zu schlucken?	1	2	3	4



EORTC QLQ - H&N35

חולים מדווחים לפעמים, שיש להם בעיות או סימפטומים המפורטים להלן. אנא ציין/ציני באיזו מידה חווית סימפטומים או בעיות כאלה במהלך שבעת הימים האחרונים. נא לענות על ידי הקפת עיגול מסביב לספרה המתאימה ביותר לגביך.

כלל לא	במידה מועטה	במידה די רבה	במידה רבה מאד	במהלך שבעת הימים האחרונים:
1	2	3	4	31. האם היו לך כאבים בפה?
1	2	3	4	32. האם היו לך כאבים בלסת התחתונה?
1	2	3	4	33. האם היתה לך רגישות בפה?
1	2	3	4	34. האם היו לך כאבים בגרון?



EORTC QLQ - H&N35

Больные иногда сообщают о наличии перечисленных ниже симптомов или проблем. Пожалуйста укажите, в какой степени эти симптомы или проблемы проявлялись у Вас на протяжении последней недели. При ответе на вопрос обведите цифру, которая лучше всего характеризует Ваше состояние.

В течение последней недели:	Не Слегка Сущест - Очень			
	было	было	венно	сильно
31. Испытывали ли Вы боли во рту ?	1	2	3	4
32. У Вас болели челюсти ?	1	2	3	4
33. Испытывали ли Вы повышенную чувствительность во рту?	1	2	3	4
34. Болело ли у Вас горло?	1	2	3	4



EORTC QLQ - H&N35

रूग्ण कधी त्यांना खालीलप्रमाणे त्रास होत आहे किंवा खालील लक्षणांपैकी काही लक्षणे दिसत आहे असे सांगत येतात. कृपा करून तुम्हाला खाली दिलेल्या लक्षणांपैकी कोणत्या बाबतीत किती त्रास गेल्या आठवड्यात झाला ते नमुद करा. ज्या प्रमाणात तुम्हाला त्रास झाला तो दर्शविणा-या आकड्याभोवती वर्तुळ करा

गेल्या आठवड्यात	अजिबात नाही	जरामे	बरेच	खूप जास्त
३१. तुमच्या तोंडात दुखत होते का?	१	२	३	४
३२. तुमच्या जबड्यात दुखत होते का?	१	२	३	४
३३. तुमचे तोंड आल्यासारखे वाटत होते का?	१	२	३	४
३४. तुमचा घसा दुखत होता का?	१	२	३	४

2. Validierung

Phasen der Entwicklung von PRO

Idee

Phase I: Issues

Phase II: Items

Phase III: Pilotierung

Phase IV: Validierung

- **Herausforderungen:**

- Änderungssensitivität vs. Test-Retest-Reliabilität

- Kürze vs. Komplexität

- usw.

- Wann ist man eigentlich fertig?

Wann ist ein Instrument “validiert”?

- **Validität** = Qualität der Inferenzen und Schlußfolgerungen, die man aus den Scores des Instrumentes zieht
- **Validierung** = Prozess, in dem wir Evidenz sammeln bzgl. Angemessenheit, Sinnhaftigkeit und Nutzbarkeit der Entscheidungen, die man basierend auf den Scores trifft

Also: eigentlich ist man nie fertig.
Es gibt nur ein „gut genug“.

Phasen der Entwicklung von PRO

Idee

Phase I: Issues

Phase II: Items

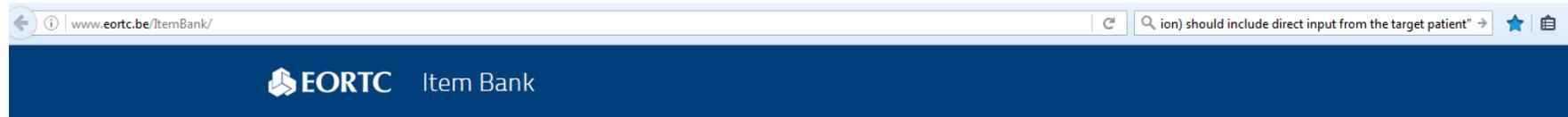
Phase III: Pilotierung

Phase IV: Validierung

- **Herausforderungen:**

- Änderungssensitivität vs. Test-Retest-Reliabilität
- Kürze vs. Komplexität
 - usw.
- Wann ist man eigentlich fertig?
- Dauer des Prozesses

- Datenbanken mit „validierten Items“ eine Lösung?



EORTC Quality of Life Group Item Bank

The Item Bank is a database of items used in fully and partially validated EORTC quality of life questionnaires.

This tool has been created by the EORTC Quality of Life Group with the primary aim of being used during the development of new instruments.

Developers can access it to see how items were formulated in previous questionnaires and to reuse existing validated items and their translations.

Access to the Item Bank can also be granted to academic and commercial users as a reference tool.

[Request access](#)

Log In

Email

Password

[Log in](#)

[Forgot your password?](#)

[No account? Request access](#)

67 questionnaires, 952 questions

Include custom questionnaires from the community

Official questionnaires

Custom questionnaires

Item classification

A

ANL Anal module
(27 questions)

B

BLM Bladder module
(30 questions)

BN Brain module
(20 questions)

BR Breast module
(45 questions)

BM Bone Metastases module
(22 questions)

BR23 Breast previous
(23 questions)

BRECON Breast Reconstruction module
(23 questions)

Welcome to the Item Library!

The **Core questionnaire** QLQ-C30 has been developed to assess the quality of life of cancer patients.

It is supplemented by disease-, symptom- and population-specific questionnaires called **modules**. You can also find general questionnaires that can be used without the Core questionnaire, called **standalones**.

Previous versions are questionnaires that have been updated.

The **CAT** groups all item banks for the Computer Adapted Testing system.

Categories

Probleme:

- Skalen validiert, aber nicht Items
- Inhaltsvalidität braucht Phase I

3. Anwendung

- **Bogen erfasst, was ich wissen möchte**
- **Akzeptanz bei Zielgruppe**
 - wenig fehlende Items (<5%)
 - wenig fehlende Fragebögen (je nach Anwendungsgebiet, ggf. niedrigere Quote in Terminalsituation)
 - Patienten bewerten Bogen als wichtig, freuen sich darüber
- **zu den besten Zeitpunkten ausgeben**
 - klinisch relevant
 - nicht zu häufig, nicht zu selten

Welches Instrument nehmen?

		FACT	EORTC	MDASI
valide	Patienten involviert in Entwicklung	+	+	(+)
	Items kombinierbar in Skalen	+	+	+
reliabel	Interne Konsistenz	+	+	+
	cross-kulturell validiert	(+)	+	(+)
	Antwortkategorien	5	4	10
	klares Zeitfenster	+	+	+
Akzeptanz	wenig fehlende Werte	(+)	(+)	(+)
	keine Boden- und Deckeneffekte	+	+	?
verfügbar	in verschiedenen Sprachen	38	91	7
	einfacher Zugang	online	online	per Email
differenziert	Anzahl Items	27	30	19
ökonomisch	Anzahl Items	27	30	19

ISOQOL (2013)

Qual Life Res (2013) 22:1161–1175

DOI 10.1007/s11136-012-0252-1

Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards

**Michael Brundage · Jane Blazeby · Dennis Revicki · Brenda Bass · Henrica de Vet ·
Helen Duffy · Fabio Efficace · Madeleine King · Cindy L. K. Lam · David Moher ·
Jane Scott · Jeff Sloan · Claire Snyder · Susan Yount · Melanie Calvert**

FDA (2021)

Core Patient-Reported
Outcomes in Cancer
Clinical Trials
Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

• Für Nutzer

- Gibt es überhaupt (ein) Instrument(e)? Wenn ja, welches nehmen?
- Welche Zeitpunkte?
- Kosten / Copyright?
- Einfache Auswertung?
- Akzeptanz bei EMA, FDA, IQWiG, G-BA, Community?

• Für Entwickler

- Ansprüche der Nutzer
- Management der Übersetzungen
- Management der Anfragen zu Scoring, Anwendung etc.

4. Interpretation

- MID

- MIC

- Daumenregeln
- Verteilungsbasiert
- Ankerbasiert

- Zielgrößen / Thresholds

- Vortrag Giesinger

- Referenzwerte

- Klinische Populationen
- Allgemeinbevölkerung

MID/MIC: Daumenregeln

- 10% Änderung der Gesamtskala
 - z. B. bei Skala 0-100: 10 Punkte = MID

MID/MIC: Verteilungsbasiert

- Standardabweichung:
 - Änderung $> \frac{1}{2} SD = MID$
- Standardfehler (SE):
 - Änderung $< SE$ ist wahrscheinlich Meßfehler und keine wirkliche Änderung
 - Änderung $> 1 SE = MID$
- Effektstärke
 - Unterschied zwischen Mittelwerten / SD des Baseline-Scores
 - z. B. $0.2 = MID$

MIC: verteilungsbasiert

	QoL evaluated in relation to	Advantages	Disadvantages
Paired t-statistics	standard error of the mean change	none	increases with N
Effect size (ES)	pre-test SD	standardised units benchmarks independent of N	decreases with increased baseline variability does not consider variability of change
Standardised response mean (SRM)	SD of change	standardised units independent of N	varies dependent of effectiveness of treatment
Responsiveness statistic (a variation of SRM)	SD of change in a stable group	standardised units independent of N takes measurement error into account	data on stable subjects often not available
Standard error of measurement (SEM)	Standard error of measurement	relatively stable across populations takes precision of the tool into account	assumes measurement error to be constant across the range of possible scores
Reliable change index (RCI)	Standard error of the measurement difference	relatively stable across populations takes precision of the tool into account	assumes measurement error to be constant across the range of possible scores

MID/MIC: Ankerbasiert

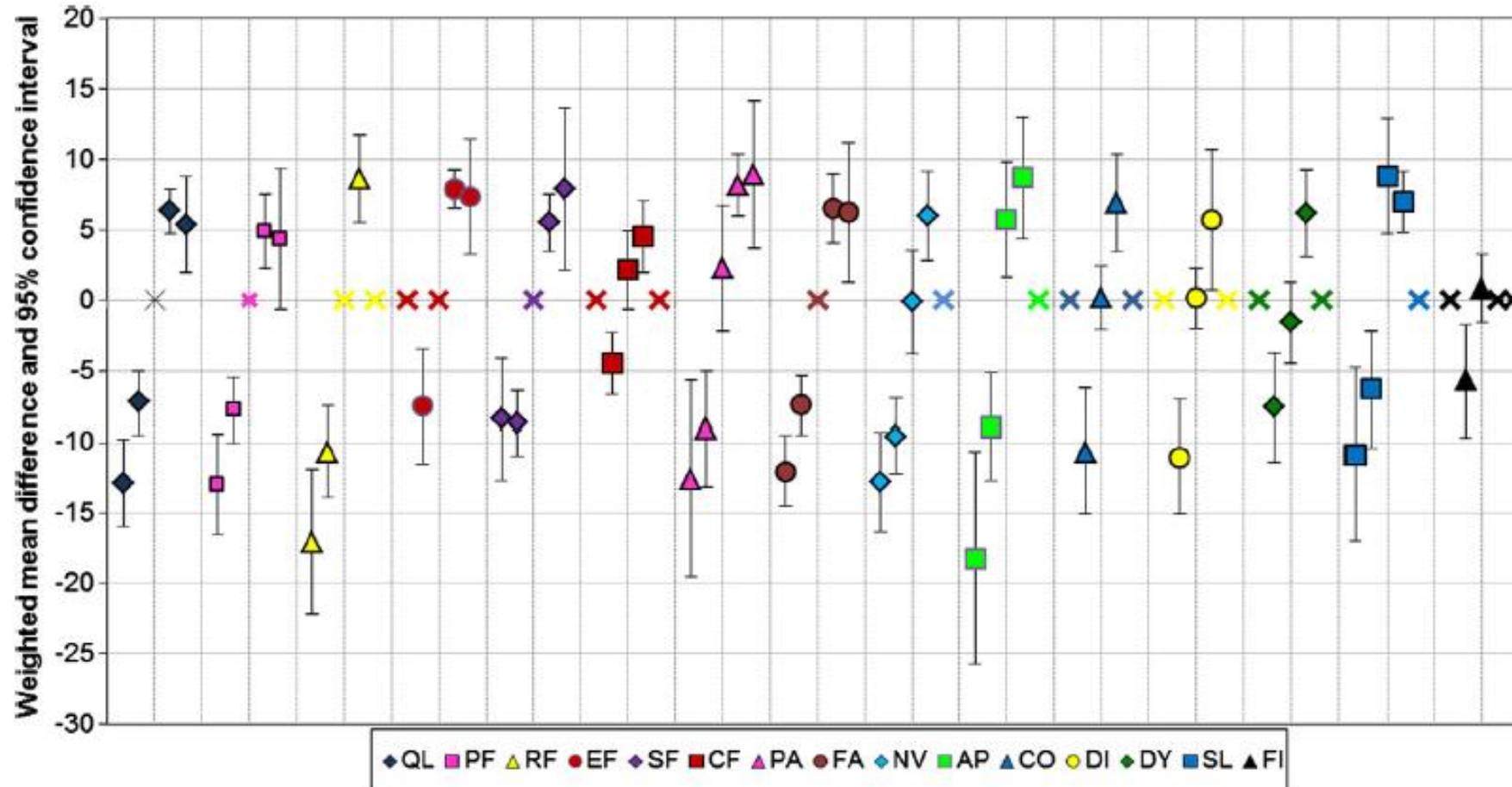
- Externe Kriterien (Berufsfähigkeit, Schlucktest, Performance Score usw.)
- Einschätzung durch Experten

MID/MIC: Ankerbasiert

- durch Experten:
 - **large:** deutlicher Effekt, unzweifelhaft klinisch relevant
 - **medium:** wahrscheinlich klinisch relevant, aber weniger großer Effekt
 - **small:** subtiler Unterschied, aber noch klinisch relevant
 - **trivial:** wahrscheinlich nicht klinisch relevant oder kein Unterschied
- Meta-Analyse publizierter Veränderungen über die Zeit

Beispiel für den EORTC QLQ-C30

n=118 Artikel, n=35 Experten



Pro Skala von links nach rechts: mittlere + kleine **Verschlechterung**, trivial, kleine + mittlere **Verbesserung**.
X = ungenügend Daten vorhanden, daher keine Berechnung

MID/MIC: Ankerbasiert

- Externe Kriterien (Berufsfähigkeit, Schlucktest, Performance Score usw.)
- Einschätzung durch Experten
- Einschätzung durch die Betroffenen selbst (Global Rating of Change)

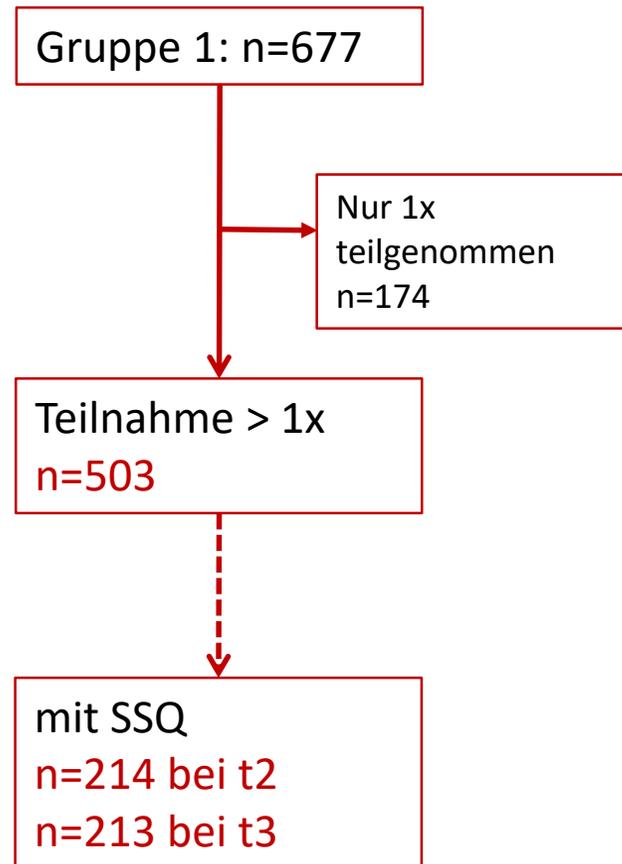
Subjective Significance Questionnaire

Im Vergleich zum letzten Mal, als ich den Fragebogen ausfüllte, ist meine Lebensqualität

sehr viel schlechter	etwas schlechter	ein bißchen schlechter	ungefähr gleich	ein bißchen besser	etwas besser	sehr viel besser

Und was soll daran jetzt so schwierig sein?

Beispiel: EORTC QLQ-HN43



Gruppe 2: n=135

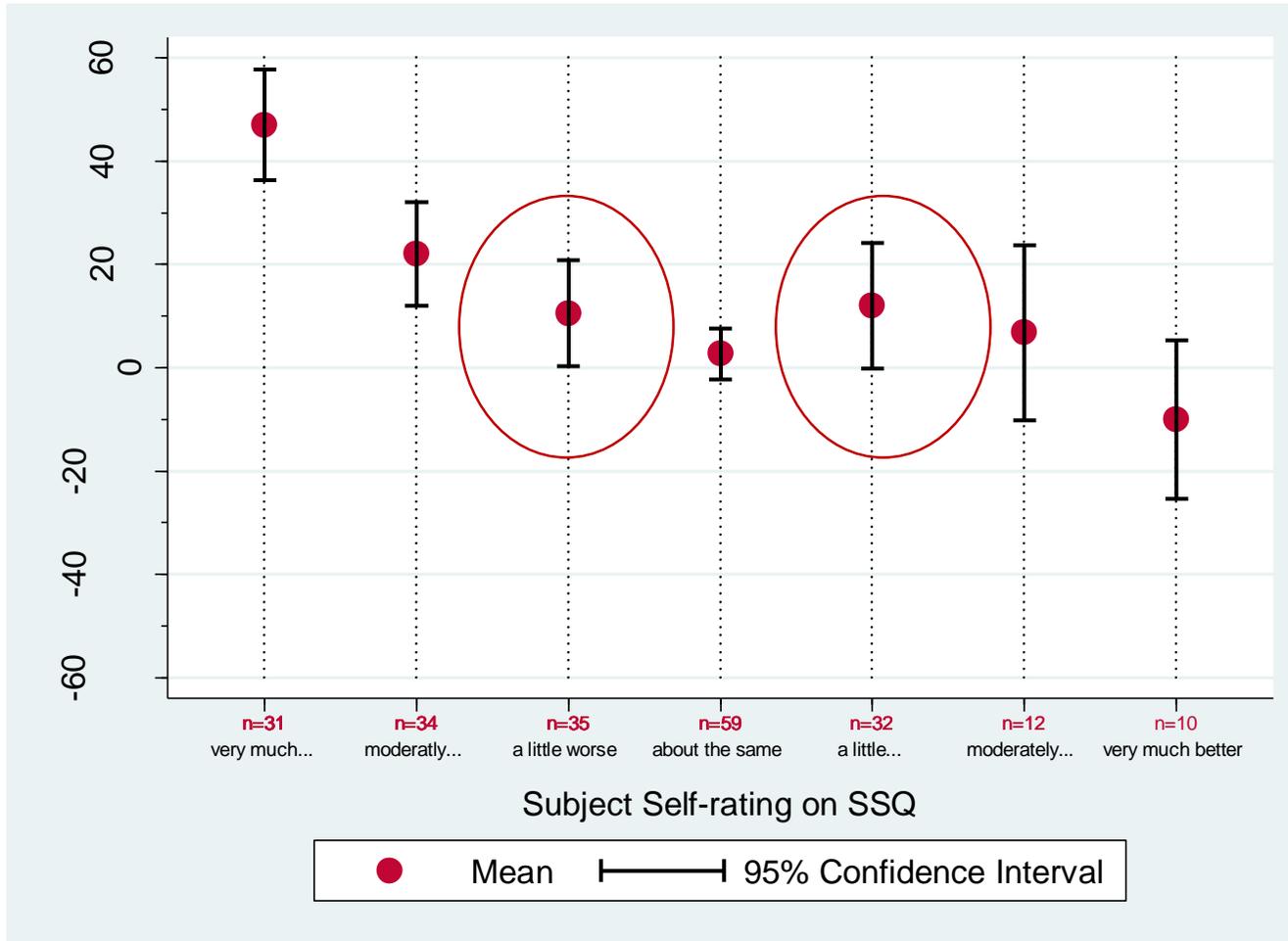
SSQ nötig für MIC
(nicht für MID)

Beispiel Schluckprobleme

... my swallowing is:

very much worse	moderately worse	a little worse	about the same	a little better	moderately better	very much better
n=31	n=34	n=35	n=59	n=32	n=12	n=10

Beispiel Schluckprobleme



die ersten 4 Patienten mit „a little worse“:

t1	t2	Delta t1-t2	SSQ
25	25	0	a little worse
0	0	0	a little worse
67	8	-58	a little worse
67	75	8	a little worse

Probleme: 1. nur ein Item
 2. Vergessen der Baseline?

Ergebnisse

	MID	MIC	
		Deterioration (t1 to t2)	Improvement (t2 to 3)
Anchor-based	mean difference of Swallowing at t2 in patients with KPS 60 vs. KPS 70 at t2	mean delta in patients who say that their swallowing is “a little worse”	mean delta in patients who say that their swallowing is “a little better”
	<i>discarded because of poor correlation</i>	11	-14
	mean difference of Swallowing at t2 in patients with KPS 70 vs. KPS 80 at t2	ROC derived cut-point	ROC derived cut-point
	<i>discarded because of poor correlation</i>	8	<i>discarded because of poor AUC</i>
		Based on predictive regression modelling	Based on predictive regression modelling
		15	-3
Distribution-based	0.5 SD of Swallowing at t2	0.5 SD of delta in Swallowing	0.5 SD of delta in Swallowing
	14	16	-12
	0.3 SD of Swallowing at t2	0.5 SD of delta in Swallowing	0.5 SD of delta in Swallowing
	10	10	-8
	SEM at t2	SEM of delta in Swallowing	SEM of delta in Swallowing
	11	12	-10

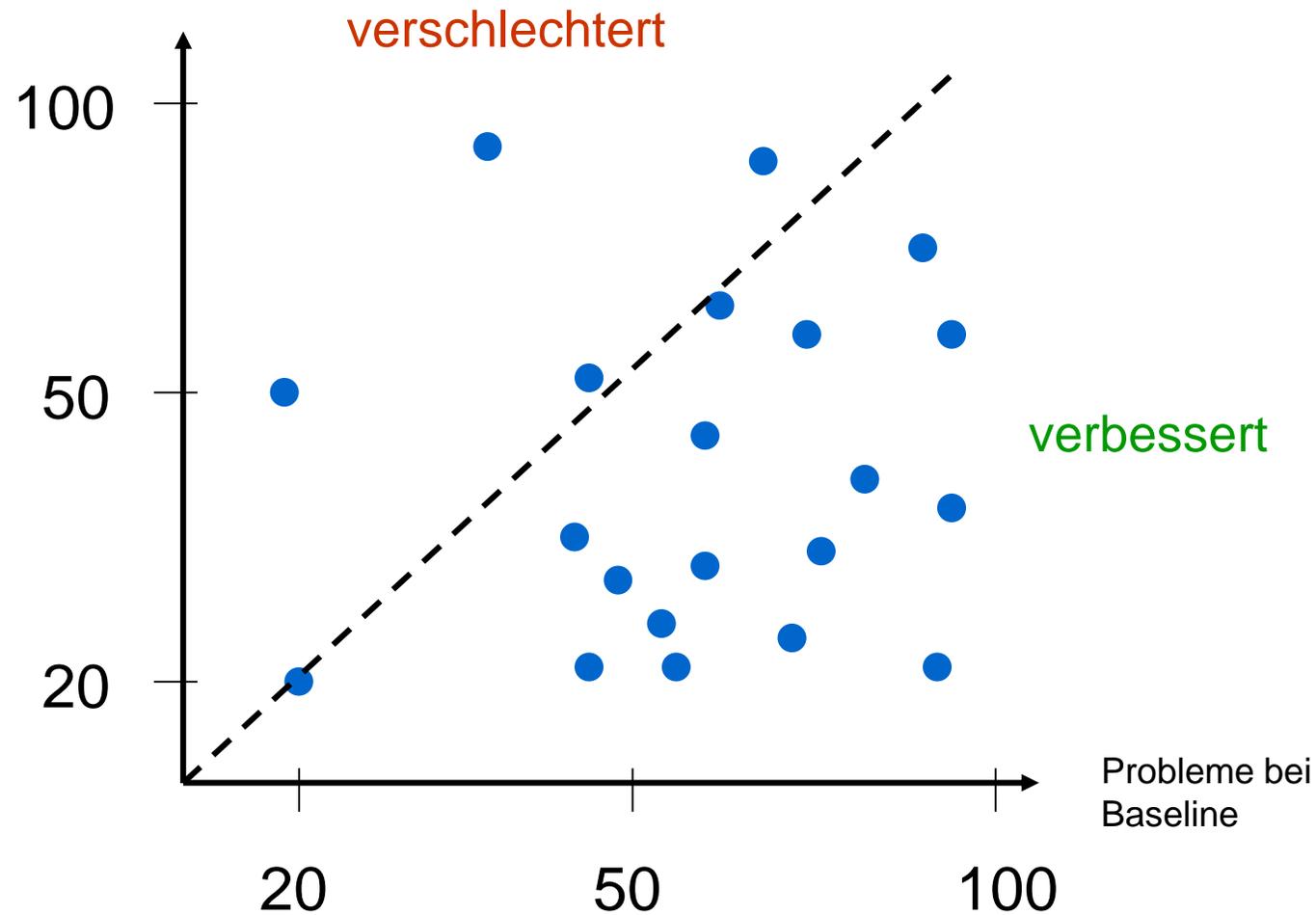
Herausforderungen MID / MIC

- Nutzer wollen einen Wert \leftrightarrow verschiedene Methoden liefern verschiedene Werte (wie integrieren?) (und manche Methoden funktionieren nicht gut)
- MID \leftrightarrow MIC
- Verbesserung \leftrightarrow Verschlechterung
- kann im Verlauf variieren
- Kein Charakteristikum eines Instrumentes, sondern einer Gruppe von Befragten

Weitere Herausforderungen

- Baseline-Unterschiede
- Boden- und Deckeneffekte

Probleme bei FUP

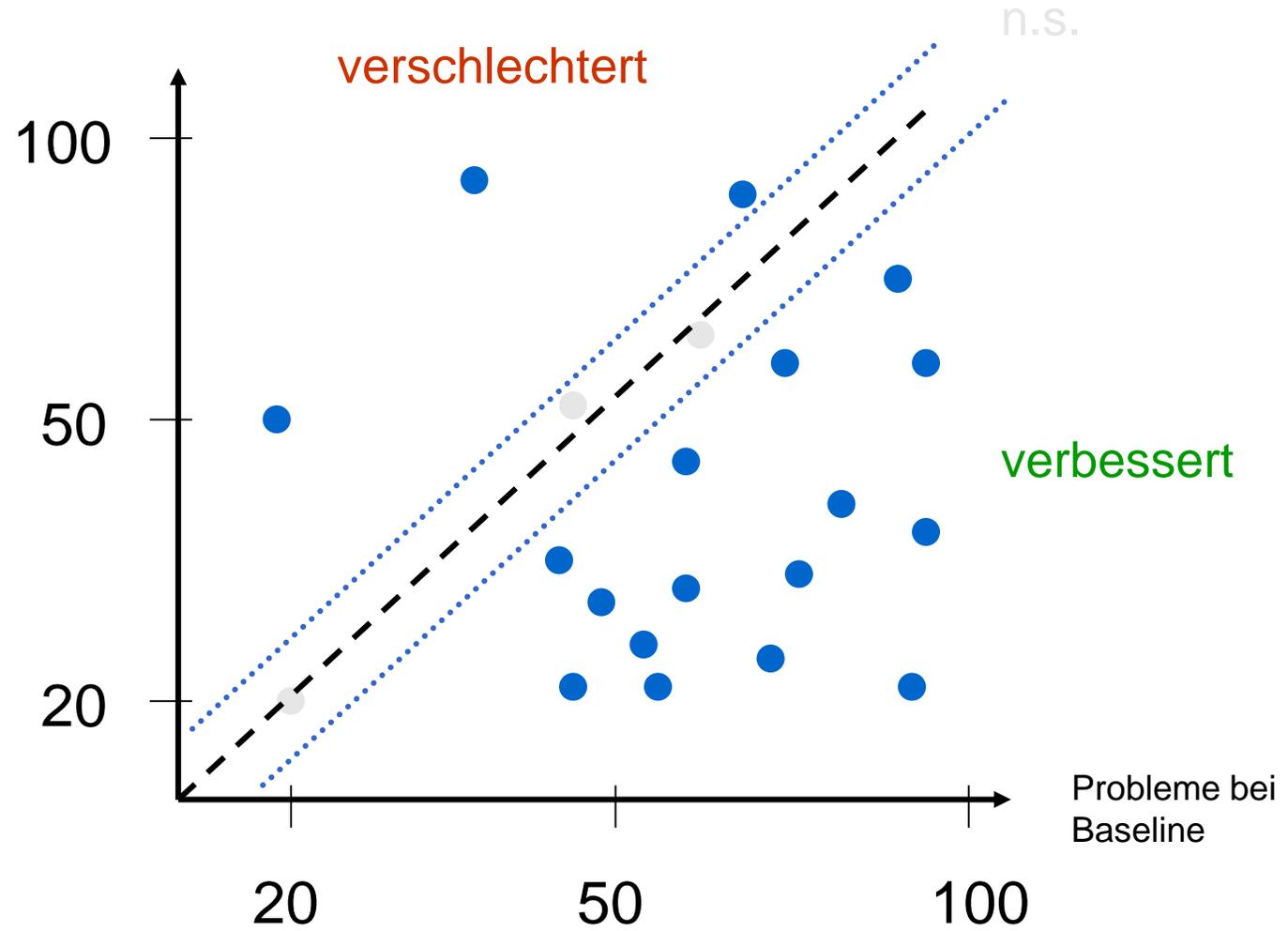


verschlechtert

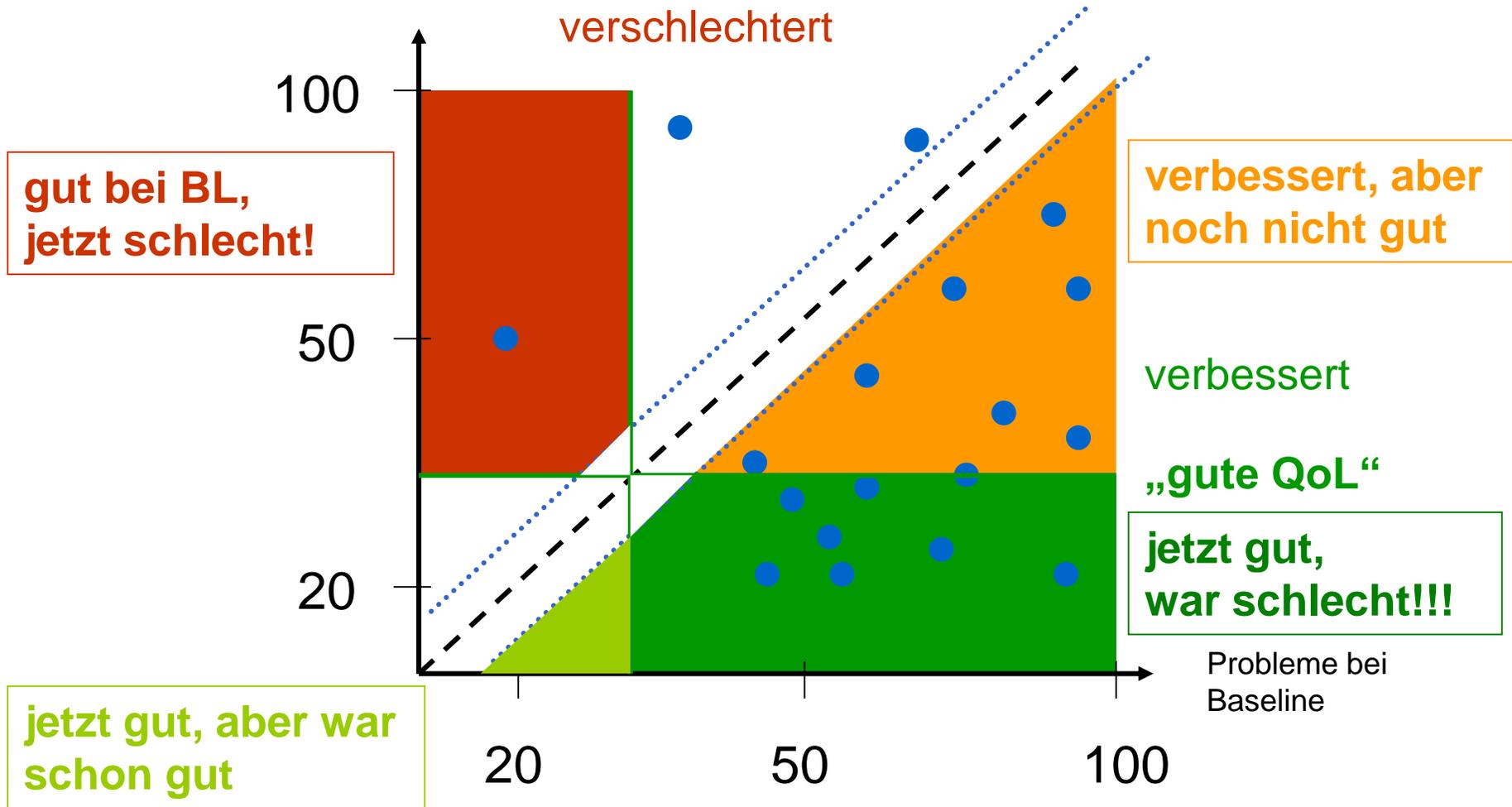
verbessert

Probleme bei
Baseline

Probleme bei FUP



Probleme bei FUP



Beispiel: Rehabilitation nach Laryngektomie

Natürlich wünscht sich jeder ein Leben frei von jeglichen Beschwerden. Wir wissen aber, dass Menschen ohne Kehlkopf nie völlig beschwerdefrei sein können.

Uns interessiert Ihre persönliche Meinung:

Welchen Wert auf der Skala sollte ein Kehlkopfloser ca. 1 Jahr nach der OP wenigstens erreicht haben, um eine akzeptable Lebensqualität zu besitzen?

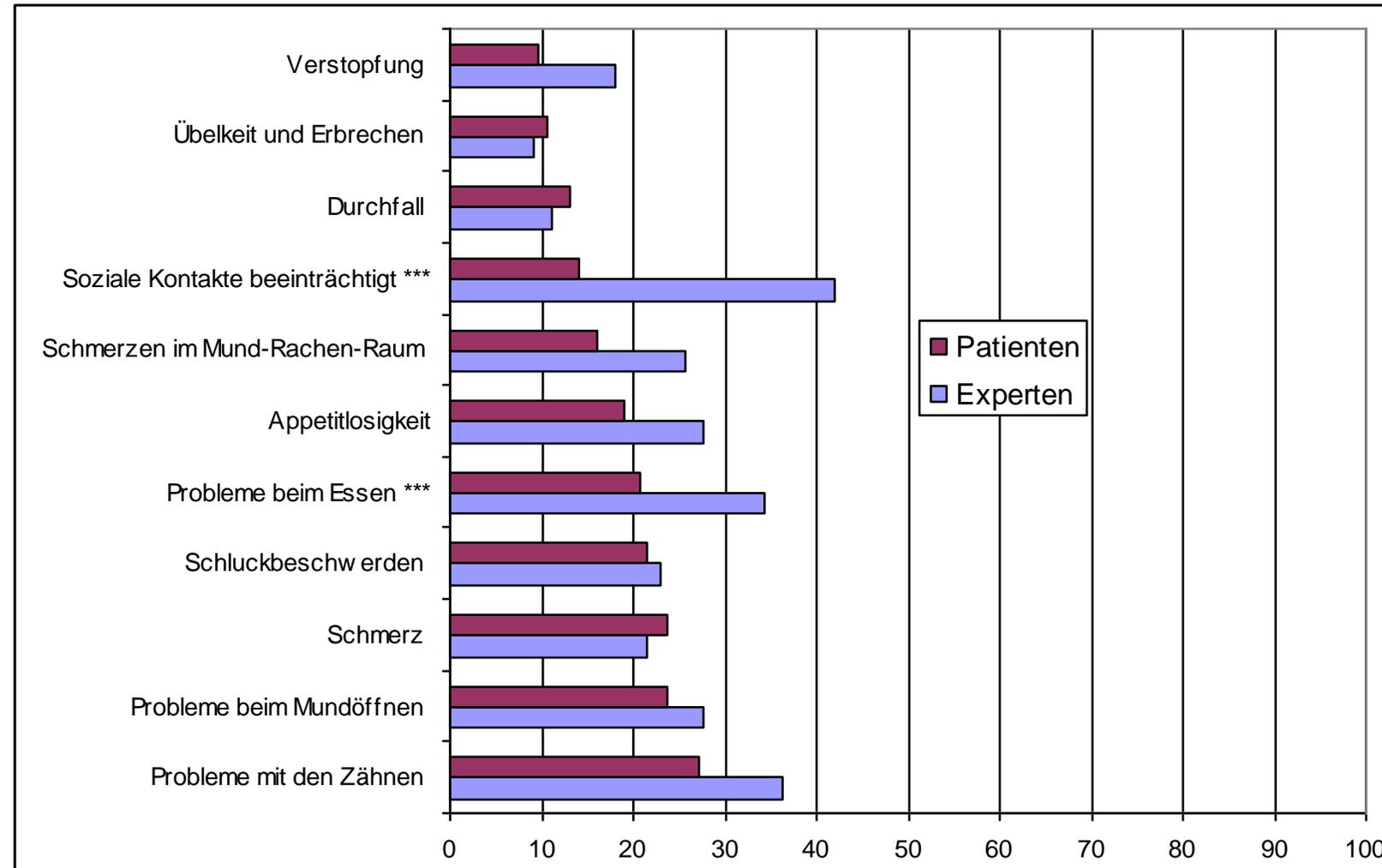
Hier ein Beispiel, um Ihnen die Bearbeitung der Fragen zu erklären:

	Überhaupt nicht 0	Wenig 1	Mäßig 2	Sehr 3
Haben Sie Probleme bei der Benutzung ihrer Schultern?		X		

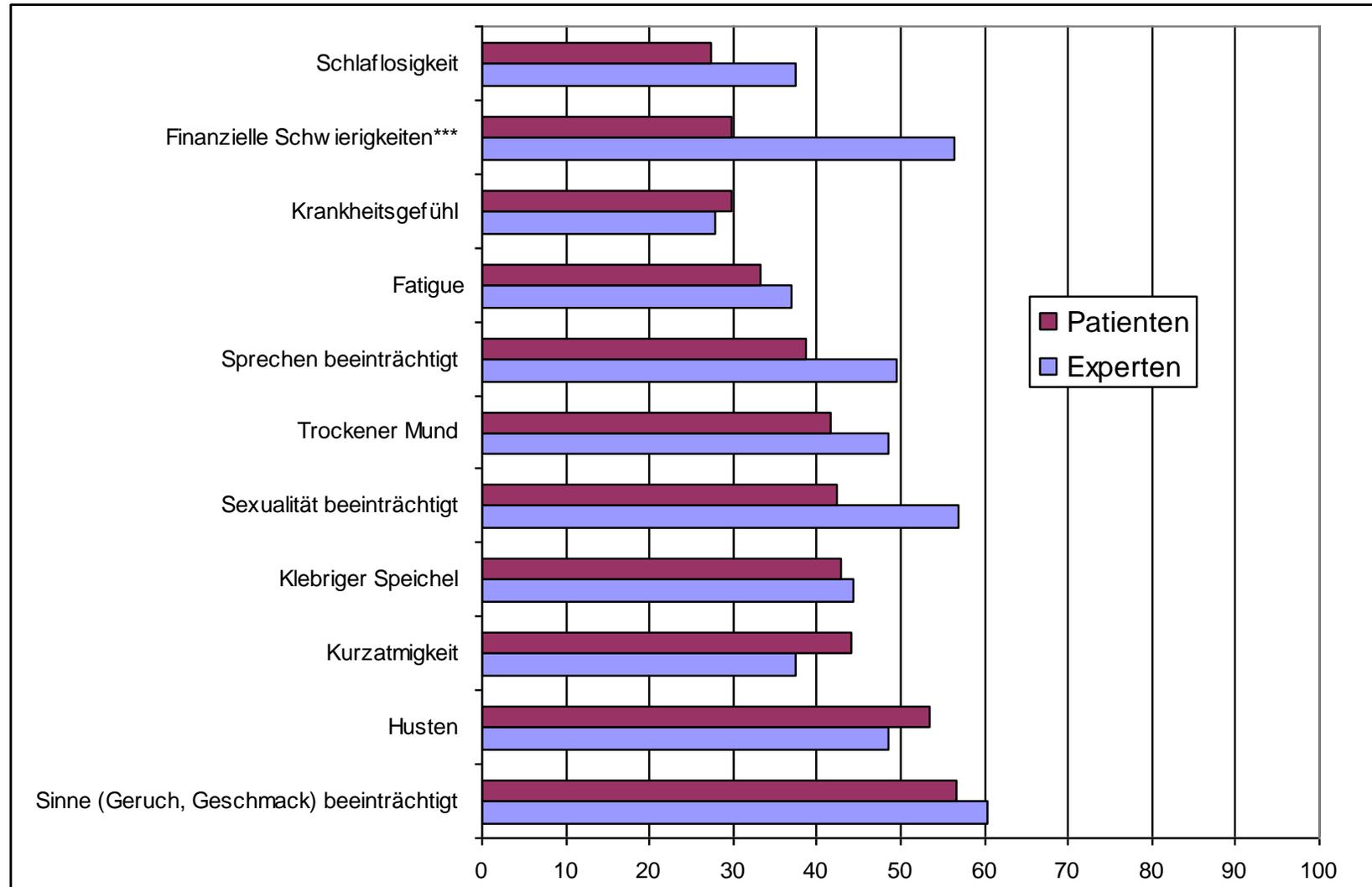
- d.h., **wenig** Probleme bei der Benutzung ihrer Schultern halten Sie für akzeptabel

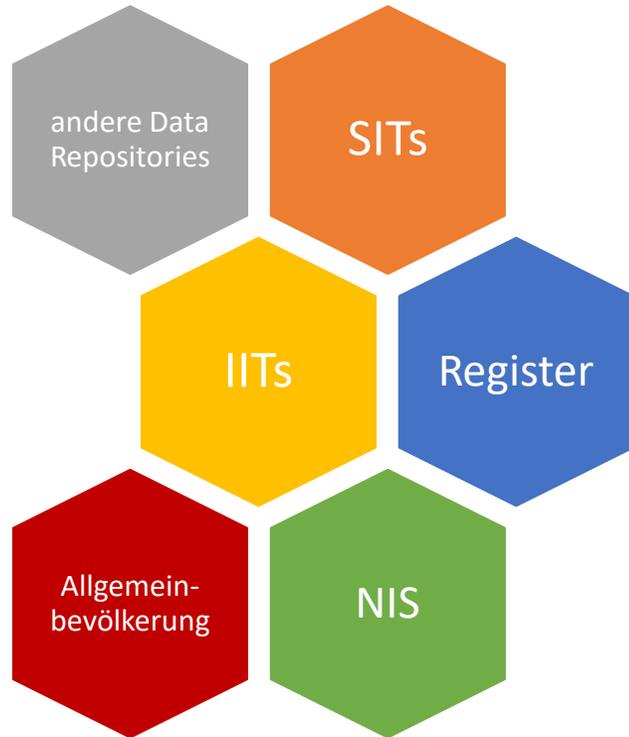
	Überhaupt nicht 0	Wenig 1	Mäßig 2	Sehr 3
s01 Bereitet es Ihnen Schwierigkeiten, sich körperlich anzustrengen (z.B. eine schwere Einkaufstasche oder einen Koffer zu tragen?)				
s02 Bereitet es Ihnen Schwierigkeiten, einen <u>längeren</u> Spaziergang zu machen?				

Beispiel: Rehabilitation nach Laryngektomie



Beispiel: Rehabilitation nach Laryngektomie





- **Herausforderungen:**

- Statische Manuale → Interaktive Datenbankabfragen
- Bereitschaft zum Teilen von Daten
- Einverständnis der Befragten
- Variablenmapping
- Harmonisierung klinischer Daten
- Aktualität der Daten
- Zugang zu Daten nur auf Plattform der Firma
- u.a.

Fazit

1. Es ist schwieriger, als man so denkt.
2. Wir müssen weiter nach Lösungen suchen.
3. Und keine Angst vor Holzwegen haben.