

# Patientenberichtete Endpunkte:

Wie können aussagekräftige minimale  
patientenrelevante Unterschiede hergeleitet werden?

Jens-Otto Andreas

18.JUN.2021



Vivienne, living with osteoporosis

# Disclaimer

- Die in dieser Präsentation geäußerten Ansichten sind meine persönlichen Ansichten und nicht unbedingt die Ansichten von UCB oder anderer Institutionen

# Agenda

## Definition

- PRO, MID / Individuelle Response

## Herleitung/Validierung von MIDs

- Verteilungsbasierte Verfahren
- Ankerbasierte Verfahren

## Credibility Tool der McMaster University, Canada

## Responseschwelle von $\geq 15\%$ der Skalenspannweite

## Ausblick

# Definition

## PRO, MID / Individuelle Response

### ■ Patientenberichtete Endpunkte (Patient-Reported Outcomes, PRO)

- „A PRO is an umbrella term used to describe any outcome evaluated directly by the patient himself/herself, without interpretation by clinicians or others, and based on patients' perception of a disease and its treatment ”  
*(EUNetHTA guideline: Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints)*

### ■ Minimal Important Difference (MID)

- Kleinster Unterschied zwischen zwei Messwerten, den ein Patient für bedeutsam hält.  
*(Brettschneider, Lüthmann, Raspe 2011; DIMDI Der Stellenwert von PROs im Kontext von HTA )*

### ■ Im Rahmen unserer Diskussion geht es um die individuelle Response eines Patienten im Sinne einer MID

- Responder Definition: “The individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit”.  
*(US Food and Drug Administration. (2009). Guidance for industry on patient-reported outcome measures: Use in medical product development to support labeling claims.)*

# Herleitung/Validierung von MIDs

## | Frage

- Gibt es überhaupt die MID für ein Instrument?
  - Ist diese eindeutig zu bestimmen?
- Ähnliche Frage in der frühen Entwicklung (Dosisfindung)
  - Gibt es eine Minimal Effective Dose (MED)?
- Ziel sollte eine möglichst gute Annäherung sein

# Herleitung/Validierung von MIDs

## | Verteilungsbasierte Verfahren

### ■ Positiv

- Berechnung z.B. mittels des Standardfehlers (SEM) oder  $\frac{1}{2}$  Standardabweichung
- Einfach zu implementieren, da im Prinzip immer berechenbar

### ■ Negativ

- Keine Patientenperspektive
- Geringe regulatorische Akzeptanz
  - FDA Guideline „Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims:  
“Distribution-based methods for determining clinical significance of particular score changes should be considered as supportive and are not appropriate as the sole basis for determining a responder definition”

# Herleitung/Validierung von MIDs

## | Ankerbasierte Verfahren

- Kriterien für ankerbasierte Verfahren:
  - Studiendesign: PRO und Anker werden zeitlich sinnvoll erhoben. Mindestens 2 Erhebungen (eine zu Beginn und eine zum Ende der Behandlung). Längsschnitt
  - Ankerwahl: Es ist mindestens ein Anker notwendig, Idealerweise ein PRO-Instrument, das die Bewertung des Zustands oder die Veränderung des Zustands durch den Patienten abbildet
    - Patient Global Impression of Severity (PGIS)
    - Patient Global Impression of Change (PGIC)
  - (Ausreichende) Korrelation zwischen Anker und Veränderung des PROs
  - Präspezifikation

# Herleitung/Validierung von MIDs

## | Ankerbasierte Verfahren

### ■ Positiv

- Berücksichtigung der Patientenperspektive
- Regulatorische Akzeptanz
  - FDA Guideline „Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims:  
“The empiric evidence for any responder definition is derived using anchor-based methods”.

### ■ Negativ

- Qualität der Validierungsstudien zu MIDs variiert (z.B. EQ-5D VAS)
- Übertragbarkeit validierter MIDs auf andere Indikationen/Populationen
- Potentiell unterschiedliche MIDs für ein Instrument



| Guyatt et.al.

---

## Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study

Tahira Devji,<sup>1</sup> Alonso Carrasco-Labra,<sup>1</sup> Anila Qasim,<sup>1</sup> Mark Phillips,<sup>1</sup> Bradley C Johnston,<sup>1,2</sup> Niveditha Devasenapathy,<sup>3</sup> Dena Zeraatkar,<sup>1</sup> Meha Bhatt,<sup>1</sup> Xuejing Jin,<sup>4</sup> Romina Brignardello-Petersen,<sup>1</sup> Olivia Urquhart,<sup>5</sup> Farid Foroutan,<sup>1</sup> Stefan Schandelmaier,<sup>1</sup> Hector Pardo-Hernandez,<sup>6,7</sup> Robin WM Vernooij,<sup>8</sup> Hsiaomin Huang,<sup>9</sup> Yamna Rizwan,<sup>10</sup> Reed Siemieniuk,<sup>1</sup> Lyubov Lytvyn,<sup>1</sup> Donald L Patrick,<sup>11</sup> Shanil Ebrahim,<sup>1</sup> Toshi Furukawa,<sup>12</sup> Gihad Nesrallah,<sup>13,14,15</sup> Holger J Schünemann,<sup>1,16</sup> Mohit Bhandari,<sup>1,17</sup> Lehana Thabane,<sup>1</sup> Gordon H Guyatt<sup>1,16</sup>

## CREDIBILITY ASSESSMENT – CORE CRITERIA

**Is the patient or necessary proxy responding directly to both the PROM and the anchor?**

Yes       No       Impossible to tell

**Is the anchor easily understandable and relevant for patients or necessary proxy?**

Definitely yes     To a great extent     Not so much     Definitely no     Impossible to tell

**Has the anchor shown good correlation with the PROM?**

Definitely yes     To a great extent     Not so much     Definitely no     Impossible to tell

≥ 0.7                    ≥ 0.5 to <0.7                    ≥0.3 to <0.5                    <0.3

*Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R, Urquhart O, Foroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang H, Rizwan Y, Siemieniuk R, Lytvyn L, Patrick DL, Ebrahim S, Furukawa T, Nesrallah G, Schünemann HJ, Bhandari M, Thabane L, Guyatt GH. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*. 2020 Jun 4;369:m1714. doi: 10.1136/bmj.m1714. PMID: 32499297; PMCID: PMC7270853.*

## CREDIBILITY ASSESSMENT – CORE CRITERIA

### Is the MID precise?

- Definitely yes     To a great extent     Not so much     Definitely no     Impossible to tell
- ≤ 10% **or** ≥ 200 patients    11-25% **or** 150-199 patients    26-49% **or** 100-149 patients    ≥50% **or** <100 patients

### Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?

- Definitely yes     To a great extent     Not so much     Definitely no     Impossible to tell

*Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R, Urquhart O, Foroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang H, Rizwan Y, Siemieniuk R, Lytvyn L, Patrick DL, Ebrahim S, Furukawa T, Nesrallah G, Schünemann HJ, Bhandari M, Thabane L, Guyatt GH. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*. 2020 Jun 4;369:m1714. doi: 10.1136/bmj.m1714. PMID: 32499297; PMCID: PMC7270853.*

# Credibility Tool der McMaster University, Canada

## Einschätzung

### ▪ **Korrelation $\geq 0.5$ zwischen Anker und PRO gefordert**

- Wird von den meisten derzeitigen Validierungspublikationen nicht erfüllt.
- In anderen Übersichten (wie z.B. Coon 2018) wird eine Korrelation von  $\geq 0.3$  als ausreichend angesehen.

### ▪ **Rapid Response von Forschungsgruppe von Prof. Cella**

- “We argue that strict application of some of these proposed criteria **may be deleterious** to PRO research.”
- “The first criterion ,..., **excludes many useful, clinically-relevant anchors** that aid interpretation.
- “... we suggest dropping the threshold **from 0.5 to something approaching 0.3**, as we and others have suggested (2-4).”
- “... a **range of important difference (ID) values is more useful** than a single, precise MID (5,6 ).”
- “In conclusion, ..., we suggest that the **criteria should be subject to further consideration** by the scientific community. Until this happens, **we advise caution when applying these criteria to anchor selection.**”

# Responseschwelle von $\geq 15\%$ der Skalenspannweite

## Methodenpapier 6.0

### ■ Keine MID

- hinreichend sicher spürbare relevante Änderung (Responseschwelle)
- Responsekriterium von **mindestens** 15 % der Skalenspannweite
  - Responderanalysen ohne weitere Prüfung des Responsekriteriums werden für die Bewertung herangezogen
- **Post hoc** spezifizierte Analysen mit einem Responsekriterium von **genau** 15 % der Skalenspannweite werden berücksichtigt
- **Präspezifizierte** Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite werden regelhaft **nicht** herangezogen

# Responseschwelle von $\geq 15\%$ der Skalenspannweite

## Methodenpapier 6.0

### ■ Positiv

- Eine allgemeingültige Responseschwelle
- Planungssicherheit
- Post hoc Analysen möglich

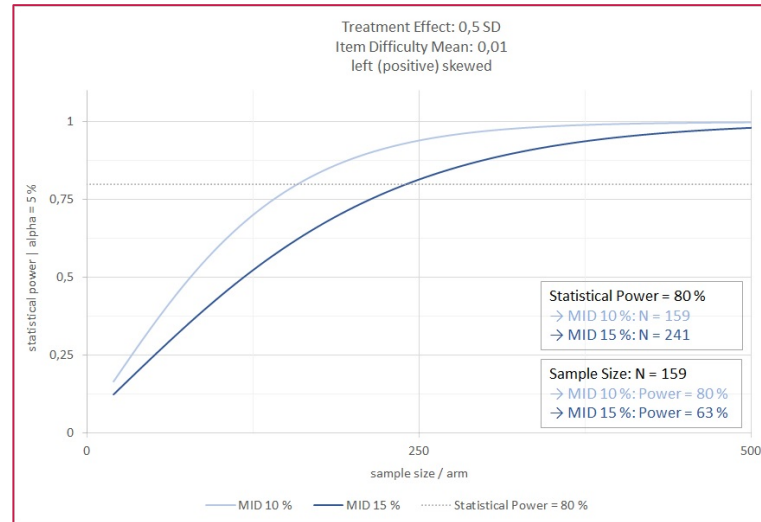
### ■ Negativ

- 15% Responseschwelle ist für eine Vielzahl von PROs größer als die bisher akzeptierte MID
  - EORTC QLQ-C30, SF36, HAQ-DI
- Vergleichbarkeit zu früheren Bewertungen nicht immer gegeben
- Definition der Skalenspannweite nicht immer eindeutig
  - Beispiel: Skalenspannweite für die Physical Component Summary (PCS) und die Mental Component Summary (MCS) des SF-36
- Risiko bei bestimmten Datenkonstellationen einen möglichen Schaden nicht zu entdecken (Powerverlust)

# Responseschwelle von $\geq 15\%$ der Skalenspannweite

## Methodenpapier 6.0

- Unter bestimmten Voraussetzungen führt eine Erhöhung der Responseschwelle von 10 auf 15 Punkte (Skala von 100) zu einem Powerverlust; insbesondere bei schiefen Baseline-Verteilungen



Vortrag MID – One Size Fits All?  
GMDS-ATF Workshop, 5. November 2020

**Und nun?**



# Ausblick

## | Weitere Entwicklungen

- **Berücksichtigung der Patientenperspektive**
- **Etablierte Responseschwellen (MIDs) sollten weiterhin akzeptiert werden**
  - Präspezifikation erforderlich
  - Konsistenz mit der Zulassung
- **Entwicklung eines Kriterienkatalogs zur Beurteilung der Validierung von MIDs**
  - Kriterien auf der Basis des Credibility Tools sollten im gemeinsamen Dialog von Industrie, Akademia, Behörden (IQWiG, BfArM, G-BA) weiterentwickelt werden
- **Weitere Untersuchung der Eigenschaften der 15% Responseschwelle**
  - Mögliche Option für Instrumente ohne vorhandene/ausreichende Validierung

# Referenzen



- Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores; Cheryl D. Coon, Karon F. Cook; *Qual Life Res* (2018) 27:33–40
- Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R, Urquhart O, Foroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang H, Rizwan Y, Siemieniuk R, Lytvyn L, Patrick DL, Ebrahim S, Furukawa T, Nesrallah G, Schünemann HJ, Bhandari M, Thabane L, Guyatt GH. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*. 2020 Jun 4;369:m1714. doi: 10.1136/bmj.m1714. PMID: 32499297; PMCID: PMC7270853
- Rapid Response: John D Peipert, Assistant Professor David Cella, Northwestern University Feinberg School of Medicine, 06 September 2020 (<https://www.bmj.com/content/369/bmj.m1714/rapid-responses>)
- Vortrag Doris Böhm, MID – One Size Fits All? GMDS-ATF Workshop, 5. November 2020
- Link zur R Shiny App zur Simulation von MID Szenarien: <https://htaor.shinyapps.io/midapp/>

# Questions?

# Thanks!