ISOQOL 2026 - ORAL PRESENTATION

A Review of Vignette-Based Methods for Estimating Meaningful Score Differences (MSDs)

Kacylia Pistoia SeeingTheta

"Meaningful score difference" (MSD)

how much a COA score must change for the change to matter to a patient.
We will use "MSD" as an umbrella term here.

"Vignette"

a hypothetical patient described before and after a change; respondents judge whether the change in a pair is meaningful.

What is the usual route, and is it consistent?

ANCHOR-BASED

Uses a patients own global assessment of change to infer meaningfulness.

DISTRIBUTION-BASED

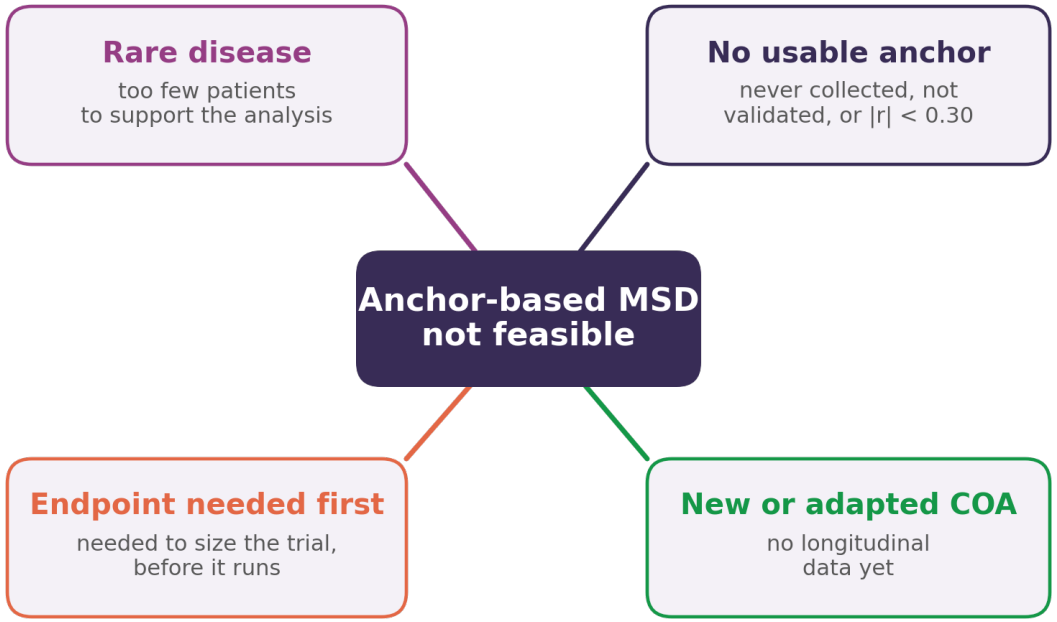
Infer lower bound of meaningful range from the statistical spread of scores in the sample.



Terwee (2021): estimates spanned **0.1 to 12.7** T-score points on a shared metric (31 studies), almost as wide within a single measure.

So, the number depends on how it’s derived.

When the usual route is closed – vignettes as a supplement



FDA · PFDD GUIDANCE 4 (2023)

*"Often, MSD is determined based on what patients would regard as a clinically meaningful within-patient change, but other approaches might also be appropriate (e.g., **those based on the patient's perception of the differences between hypothetical vignettes representing different degrees of symptom severity or functioning**)."*

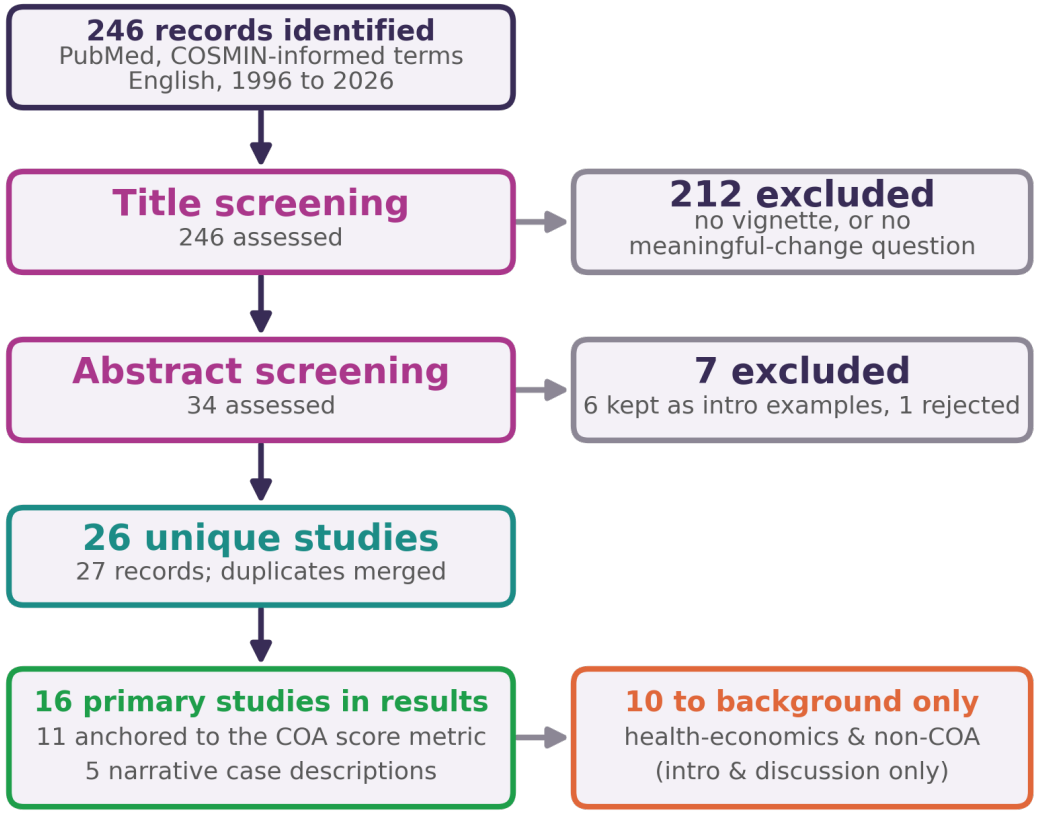
TERWEE (2021)

*"We recommend to use the predictive modeling method, possibly **supplemented with the vignette-based method**, in future MIC studies."*

Vignettes are offered as *"other approaches"* (FDA, 2023) to *"supplement"* anchor-based methods (Terwee, 2021).

Does a methodology for this exist?

A targeted review to see if a standard method exists



PubMed, 1996-2026, using COSMIN-informed terms across three concept blocks.

Full-text screening kept 16 studies in the results (vignettes on a COA score)

Single database (PubMed) - a fuller review is still needed.

The method groups

The 16 COA studies cluster into four method groups

Bookmarking & standard setting

A panel places bookmarks at score-category boundaries (usually IRT-derived).

Qualitative vignette interviews

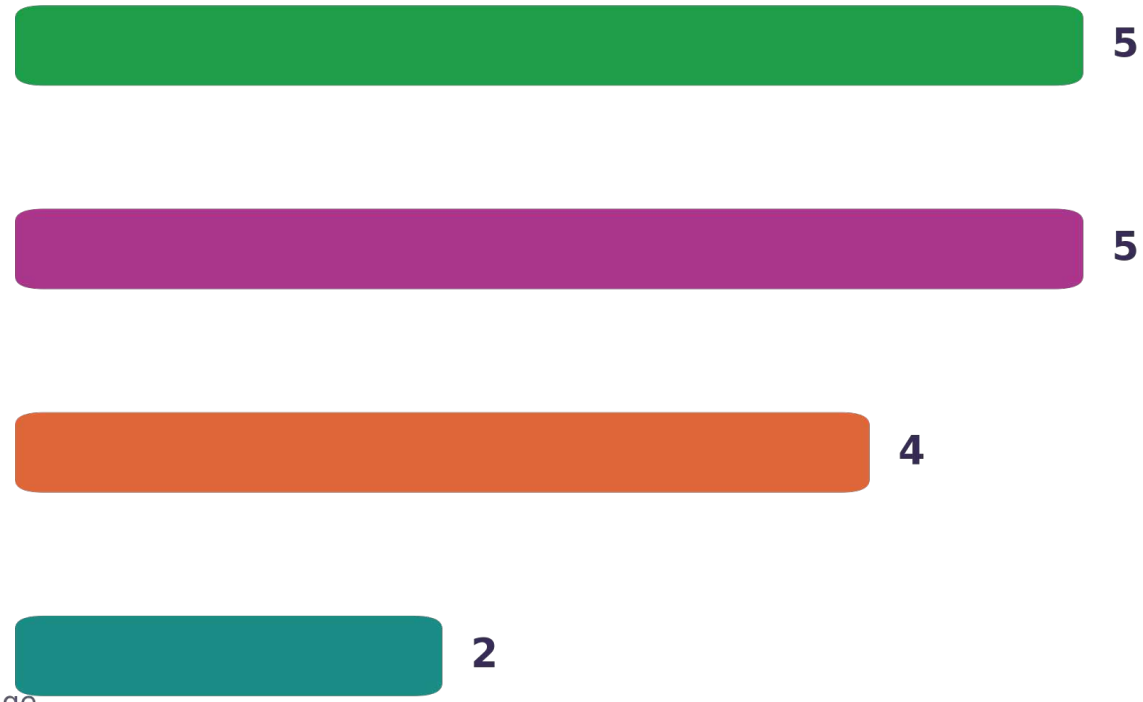
Narrative case descriptions, rated and discussed.

Delphi & structured consensus

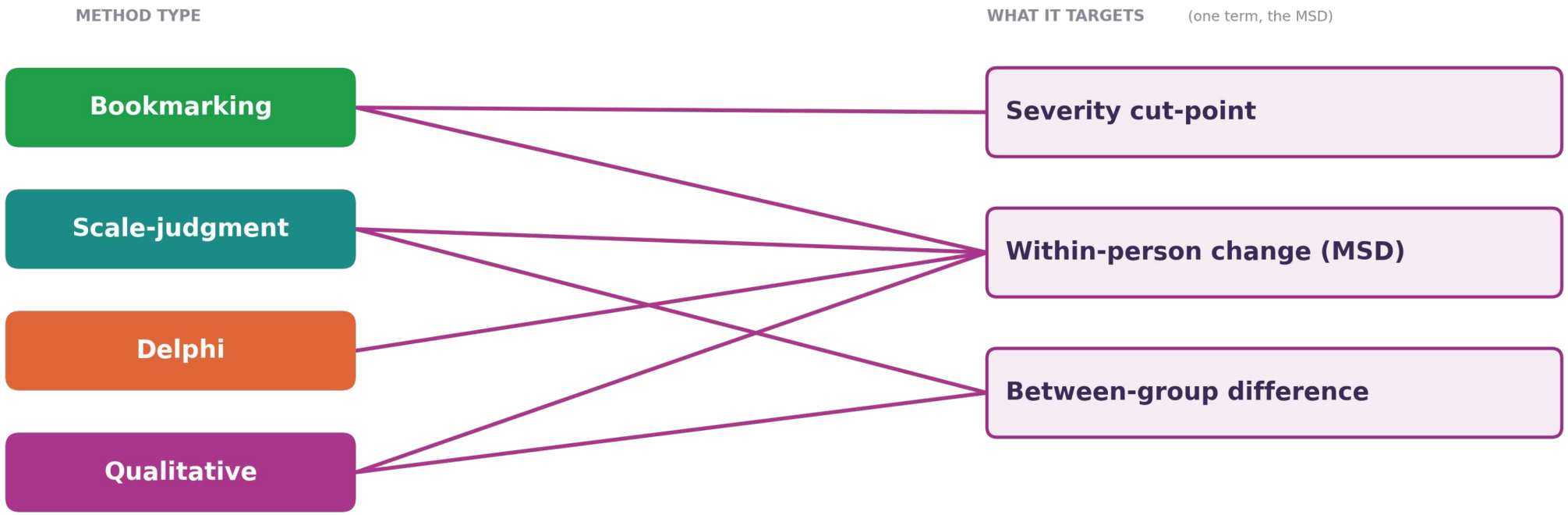
Iterative rating rounds with a pre-set agreement rule.

Scale-judgment

Judges compare the same hypothetical patient at two times and rate the change.

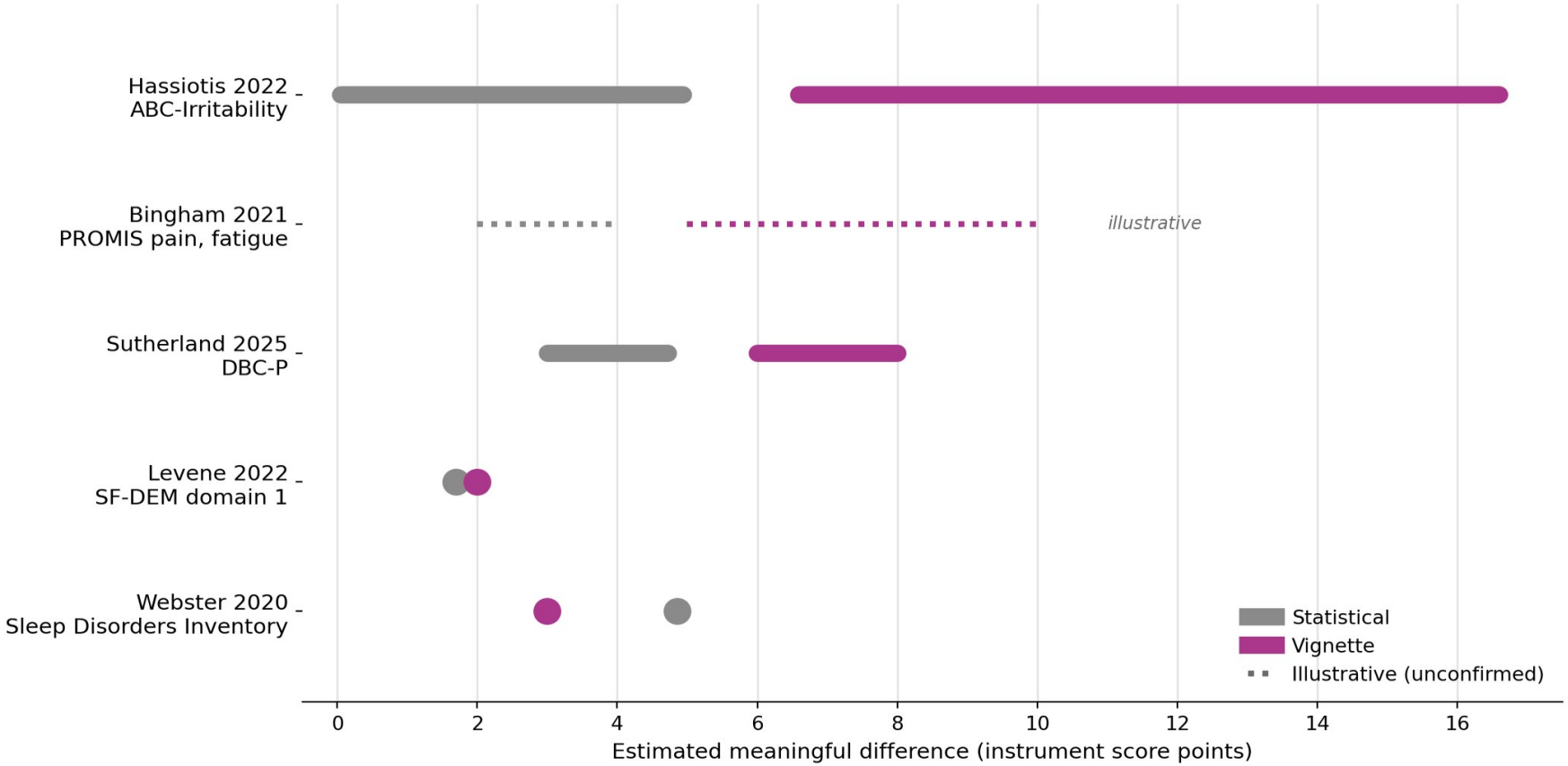


Finding 1: one term, different targets



"Meaningful score difference (MSD)" is used for three different targets across the 16 studies.
One method can target more than one.
So the first question is "which target?", not "how many points?"

Finding 2: vignette and quantitative anchor-based estimates diverge



Often higher, sometimes aligned, occasionally lower; no study explains why.
Both can't be compared; they use different study designs.
Vignette methods are often used in situations where anchor-based approaches were not feasible.
The field needs a vignette vs quantitative anchor-based estimates validation study.

Gaps - 4 themes in a young field

1

No shared terminology

of what counts as a vignette - an IRT score profile and a clinical case description carry the same name.

2

The target is often unclear

“meaningful difference” - but in what?

3

No quality standard

specific to these methods - COSMIN does not grade interpretability as a property.

4

Still early days

a new approach, limited use so far - 16 studies, mostly PROMIS rheumatology and rare diseases; the field is still learning.

What can we borrow from health economics?

	Health economics	COA vignette practice today
Constructing the states BORROW	Standard structure, tested for hidden severity cues	Written from scratch, no conventions
Elicitation task BORROW	Validated tasks with known limits	Ad hoc rating and ranking
Scale of administration LEAVE	Large samples, often 1,000+ online	Qualitative survey, ~100 patients max
Quality checks BORROW	Tests for inconsistency and order effects	Design effects untested
What you get LEAVE	A threshold with a confidence interval	A point value, no error rate

TAKE: the machinery: construction conventions · validated elicitation tasks · consistency checks

KEEP: our constraints value a change within a person · higher regulatory bar · patients who live with the condition (small samples, ~100 even in a quantitative study)

Where the field might go

A few directions for the field to adopt together.

01

Agree shared terminology

for vignette methods and their targets.

02

State the target up front

say what “meaningful” means before estimating it.

03

Borrow the health-economics toolkit

for constructing and valuing hypothetical states.

04

A direct comparison

vignette vs quantitative anchor-based estimates on the same data would help the field adopt these methods with confidence.

Thank you.

CONTACT

Kacylia Pistoia
SeeingTheta
kacylia@seeingtheta.com



Come and meet Iris and the team at the SeeingTheta booth.

ACKNOWLEDGEMENTS

With thanks to our co-authors and collaborators, including Aaron Yarlas, Steve Kymes and Montserrat Vera Llonch (Ionis).

THE SEEINGTHETA TEAM



Piper FROMY
Psychometrics, Methodology and Strategic Consulting



Elisabeth BERGER
Statistics and Programming



Kacylia PISTOIA
Psychometrics and Methodology



Maryam BAYRAM
Psychometrics and Methodology



Nikola MARKIEWICZ
Psychometrics and Methodology



Marie-Jeanne HOLVECK
Statistics and Programming