Interpretation Clinical significance: what does it mean?

Patrick Marquis, MD, MBA
Mapi Values - Boston

DIA workshop
Assessing Treatment Impact Using PROs:
Challenges in Study Design, Conduct and Analysis
Paris May 10-11, 2004

Context

- Context of the interpretation of clinical trial results assuming:
  - Absence of issue regarding trial design, quality and analysis
  - PRO questionnaire developed and validated according to recognized standards
  - Hit on primary end-point or clinical efficacy end-point

Question

- Statistical significance reached however…
- Null hypothesis may be rejected for very small differences by large sample sizes
- Differences may be of no practical importance

➡ Clinical meaning and practical value?

ClinSig definition

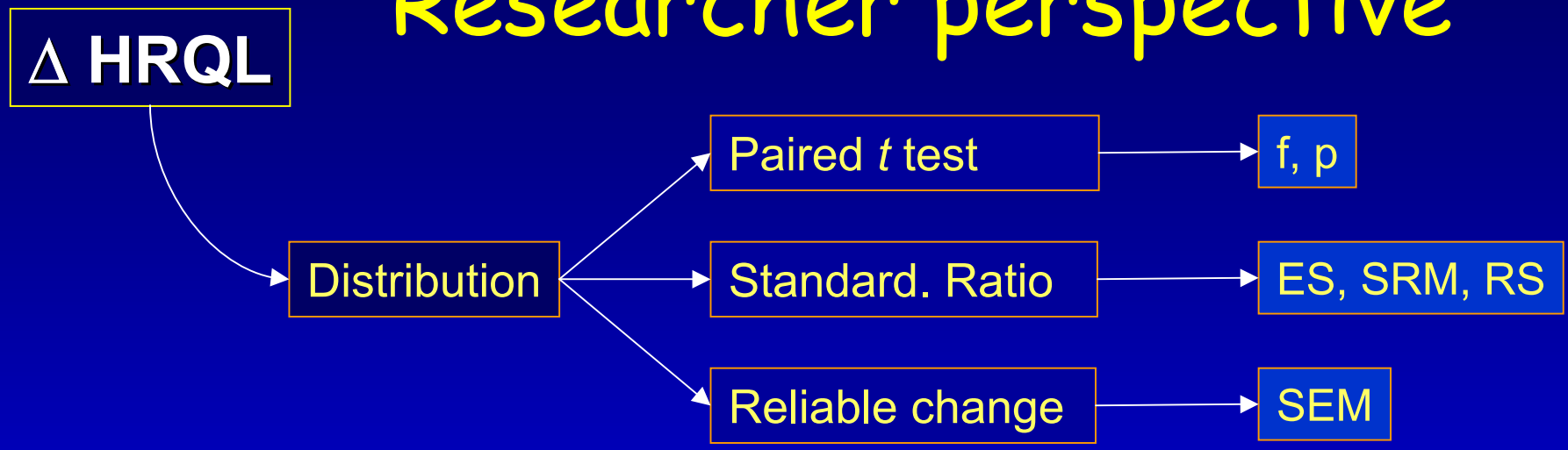
- Patient perspective
  - A difference large enough to have an implication in the patient management
  - A value that patients recognize as minimal important difference (MID)
- Clinical perspective
  - Smallest effect size leading clinicians to recommend a treatment

ClinSig in PRO research = like any other end-points

Specificities

- Concepts vary across questionnaires
- Scale label not always informative
- Number of items and scaling vary
- Differences in anchors, direction, and magnitude of change

Researcher perspective



Standardized Ratio

- Helpful to understand the size of the change based on a standardized unit
- Commonly used in clinical research
- Easy to interpret
- Interpretation guidelines (Cohen)

  .2 to .49 small / .5 to .79 moderate / .8+ large

  For sample size estimates

Role of effect size

- Relative efficiency of measures to detect treatment effect for leflunomide versus placebo in RA (Tugwell, 2000)

	Mean	ES
• Tender joint count	- 4.69	- 0.59
• Swollen joint count	- 2.78	- 0.44
• Pain intensity	- 17.51	- 0.65
• ESR	- 8.82	- 0.41
• CRP	- 1.09	- 0.47
• HAQ disability index	- 0.42	- 0.80
• SF-36 MCS	- 0.65	- 0.06
• SF-36 PCS	- 8.52	- 0.67
• SF-36 bodily pain	- 15.76	- 0.73

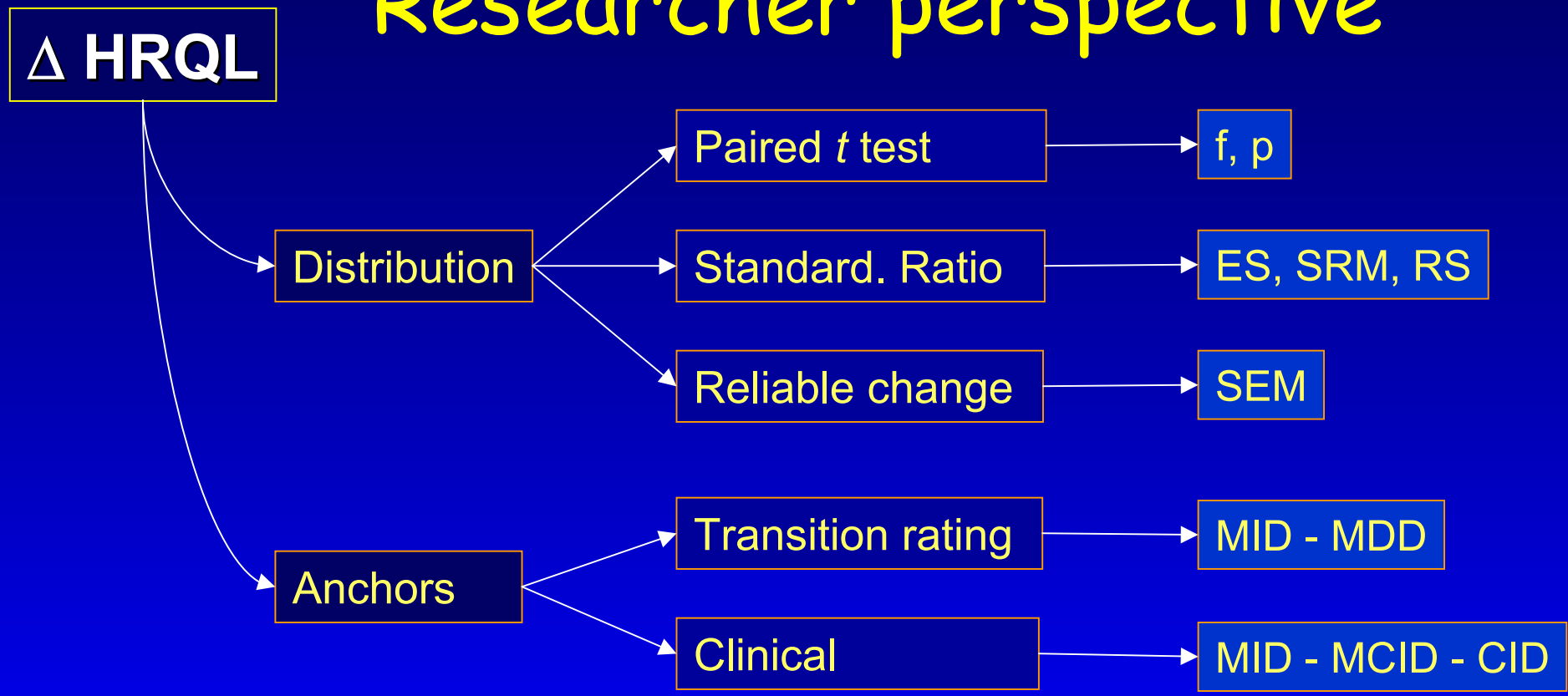
Standard Deviation as a key element

- Increased variance due to confounding factors such as gender and age
  - ⇨ ES should integrate these confounders
- In the SEM, the higher the reliability:
  - ⇨ The lower the true effect
  - ⇨ The lower the corresponding ES

SD= 20

r=.70	⇨ SEM= 11	⇨ ES=.55
r=.80	⇨ SEM= 8.9	⇨ ES=.45
r=.90	⇨ SEM= 6.3	⇨ ES=.32

Researcher perspective



Transition rating

- Within-patient transition rating (magnitude of their global change)
- Between-patient difference rating (magnitude of perceived differences)

Within-patient transition rating

- Patient global rating of change after treatment (Juniper 1996)

7 A very great deal change
6 A great deal change
5 A good deal change
4 Moderately change
3 Somewhat change
2 A little change
1 Almost the same
0 No change

Mean HRQL change of patients considered as the MID = 0.5 for a 1 to 7 scale

ADVAIR FDA Claim

"The subjective impact of asthma on patients' perception of health was evaluated through use of a validated instrument called the Asthma Quality of Life Questionnaire (AQLQ). Patients receiving ADVAIR DISKUS 100/50 had clinically meaningful improvements in overall asthma-specific quality of life as defined by a difference between groups of >= 0.5 points in change from baseline AQLQ scores (difference in AQLQ score of 1.25 compared to placebo)."

Transition rating as the anchor?

- Easy to implement and intuitive
- Consistency with other methods
- Used for most questionnaires

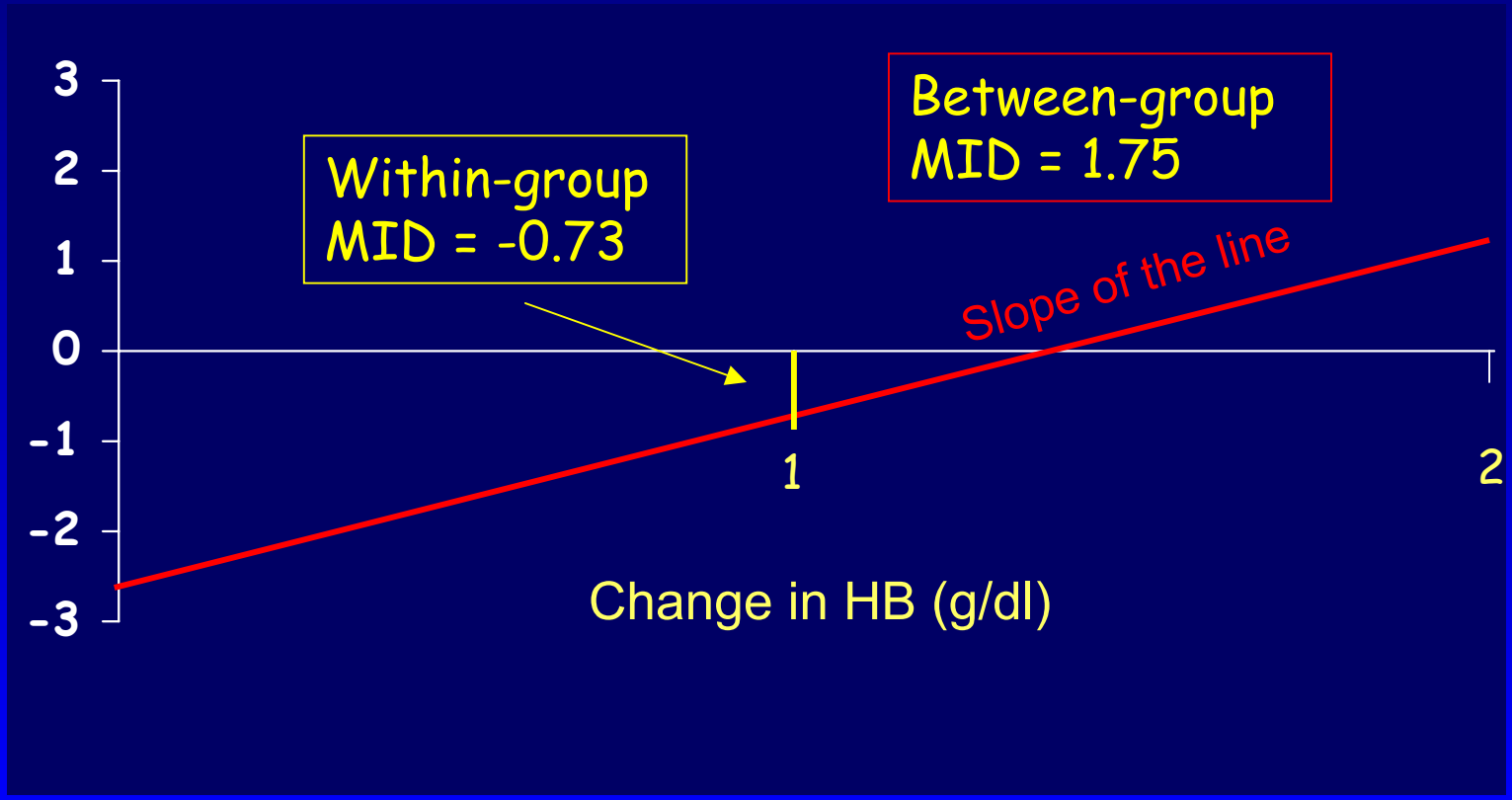
Transition rating as one anchor

- Issues
  - Single measure
  - Pattern of correlation pre-post test not usually as expected
  - Arbitrary decisions and sensitivity to wording
  - Absence of clinical anchor / interpretation
  - Mainly developed for within-patient changes but used for between-patient decision

Clinical anchor

- Meaningful clinically and for patients
- "Some" level of correlation with the clinical anchor
- Threshold or calibration needed for the clinical anchor

Clinical anchor



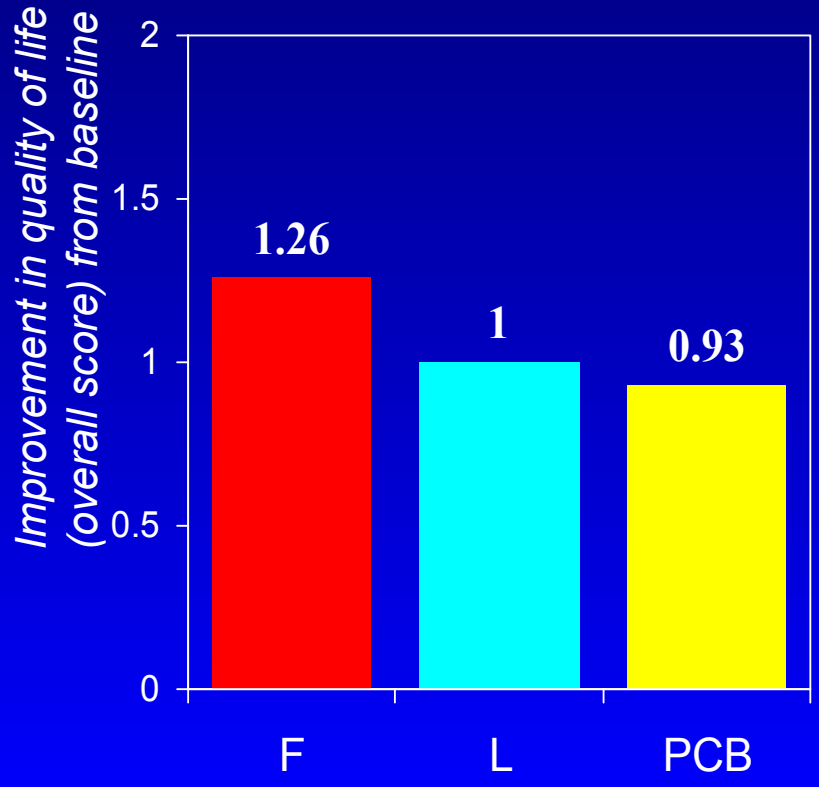
Patrick D et Al. Eur J Cancer 39(2003) 335-345

Application of MID - MDD - CID

- Benchmark for interpreting change
  - Not a hard threshold
- Patient Benefit from treatment
  - Responder analysis
- NNT Number Needed to treat

- Interpretation of change over different time points
- Sample size calculation

Interpretation issue in QoL



Control of SAR symptoms

Improvement of HRQL

- F vs PCB p= .005
- F vs L p= .03

(Clin Exp Allergy 2000; 30: 891-899)

Percentage reaching the MID

- F 74%
- L 64%
- PCB 61%

NNT results

	F		
L	Improved (0.74)	Unchanged (0.23)	Deteriorated (0.03)
Improved (0.64)	0.4691	0.1494	0.0174
Unchanged (0.26)	0.1919	0.0611	0.0071
Deteriorated (0.10)	0.0768	0.0244	0.0028

NNT = 1/(Net Benefit)

Net Benefit = 0.2931 – 0.1739 = 0.1192

NNT = 8.4 for F vs L

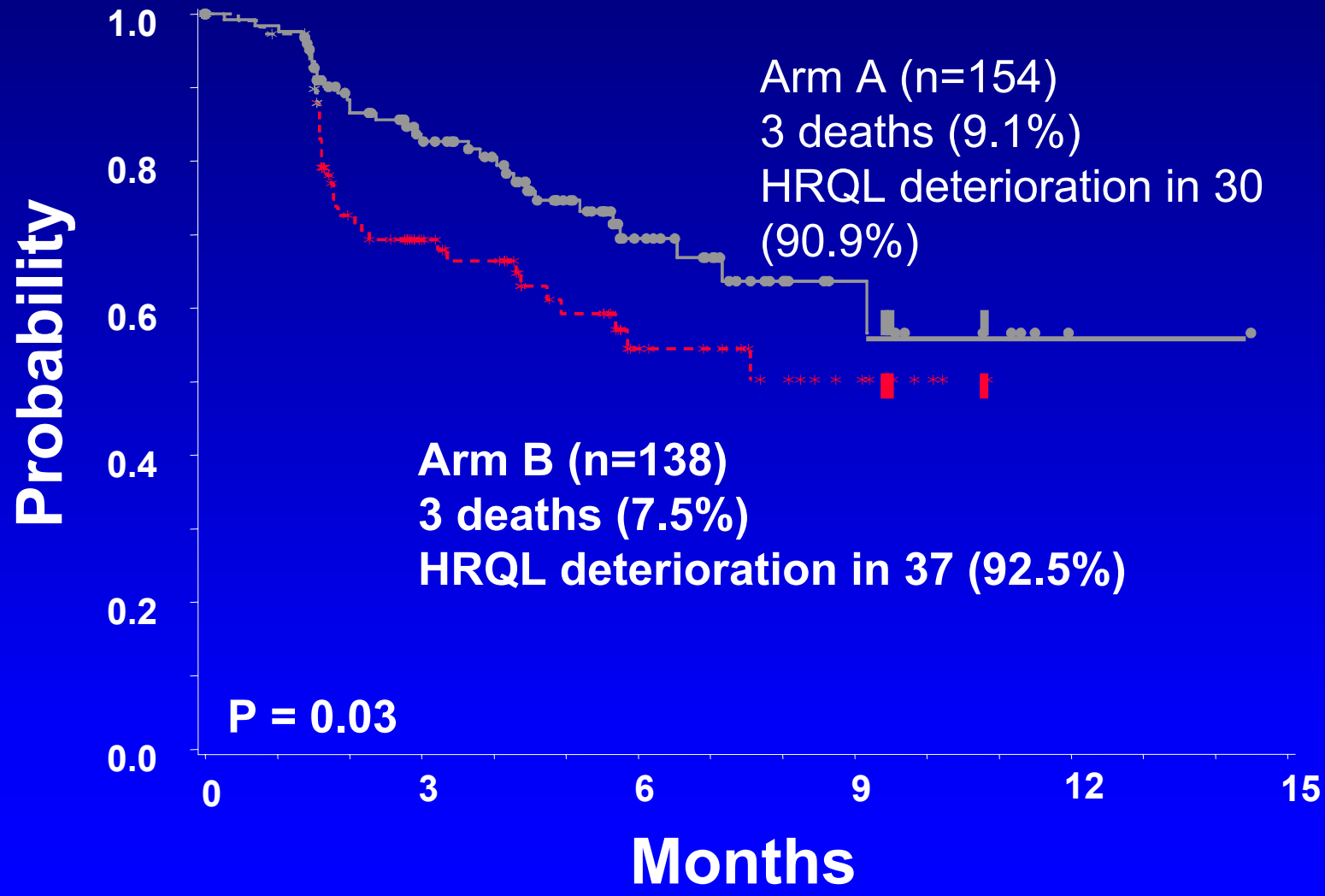
NNT = 7.2 for F vs PCB

NNT = 57.2 for L vs PCB

Interpretation of the NNT

- About 7 patients should be treated with F to get an additional meaningful HRQL benefit over placebo
- About 8 patients should be treated with F to get an additional meaningful HRQL benefit over L
- About 57 patients should be treated with L to get an additional meaningful HRQL benefit over placebo

Integration of multiple time points Time to definitive 5-point HRQL deterioration

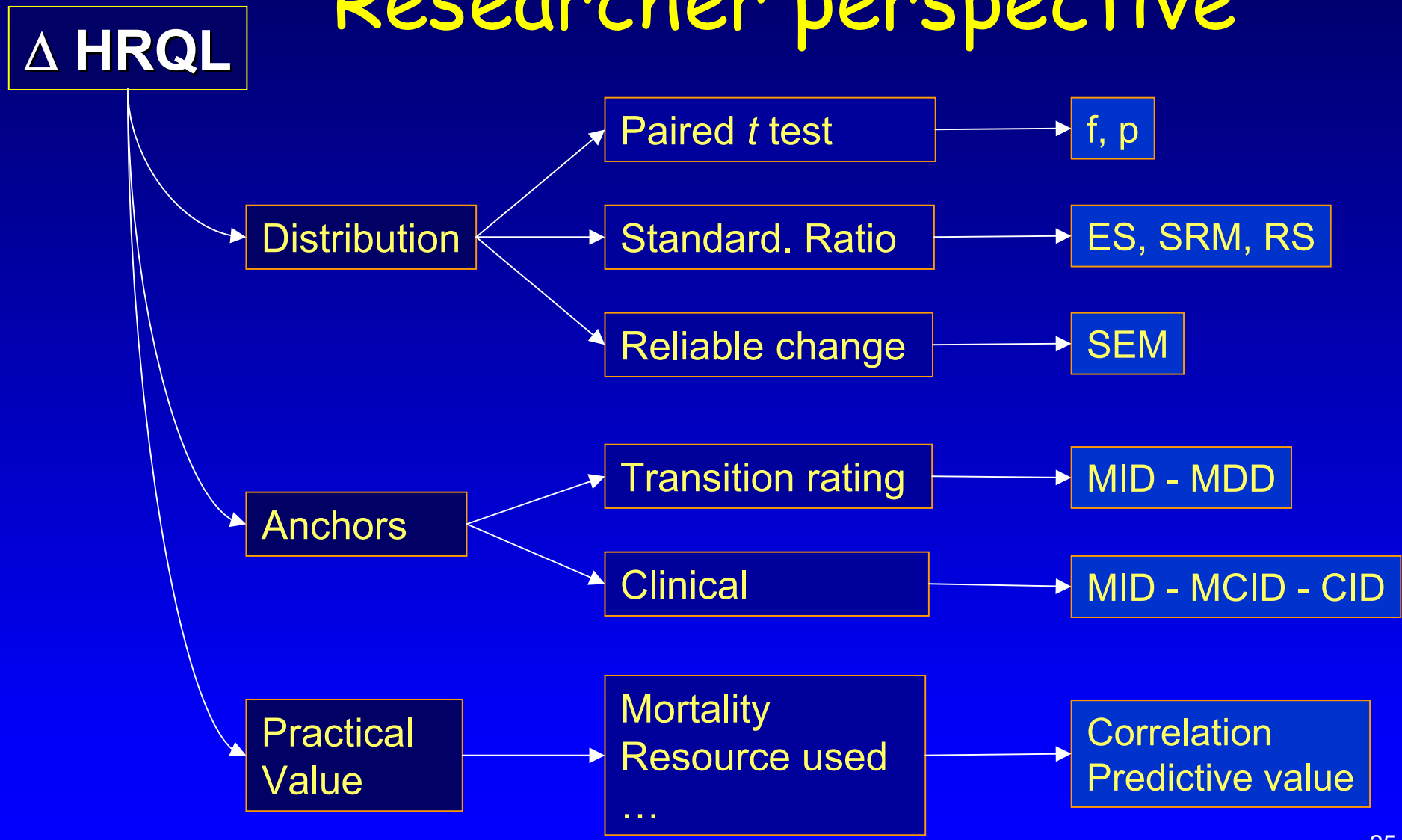


Universality of half-SD?

- Consistency of MID determined by anchor-based methods with ES?
- 38 studies reviewed
- For 32 studies MID closed to half-SD (.495 ±.155)
  - No influence of the number of response options and type of questionnaires
- Larger ES for negative changes & acute conditions
- Smaller ES for population-based estimation

Norman G, Med Care 2003 41,5, 582-592

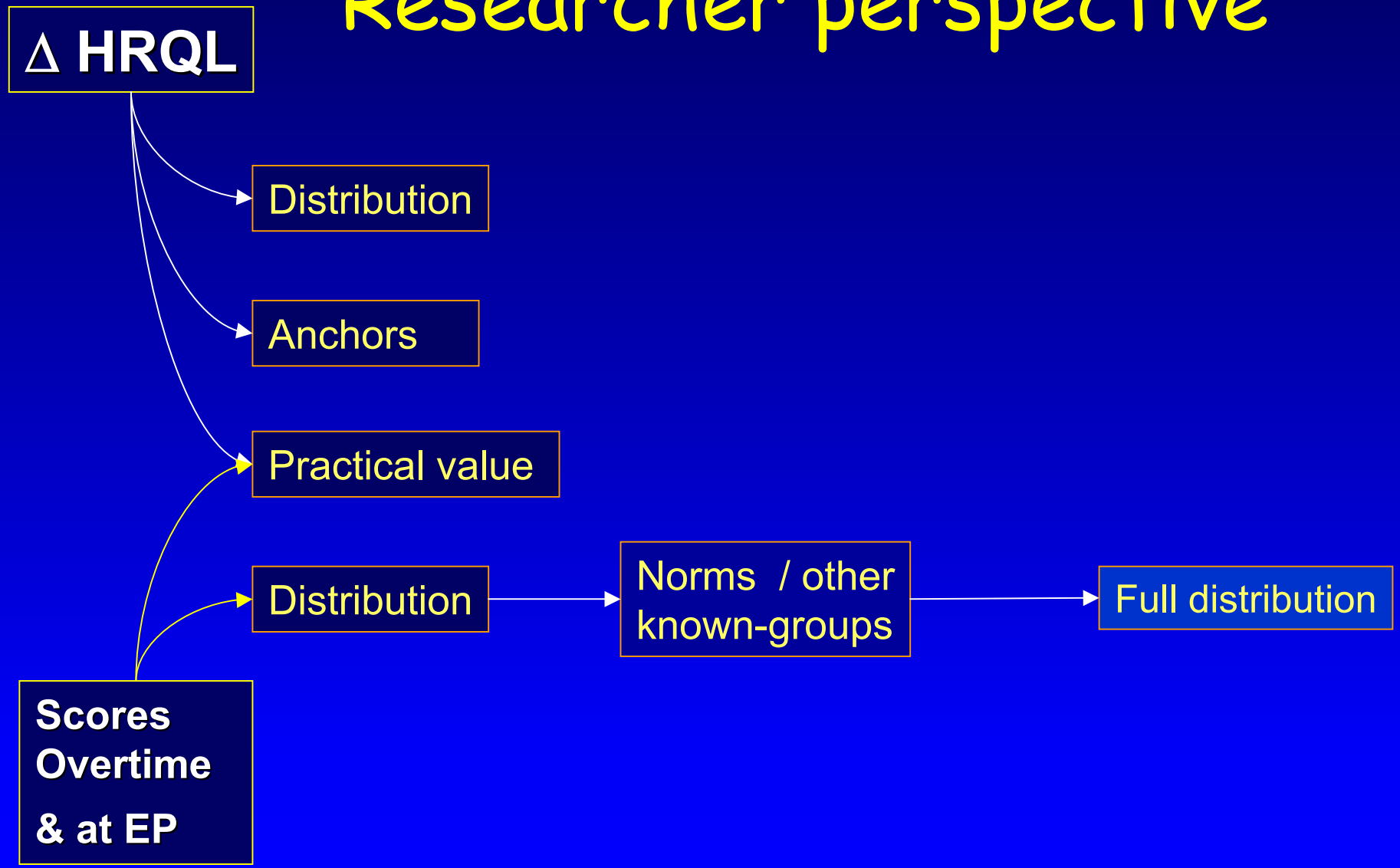
Researcher perspective



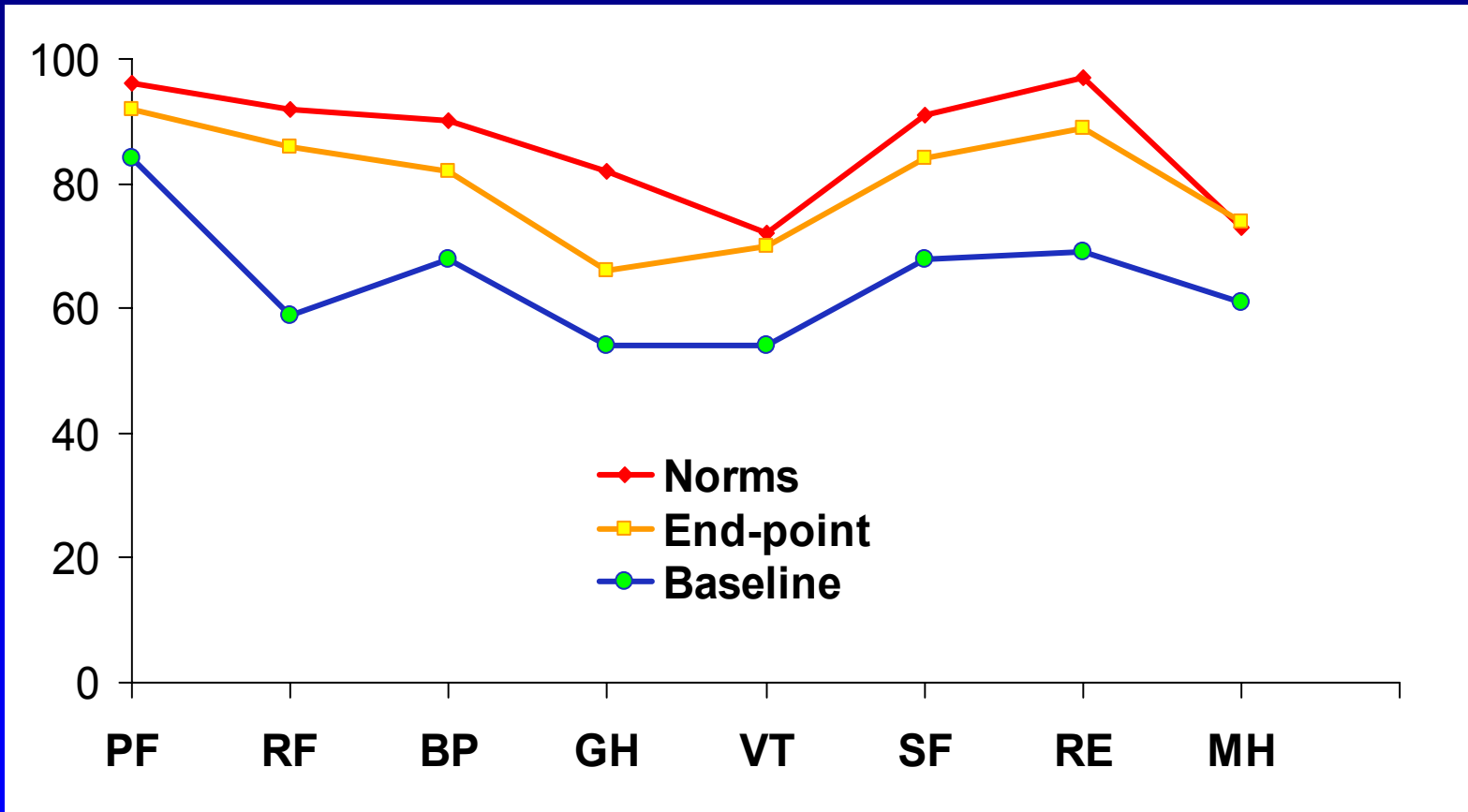
Practical Value / Added value

- Prediction of mortality, resource utilization
  - See first session
- Strong message
- Final confirmation of questionnaire validity
- Limitation to define the "minimum" different concept
- Used for establishing the robustness of the questionnaire
- Not commonly used for RCT interpretation

Researcher perspective



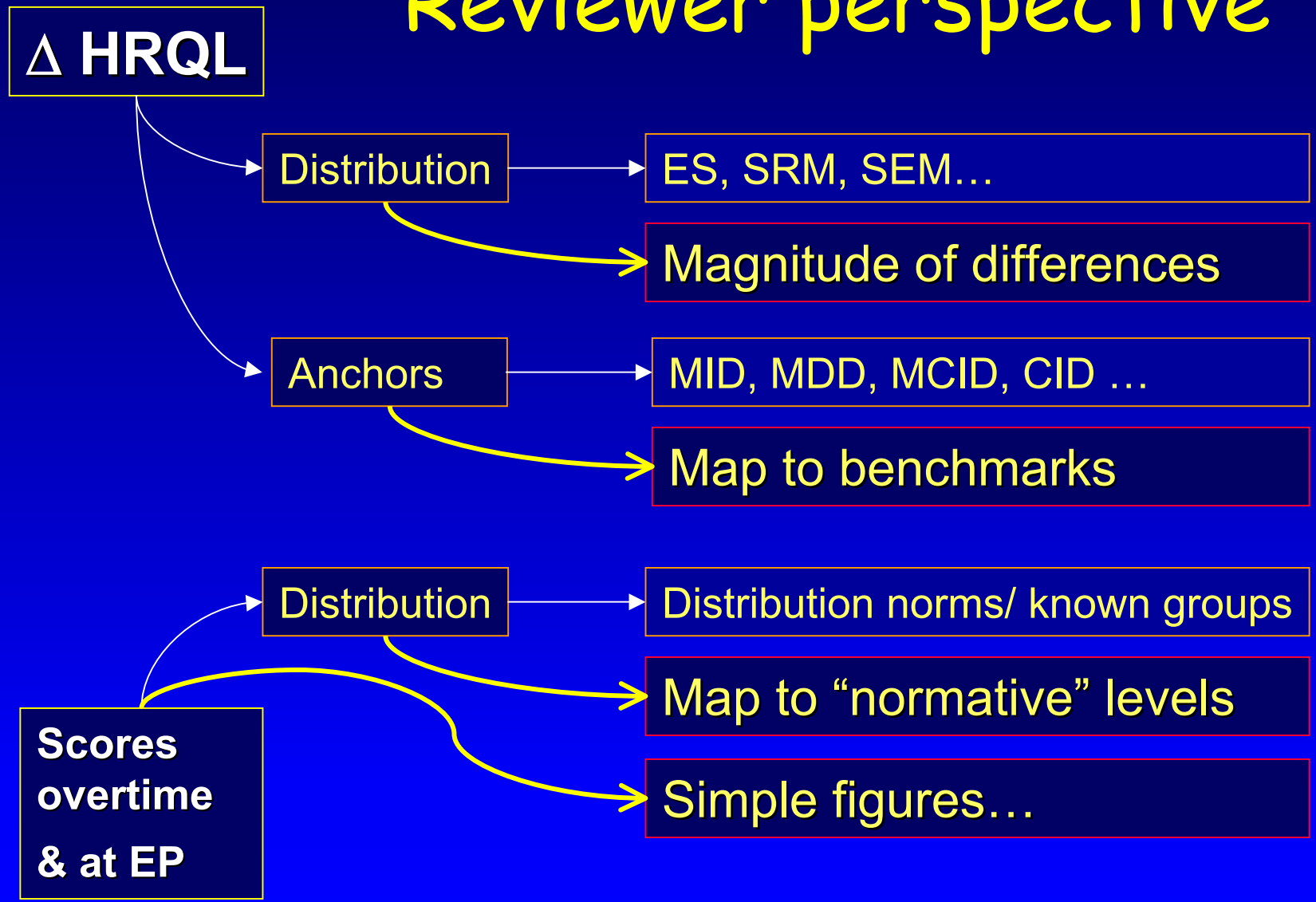
Treatment effect in PAR compared to normative data



Decision-making perspective

- Drug approval
- Label development
  - Not the indication but useful additional information for prescribers that should be mentioned in the SPC
- Promotional material approval
  - Evidence supporting the promotional claim

Reviewer perspective



Recommendations

- Establish links between clinical variable, PRO data and clinical significance
- Consistency with other results key factor
- Integrate clinical context as will always be different
  - Cancer
  - Asthma
  - Anemia
  - Irritable Bowel Disease

Conclusion

- No universally accepted approaches to determine the ClinSig
- No single approach is perfect
- Half-SD seems "safe" but
  - More research needed
  - Smaller ES can be meaningful +++
    see Cohen, Miller, and SEM
  - Comparison of active treatments
  - Increased variance due to confounders

Facilitate the task of reviewers!

- Provide supportive evidence:
  - Documentation of the development
    - ⇨ "face value"
  - Documentation of the validation with executive summary
    - ⇨ "proof of trust"
  - Interpretation guidelines (different approaches)
    - ⇨ "meaning"
- Standardization & quality of PRO submission…

Try to be … a reviewer just for one day

Points of Discussion

- Is the clinical significance only a matter of method and technique?
- Should it be universal (mathematical model) and or specific to the clinical context?
- Is the search for thresholds mandatory?
- How to choose the right anchor for determining the MID?
- Are post-doc analyses acceptable to further analyze the clinical significant?