Is “15% of the scale range” Universally Applicable to Define “MID” and Clinical Relevance of Patient-Reported Treatment Benefits?

Group Discussion with the ISPOR Clinical Outcome Assessment Special Interest Group (COA SIG)

SPEAKERS
Olivier Chassany, MD, PhD
Greater Paris University Hospitals (AP-HP), Paris, France
Katja Rudell, PhD, MSc
Parexel International, Uxbridge, London UK
James Shaw, PhD, PharmD, MPH
Bristol-Myers Squibb, Lawrenceville, NJ, USA
Academic perspective

Prof. Olivier Chassany
Patient-Report Outcomes Research (UMR 1123), Université de Paris, Inserm
Health Economics Clinical Trial Unit (URC-ECO), Hospital Hotel-Dieu, AP-HP, Paris
Declaration of recent interests

• Sanofi, Pfizer, Gilead, BMS, IQVIA, Boehringer
Example of a dossier submitted to the Federal Joint Committee (G-BA), using the usual MID (i.e. 5 points for PCS and MCS of the SF-36)

Resolution of the Federal Joint Committee (G-BA) on an Amendment of the Pharmaceuticals Directive (AM-RL): Annex XII – Benefit Assessment of Medicinal Products with New Active Ingredients According to Section 35a SGB V Secukinumab (Reassessment on the Basis of New Scientific Findings (Psoriatic Arthritis)) - 18 February 2021
German HTA agency IQWiG (Institute for Quality and Efficiency in Health Care) updated their general methods paper on health benefit assessment (05 Nov 20)

- Pharmaceutical manufacturers must submit benefit dossiers to facilitate HTA appraisals to G-BA (Federal Joint Committee) in Germany for new pharmaceutical, new combination or new label.

- For all non-orphan drugs the G-BA delegates the benefit assessment to the IQWiG. The IQWiG conducts the dossier evaluation using the criteria set out in the IQWiG general methods paper.

- The entire HTA is based on patient relevant endpoints, which include Clinical Outcomes Assessments and Patient Reported Outcomes (PRO). The final decision on the added benefit is made by the G-BA.

https://www.iqwig.de/methoden/general-methods_version-6-0.pdf?rev=194070
https://www.g-ba.de/english/benefitassessment/
German HTA agency IQWiG (Institute for Quality and Efficiency in Health Care) updated their general methods paper on health benefit assessment (05 Nov 20)

• A within-patient relevance threshold was established to allow IQWiG make quick decisions about benefits of therapies that were using PRO rating scales.
• Based on a systematic literature review that covered 8 therapy areas*, the overall threshold of treatment benefits ranged between +1-38% change
• The new rule is to present results: % of patients who improve their scores by at least 15% of the width of an overall scale (e.g. 15 points on a 0-100 scale).
  • Patients exceeding that threshold are considered to perceive a meaningful change (responder)

* rheumatism, orthopaedics, paediatrics, fatigue, oncology, cardiovascular, COPD, urogenital, musculoskeletal diseases and miscellaneous diseases

https://www.iqwig.de/methoden/general-methods_version-6-0.pdf?rev=194070
https://www.g-ba.de/english/benefitassessment/
IQWiG skepticism regarding minimal important difference (MID) one of the standard threshold for assessing benefit

IQWiG criticizes MIDs:
• validation not scientific sound
• high variability
• Lack of standard to assess quality of MID
• Lack of quality in MID reporting
• MIDs were not accepted anymore by IQWiG, but often by the German Regulatory Agency that assessed the efficacy and safety of the product (G-BA) due to consistency reasons.

IQWiG skepticism regarding minimal important difference (MID) one of the standard threshold for assessing benefit

Discussion (June 2020) with stakeholders:
• IQWiG explained their 15% approach, which is not meant to be a MID but a response threshold, “where we are sufficiently sure that we represent a change that is noticeable for patients”.

This is the nub of matter:
• “noticeable” is exactly what means the MID, a minimal change that patients are able to notice anchored to a patient global impression of change
• The challenge is to decide what level of change is large enough to define a patient as responder

A universally response threshold of 15% of the scale range is appealing but too simple
A unique 15% of change cannot fit all the responder definitions depending on the PRO endpoint used

- Responder definition (i.e., relevant change) depends on the concept captured by the tool
- 15% of scale range rule:
  - May not be large enough for primary symptoms and for endpoints which measure direct impact on daily life activities (e.g., impact of pyrosis improvement on diet)
  - Is likely to be not achievable for more indirect PRO such as Health-Related Quality of life, for which the improvement is smaller, especially, when the improvement on the clinical primary endpoint is already small.
- We cannot expect an improvement larger on HRQL scores than on the direct marker of the evolution of the disease.
Applying the 15% threshold of the scale range

<table>
<thead>
<tr>
<th>Health-related quality of life</th>
<th>Secukinumab</th>
<th>Adalimumab</th>
<th>Secukinumab vs Adalimumab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint</td>
<td>N</td>
<td>Patients with event n (%)</td>
<td>N</td>
</tr>
<tr>
<td>SF-36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improvement by 15 % of the range of the scale^2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Component Score (MCS, improvement of 9.6 points)</td>
<td>110</td>
<td>46.1 (41.9)</td>
<td>101</td>
</tr>
<tr>
<td>Physical Component Score (PCS, improvement of ≥ 9.4 points)</td>
<td>110</td>
<td>42.8 (39.0)</td>
<td>101</td>
</tr>
<tr>
<td>Improvement of ≥ 5 points^2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Component Score (MCS, improvement of 5 points)</td>
<td>110</td>
<td>68.5 (62.3)</td>
<td>101</td>
</tr>
<tr>
<td>Physical Component Score (PCS, improvement of ≥ 5 points)</td>
<td>110</td>
<td>66.9 (60.8)</td>
<td>101</td>
</tr>
</tbody>
</table>

Resolution of the Federal Joint Committee (G-BA) on an Amendment of the Pharmaceuticals Directive (AM-RL): Annex XII – Benefit Assessment of Medicinal Products with New Active Ingredients According to Section 35a SGB V Secukinumab (Reassessment on the Basis of New Scientific Findings (Psoriatic Arthritis)) - 18 February 2021

Diff: 17.3%  
Diff: 23.2%
Applying the 15% of the scale range threshold on a generic heath status scale is counter productive

Very unlikely to show a difference between groups based on a such high responder definition (i.e. 15 mm of EQ-5D 0-100 mm VAS scale) unless the treatment has a dramatic benefit

<table>
<thead>
<tr>
<th>Morbidity</th>
<th>Secukinumab</th>
<th>Adalimumab</th>
<th>Secukinumab vs adalimumab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint</td>
<td>N</td>
<td>Patients with event n (%)</td>
<td>N</td>
</tr>
<tr>
<td>Health status (EQ-5D VAS, improvement of ≥ 15 mm, ± 15 % of the range of the scale)</td>
<td>110</td>
<td>58.8 (53.5)</td>
<td>101</td>
</tr>
</tbody>
</table>

Resolution of the Federal Joint Committee (G-BA) on an Amendment of the Pharmaceuticals Directive (AM-RL): Annex XII – Benefit Assessment of Medicinal Products with New Active Ingredients According to Section 35a SGB V Secukinumab (Reassessment on the Basis of New Scientific Findings (Psoriatic Arthritis)) - 18 February 2021
No scientific reason why the responder definition should be different for drug approval regulators and for HTA assessors

This “universal” 15% of the scale range rule applied for HTA assessment, contradicts the EMA and FDA guidelines which recommend sensitivity analyses using several definitions of responders, and/or e.g. to present results of change as a cumulative distribution function (CDF) across the range scale, which is more informative than a single responder definition.

<table>
<thead>
<tr>
<th>Linaclotide vs placebo</th>
<th>12-wk A</th>
<th>12-wk B</th>
<th>26-wk B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal pain responder (30%)</td>
<td>55 vs 42% Δ = 13%</td>
<td>54 vs 39% Δ = 15%</td>
<td>54 vs 36% Δ = 18%</td>
</tr>
<tr>
<td>IBS relief responder</td>
<td>37 vs 19% Δ = 16%</td>
<td>39 vs 17% Δ = 16%</td>
<td>37 vs 17% Δ = 20%</td>
</tr>
<tr>
<td>Abd pain responder (40%) Abd pain responder (50%)</td>
<td>48 vs 34%</td>
<td>39 vs 26%</td>
<td>← Sensitivity analyses</td>
</tr>
<tr>
<td>Sustained</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pain responder</td>
<td>53 vs 41%</td>
<td>34 vs 18%</td>
<td></td>
</tr>
<tr>
<td>• IBS relief responder</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion - The 15% of the range scale rule cannot be the universal solution for interpreting PRO results

It may be easier to base a decision on a single binary (> 15%) presentation of results, but it reduces considerably the information usually needed to review a dossier with PRO/COA data, being for MA or for HTA

The added value of PROs is also based on the review:
- Evolution of scores over time (if repeated)
- Consistency:
  - Across different PROs (symptoms, HRQL, work productivity, satisfaction...)
  - With the other endpoints of study
  - Across studies

The IQWIG rule:
- is not tailored to each of the concepts (direct vs indirect)
- is just one responder definition among several possible
- does not resolve the issue of whether the difference in responders between groups (comparative trial or indirect comparison) is relevant or not