Assessments

IMPLICATION OF ALTERNATIVE MINIMAL CLINICALLY IMPORTANT DIFFERENCE THRESHOLD ESTIMATION METHODS ON TECHNOLOGY ASSESSMENT

Diana Brixner
Department of Pharmacotherapy, University of Utah
diana.brixner@utah.edu

Eli O. Meltzer
Allergy & Asthma Medical Group & Research Center

Kellie Morland, Cathryn A. Carroll
Xcenda, LLC

Ullrich Munzel
MEDA Pharma GmbH

Brian J. Lipworth
Institute of Health and Wellness, University of Glasgow

Objectives: Various minimal clinically important difference (MCID) threshold estimation techniques have been applied to seasonal allergic rhinitis (SAR). The objectives of this study are to (i) assess the difference in magnitude of alternative SAR MCID threshold estimates and (ii) evaluate the impact of alternative MCID estimates on health technology assessment (HTA).

Methods: Data describing change from baseline of the reflective Total Nasal Symptom Score (rTNSS) for four intranasal SAR treatments were obtained from United States Food and Drug Administration-approved prescribing information. Treatment effects were then compared with anchor-based MCID thresholds derived by Barnes et al. and thresholds obtained from an Agency for Healthcare Research and Quality (AHRQ) panel.

Results: The change in rTNSS score from baseline, represented as the average of the twice-daily recorded scores of the rTNSS, was -2.1 ($p < .001$) for azelastine hydrochloride 0.10%, 1.35 ($p = .014$) for ciclesonide, and -1.47 ($p < .001$) for fluticasone furoate. The change in the rTNSS score from baseline, represented by sum of the AM and PM score, was -2.7 for MP-AzeFlu ($p < .001$). The rTNSS change from baseline for each product was compared with anchor-based MCID threshold and the AHRQ panel estimates. Comparison of the observed treatment effect to the anchor-based and AHRQ panel MCID thresholds results in different conclusions, with clinically important differences being inferred when anchor-based estimates serve as the reference point.

Conclusion: The AHRQ panel MCID threshold for the rTNSS was twelve times larger than the anchor-based estimates resulting in conflicting recommendations on whether different SAR treatments provide clinically meaningful benefit.

Keywords: Treatment outcome, Outcome assessment, Seasonal allergic rhinitis, Intranasal corticosteroid, Intranasal combination

In 1993, Kozma and colleagues advanced a new framework for outcomes research known as the "ECHO" model (1). Using this new framework, the authors cautioned the medical community against relying predominantly on efficacy data generated from clinical trial programs. Rather, the authors suggested that medical decision making should simultaneously consider the economic, clinical, and humanistic impacts of alternative treatment options.

During the same time period, significant advancements were being made in humanistic outcomes research. Due to the contributions of interdisciplinary teams around the world, a proliferation of studies designed to develop and validate patient reported outcome (PRO) measures occurred. These scientifically validated measures provided the conduit for gathering general or disease related patient reports on various aspects of either their disease or treatment. With the availability of patient reported outcomes data, thresholds quantifying a clinically meaningful improvement could be developed. These thresholds are now known as minimal clinically important difference (MCID) thresholds.

Alternative Methods to Derive MCID Thresholds

In recent years, the analytic rigor required to support MCID threshold estimation has increased. The work of numerous interdisciplinary teams including biostatisticians, epidemiologists, physicians, and others have culminated in the identification of three different and unique strategies for MCID development (2–13). Table 1 provides a summary of the various techniques to derive MCID threshold estimates and provides examples of them.

Two quantitative strategies, the distribution-based and anchor-based MCID threshold estimation methods, have been advanced more recently in the scientific literature.

Table 1. General Considerations Related to MCID Threshold Estimate Derivation across Multiple Therapeutic Areas

Threshold estimation technique	MCID threshold estimate derived through this technique	Advantages	Disadvantages
Technical expert panel estimates (13)	3.6 units	Uses expert opinion to guide determination of MCID	Variation in estimate can occur based on the expert panel defining the MCID
Distribution-based MCID estimates (14)	0.59 units	Assesses the measurement precision of a given PRO measure	Sample-dependent and difficult to ascertain the clinical relevance for a given change
Anchor-based MCID estimates (14)	0.23–0.28 units	Uses validated PRO data as the anchor for the estimate and is not sample-dependent	Computationally more involved

MCID, minimum clinically important difference; PRO, patient reported outcome.

Distribution-based MCID estimates report the change in a PRO measure based on the normal variability observed in the data and, in general, assume that a change greater than X times the baseline standard deviation is clinically meaningful (14). The utility of distribution-based methods may be limited because they are based on statistical reasoning only. Also, variations in standard deviations may be observed when PRO data are obtained from heterogeneous patient populations. Because of these limitations, anchor-based threshold estimates are generally perceived as the more rigorous of the two techniques (14). The third strategy involves the development of MCID threshold estimates by technical expert panelists. Generally this method uses clinical experience to derive the estimates instead of patient reported outcome data. When all three types of threshold estimates are available, thresholds developed based upon clinical expert opinions have been recommended as a supplementary strategy to existing anchor- and distribution-based estimates of MCID (14).

MCID Estimates and Technology Assessment of SAR Treatment Options: A Case Study

The Total Nasal Symptom Score (TNSS) is a PRO measure required by marketing authorization authorities (eg, the United States Food and Drug Administration) to assess patient perceptions of the benefits of alternative treatments for seasonal allergic rhinitis (SAR) (15). The reflective Total Nasal Symptom Score (rTNSS) measures the overall effectiveness of a treatment in controlling symptoms of a pre-specified (e.g., 12 hr) period of time, and AM and PM reports are typically provided by patients. Patients report their symptom severity on a scale of zero (no symptoms) to three (severe symptoms) for up to four different symptom types. The four symptom types assessed by the rTNSS include rhinorrhea, nasal congestion, nasal itching, and sneezing.

Computation of treatment effectiveness using the TNSS as the primary endpoint for clinical trials for SAR treatments can be derived through various scoring mechanisms. In gen-

eral, however, treatment effectiveness is defined as the change in TNSS score from baseline. With respect to calculating the score at any time point in the study, one technique uses an average of the score recorded by the patient in the morning and the evening. Another method of scoring involves summing of the morning and evening patient reported data.

Using patient reported data derived from the TNSS, Barnes and colleagues derived MCID estimates using several analytical techniques. A more in-depth discussion of data sources and statistical methods used by Barnes et al. to generate MCID estimates using data collected from patients with SAR is described elsewhere (14). However, one of the MCID thresholds derived used the preferred direct anchor-based approach with estimated MCIDs ranging from 0.28 units (95% confidence interval [CI]: -0.18–0.73) and 0.23 units (95% CI: -0.16–0.62). This estimate is used to determine whether four different intranasal steroids' efficacy data as found in the approved prescribing information would likely be of a magnitude that patients would perceive the improvements to be clinically meaningful.

Table 2 provides a summary of the data that is used to complete the comparison. The focus of the data presented is solely on current representatives of intranasal steroid and antihistamine treatments. A wider complementary analysis representing further treatment options for SAR is described elsewhere (16). When the absolute value of treatment effect is larger than the absolute value of the estimated MCID thresholds the treatment is assumed to not only provide a statistically significant improvement in patient's symptoms but also a clinically meaningful improvement as well. Data provided in Table 2 illustrate that all treatments advance both a statistically significant and clinically meaningful improvement in symptom relief for SAR patients.

In July 2013, the Agency for Healthcare Research and Quality (AHRQ) released a comparative effectiveness review of SAR treatments (17). Despite the availability of anchor-based estimates in 2010 (14), the AHRQ reviewers used consensus-based technical panel MCID threshold estimates (17). Two panel members considered a 4-point change

Table 2. Alternative MCID Threshold Estimates

Product	N	Change from baseline of active ingredient	Change from baseline of vehicle placebo control	Difference in mean change[b]	P-value	MCID (14)	Equal or exceed the MCID threshold?
Azelastine 0.10%	169	-4.2	-2.1	-2.1	<0.001	0.28	Yes
Ciclesonide 200 mcg	148	-5.26	-4.38	-1.35	0.014	0.28	Yes
Fluticasone furoate 27.5 mcg	127	-3.84	-1.83	-2.07	<0.001	0.28	Yes
MP-AzeFlu[a]	207	-5.6	-2.9	-1.35[c]	<0.001	0.28[c]	Yes

[a]Assessed using a 24-point scale and normalized to a 12-point scale for purposes of comparison.
[b]Clinically important differences occur when the *absolute value* of the mean change in rTNSS exceeds the anchor-based MCID threshold estimate.
[c]When multiple values of treatment effect/benefit were reported in the approved prescribing information, the most conservative (lowest) estimate was included and all others were excluded from the evaluation.

meaningful and one advisor considered a 2-point change meaningful for the TNSS. The panel ultimately concluded that 30% of the maximum score (i.e., a MCID of 3.6) was an appropriate threshold for SAR technology assessment. The estimates of this panel may have likely been informed by federal documents guiding the development of drugs for allergic rhinitis and prior research by Bousquet and colleagues (15;18). A closer look at prior research reveals that the primary focus was on the assessment of responsiveness.

While responsiveness and MCID threshold estimation may be related, the two concepts should not be inferred as being interchangeable as one focuses on individual patient effects and the other compares population means. Irrespective of the underlying rationale for the selection, a 30% (3.6 units) MCID threshold derived from the AHRQ panel substantially exceeds the anchor-based MCID threshold derived through rigorous statistical testing of patient reported rTNSS data. Application of the 3.6 threshold in a manner consistent with the previous evaluation would suggest that none of the evaluated intranasal products provide clinically meaningful benefits in patient populations for which they have been approved and routinely used in clinical practice.

DISCUSSION

Healthcare decision makers are faced with a difficult task on a daily basis: assessing the quality of evidence for multiple therapies and interpreting the plethora of data to guide reimbursement and coverage determinations that affect patients' lives, physically and monetarily. Within the United States (US), payers may rely upon data provided within Academy of Managed Care Pharmacy Format for Formulary Submissions (19) and supplemental systematic reviews and additional reports. External to the United States, HTA organizations such as the National Institute for Health and Care Excellence in the United Kingdom and the Institute for Quality and Efficiency in Health

Care in Germany provide guidance about whether the value of an individual product may warrant reimbursement (20;21).

Over the past few decades, researchers have sought to obtain high quality patient reported data and quantify MCID thresholds that can be used to provide clinicians and reimbursement authorities with a secondary reference threshold to evaluate the clinical effectiveness of alternative therapies (12). The MCID, first defined by Jaeschke et al., is "the smallest difference in score in the domain of interest which patients would perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (3). Thus, MCIDs may be an important component of medical decision making if populations of patients are to maintain sufficient access to treatment options most likely to provide them with meaningful improvement in symptoms.

Furthermore, the case study above demonstrated that reimbursement decisions may vary based upon the MCID threshold estimation technique used. To mitigate the risk of denying patient access to treatment options that may result in clinically important improvements in their treatment, all available methods should be considered and presented in health technology assessment reports. A critical assessment should discuss differences in obtained results and check whether the applied MCID is sensitive for detecting relevant effects.

Before clinicians can routinely use MCID to guide decisions specific to treatment plans of individual patients, much work is needed by the research community in collaboration with payers and reimbursement authorities. There may be merit in bringing experts from around the globe together to discuss many of the issues identified in this study. For example, this study identified that the TNSS has favorable endorsement for marketing authorization and regulatory bodies for use in pivotal clinical trials. Once a PRO gains this favorable endorsement, should there be an immediate focus on developing MCID threshold estimates in a uniform manner to guide

subsequent reimbursement/payer decisions? Also, what entity, if any, should assume responsibility for monitoring ranges of MCID threshold estimates and commissioning action when large variations exist? Finally, what group should be identified as monitoring the medical literature and designated terms associated with MCID threshold estimation? Encouraging standardization around such terminology may support ease of access of such data when it is needed by clinicians or reimbursement authorities.

As with any evaluation, this review and associated case study have certain limitations. First, the statistical analysis supporting the derivation of MCIDs using alternative methods is simply reviewed. Additional detail is provided for the review in the study by Barnes et al. (14). In addition, the therapeutic alternatives in the evaluation are not complete and only a few intranasal treatment options for SAR are considered. The authors make no apology for the simplicity. Quality, transparency and simplicity in decision making should always be a desired goal in medicine. The narrow focus on a select few current treatment options for SAR was deemed adequate to illustrate the methodological points and was motivated by the authors' awareness that additional research was ongoing to support further therapeutic options for the treatment of SAR (16).

CONCLUSION

Medical decision making today is greatly enhanced because of high quality outcomes research. Clinical trials continue around the global and provide healthcare decision makers insights on the potential advantages or disadvantages of alternative treatment options for a given disease state. Reliance, however, on this information solely for the purposes of medical decision making continues to be an area of public debate.

With the proliferation of high quality humanistic outcomes research data, medical decision makers may now have at their disposal high quality evidence to better understand the influence of alternative technologies on patient reported health status. The translation of PRO data into methodologically sound MCID estimates and the consistent application of rigorously developed MCID threshold estimates are important steps in attaining the appropriate balance between clinical and humanistic outcomes. Individuals and entities conducting health technology assessment have a professional and ethical responsibility to use MCID estimates derived from rigorous statistical analysis of population based patient reported outcomes when available.

CONFLICTS OF INTEREST

REFERENCES

1. Kozma CM, Reeder CE, Schulz RM. Economic, clinical, and humanistic outcomes: A planning model for pharmacoeconomic research. *Clin Ther*. 1993;15:1121-1132.
2. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*.1989;27:S178-S179.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-415.
4. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol*. 1994;47:81-87.
5. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: A comparison of two techniques. *J Clin Epidemiol*. 1996;49:1215-1219.
6. Sloan JA. Assessing the minimally clinically significant difference: Scientific considerations, challenges, and solutions. *COPD*. 2005;2:57-62.
7. Make B. How can we assess outcomes of clinical trials: The MCID approach. *COPD*. 2007;9;191-194.
8. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: A review of concepts and methods. *Spine J*. 2007;7:541-546.
9. Revicki DA. Interpreting and reporting results based on patient-reported outcomes. *Value Health*. 2007 S116-S124.
10. Revicki D. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102-109.
11. Barrett B. Sufficiently important differences: Concepts, caveats and challenges. *Med Decis Making*. 2013;8:869-874.
12. Hedayat AS, Want J, Xu T. Minimum clinically important difference in medical studies. *Biometrics*. 2015;71:33-41.
13. McGlothlin AE, Lewis RJ. Minimal clinically important difference: Defining what really matters to patients. *JAMA*. 2014;312:1342-1343.
14. Barnes ML, Vaidyanathan S, Williamson PA, Lipworth BJ. The minimal clinically important difference in allergic rhinitis. *Clin Exp Allergy*. 2010;40:242-250.
15. Food and Drug Administration. Guidance for industry. *Allergic rhinitis: Clinical development programs for drug products*. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071293.pdf. Published June 14, 2000. Accessed February 20, 2015.

16. Meltzer EO, Wallace D, Dykewicz M, Shneyer L. Minimal clinically important difference (MCID) in allergic rhinitis: Agency for Healthcare Research and Quality or anchor-based thresholds? *J Allergy Clin Immunol Pract*. 2016;4:682-688.

17. Agency for Healthcare Research and Quality. *Treatments for seasonal allergic rhinitis. Effective health care program*. Comparative Effectiveness Review Number 120. 2013. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

18. Bousquet PJ, Combescure C, Klossek JM, Daurès JP, Bousquet J. Change in visual analog scale score in a pragmatic randomized cluster trial of allergic rhinitis. *J Allergy Clin Immunol*. 2009;123:1349-1354.

19. FMCP Format Executive Committee. The AMCP format for formulary submissions version 3.0. *J Manag Care Pharm*. 2010;16(Suppl A): 1-30.

20. National Institute for Health and Care Excellence. *About*. http://www.nice.org.uk/about.Accessed December 16, 2014.

21. Institute for Quality and Efficiency in Healthcare. *General methods*. https://www.iqwig.de/download/IQWiG_Methoden_Version_4-2.pdf. Published April 22, 2015. Accessed February 20, 2015.