

EDITORIAL

Clinical & Experimental Allergy

From statistical significance to clinical relevance

This editorial discusses the findings of the paper by Barnes et al. [13]

R. Gerth van Wijk

Department of Internal Medicine, Erasmus MC's, Rotterdam, The Netherlands

The evidence for efficacy of an intervention can be obtained from well-designed randomized-clinical trials and systematic reviews. The current trend is to prove efficacy in large clinical trials. Because of the size of these studies, there is an obvious risk that small but statistically significant inpatient changes or between-patient differences are being observed. The question, however, remains as to whether such observations are relevant for patients and clinicians. Demonstrating statistical significance can be seen as a necessary first step only.

Regulatory bodies are also adopting this view. In the recent Guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases, the European Medicine Agency (EMA) states that 'Regardless of the choice of the primary efficacy parameter, the applicant should provide a definition of a clinically meaningful effect in the primary efficacy endpoint and the basis for choosing this value. A merely statistical significant effect might not be sufficient' [1].

In last 40 years, patient-related outcomes (PROs) have been included as outcome measures in clinical trials [2]. PROs represent reports coming from the patients. Examples of PROs are symptom scores, visual analogue scores (VAS) and health-related quality of life measurements. Both the FDA and the EMA underwrite the importance of PROs. In studies of new drugs, patient evaluation should be performed with this kind of PROs. These developments represent a shift from disease-centred to patient-centred medicine. The use of PRO measures calls for an approach to determine to what extent a change in outcome is really meaningful for the patient. With the introduction of methods to assess health-related quality of life, the concept of minimal clinically important difference (MID or MCID) was put forward. The MID has been defined as the smallest difference in the scores of a PRO measure that is

perceived by patient to be beneficial or harmful and that would lead the clinician to consider a change in treatment [3, 4].

Basically, two sets of methods are commonly used to determine the MID. Anchor-based methods use an external reference for comparison with the PRO under study. This indicator can be patient based; however, laboratory and physiological measures or clinician's ratings are also possible. In many studies, global assessments of change in the health status are taken as the anchor for the PRO of interest. Another approach comprises the use of distribution-based methods. For instance, a 0.5 SD [5] or the SEM [6, 7] have been suggested to approximate an MID for some instruments. The advantage of anchor-based methods is that the outcome of the analysis is much more understandable for the patient and the clinician. However, the anchor-based approach requires studies that include the external reference for comparison. The distribution-based approach has the advantage that this can be used in the absence of an anchor. This advantage is also the weak part of distribution-based methods as the MID has no relation with the patients. In general, there is a preference for the use of anchor-based approaches. In the absence of an anchor-based MCID, distribution-based methods may be used. Moreover, distribution-based methods can complement the anchor-based approach.

In clinical trials of the efficacy of intervention in patients with allergic rhinitis, a few outcome measures are commonly used. These comprise daily symptom scores, VAS and combined symptom-medication scores (mostly used in immunotherapy trials). In 1991, Juniper and Guyatt [8] introduced the rhinitis-related quality-of-life questionnaire (RQLQ). Although, since then, many other instruments for patients with rhinitis have been developed, the seven-point RQLQ and the mini-RQLQ [9] are the most commonly used HRQL instruments. For these two instruments, a MID of 0.5 and 0.7, respectively, have been suggested. Sparse data are available on the clinical relevance of other PROs used in rhinitis trials. In an attempt to express the ARIA categories of severe/moderate

Correspondence:

Roy Gerth van Wijk, Department of Internal Medicine, Erasmus MC's, Gravendijkwal 230 3015 CE Rotterdam, The Netherlands.
E-mail: r.gerthvanwijk@erasmusmc.nl

and mild rhinitis in terms of VAS, Bousquet et al. [10] demonstrated that a 0–10 cm VAS below 5 cm might be classified as mild rhinitis, whereas those with a VAS of over 6 cm could be considered as having moderate/severe rhinitis. Recently, the same group analysed the VAS in comparison with the symptom scores and the RQLQ in a two-week clinical trial [11]. They used ROC curves and cost function to assess VAS cut-off scores to distinguish between patients with no clinical improvement and patients with improvement in the symptoms, quality of life or both. The optimal cut-off in the VAS change separating the patients without improvement from those with improvement is 0.30 cm. Using the cost of false-positive and false-negative results, a difference of > 1 cm is significant. However, when patients show an improvement in both symptoms and RQLQ scores, the median improvement on the VAS is 7 cm.

In the search for clinically relevant measures, the number of well days has been proposed and used in immunotherapy trials [12]. This outcome represents the days without symptoms or days without the need for rescue medication. Its advantage is obvious. It can be well understood by patients and clinicians.

In this issue of the journal, Barnes et al. [13] present calculated MCIDs for a series of outcome measures used in rhinitis trials. One has to realize that a single MCID value for a PRO instrument does not exist across all patient samples. The MCID may be different according to the population studied [14]. For instance, it is easy to understand that a meaningful change in outcome may be different for a patient with severe asthma compared with subjects having mild to moderate asthma. The authors attempt to provide more precise estimates of the MCIDs using the data of nine randomized placebo-controlled clinical trials in intermittent and persistent allergic rhinitis. They used both anchor- and distribution-based methods. Different ways of analysis were used to include all available data. Meta-analysis techniques allowed the authors to pool data. Interestingly, the authors also examined peak nasal inspiratory flow (PNIF) and nasal NO. They succeeded in calculating an MCID for PNIF but they failed to determine the MCID for nasal NO. It is more difficult to establish an MCID for an outcome measure that has no particular meaning for the patient. Moreover, assessment of nasal NO has technical difficulties, thereby hampering its use as a tool to monitor rhinitis.

The study of Barnes et al. is important as it is an attempt to stop a gap. Research into commonly used outcome measures in trials with patients with allergic rhinitis is sparse. Moreover, tests on nasal function such as PNIF, rhinomanometry, acoustic rhinometry and nasal challenge tests are less well established as tools for randomized-controlled trials. PROs as symptom scores, VAS, medication scores and HRQL instruments are therefore the outcome measures of choice. Apart from the tools assessing HQLQ,

these instruments are not well validated. Also, EMEA mentions this lack of validated symptoms and symptom/medication scores [1]. Calculation of MCIDs has seldom been carried out in rhinitis and rhinosinusitis patients.

In this area, more research is needed. Based on their calculations, Barnes proposes another MCID for the mini-RQLQ, than Juniper did. This might be confusing for the reader, but the differences arise from a different interpretation of the anchor's threshold that determines the MCID. It is difficult to say which view is most correct. Taking into account that the MCID may also vary in different clinical settings, it is possible that eventually the MCID should be represented as a range of values. The challenge is to limit this range as much as possible to identify a single value or a narrow range. The approach of Barnes intends to meet this challenge.

Other ways of expressing meaningful changes in clinical trials have been put forward. Barret proposed the concept of a sufficiently important difference (SID). He defined SID as the smallest amount of patient-valued benefit that an intervention would require to justify associated costs, risks and other harms [15]. Bousquet defined relevant differences from an investigator's point of view. He took into account the change that placebo interventions usually generate in trials with allergic rhinitis. For this reason, a relevant change in RQLQ was estimated as 1.0, whereas a meaningful change in RTSS4 amounted to 3.0 [11].

The inclusion of MCIDs in clinical trials has several advantages [16]. Not only is the patient's perspective taken into account, clinicians are more able to interpret the results of a trial and to understand the success or the lack of success of an intervention. MCIDs help to estimate the sample size and it is also easy to calculate the proportion of patients who achieve the MCID. From this proportion, the number needed to treat can be derived.

Therefore, accepted ways to define and include clinical relevance will enhance the quality of clinical trials and contribute to a better positioning of new drugs and interventions into clinical practice.

References

- 1 European Medicine Agency. Guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases. Available at <http://www.emea.europa.eu/pdfs/human/ewp/1850406enfin.pdf>, 2009 (accessed 24 November 2009).
- 2 Baiardini I, Bousquet P, Brzoza Z *et al.* Recommendations for assessing patient-reported outcomes including health related quality of life in clinical trials on allergy: a GA2LEN taskforce position paper. *Allergy* 2009; in press.
- 3 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; 40:171–8.
- 4 Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting G. Methods to explain the

- clinical significance of health status measures. *Mayo Clin Proc* 2002; **77**:371–83.
- 5 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41**:582–92.
 - 6 Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; **37**:469–78.
 - 7 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; **52**:861–73.
 - 8 Juniper EF, Guyatt GH. Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis. *Clin Exp Allergy* 1991; **21**:77–83.
 - 9 Juniper EF, Thompson AK, Ferrie PJ, Roberts JN. Development and validation of the mini rhinoconjunctivitis quality of life questionnaire. *Clin Exp Allergy* 2000; **30**:132–40.
 - 10 Bousquet PJ, Combescure C, Neukirch F *et al.* Visual analog scales can assess the severity of rhinitis graded according to ARIA guidelines. *Allergy* 2007; **62**:367–72.
 - 11 Bousquet PJ, Combescure C, Klossek JM, Daures JP, Bousquet J. Change in visual analog scale score in a pragmatic randomized cluster trial of allergic rhinitis. *J Allergy Clin Immunol* 2009; **123**:1349–54.
 - 12 Worm M. 'Well days' after sublingual immunotherapy with a high-dose 6-grass pollen preparation. *Allergy* 2009; **64**: 1104–5.
 - 13 Barnes ML, Vaidyanathan S, Williamson PA, Lipworth BJ. The minimal clinically important difference in allergic rhinitis. *Clin Exp Allergy*, doi: 10.1111/j.1365-2222.2009.03381.x.
 - 14 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; **61**:102–9.
 - 15 Barrett B, Brown R, Mundt M *et al.* Using benefit harm tradeoffs to estimate sufficiently important difference: the case of the common cold. *Med Decis Making* 2005; **25**:47–55.
 - 16 Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes* 2006; **4**:69.