

Quality of life questionnaires: Does statistically significant = clinically important?

A statistically significant difference between treated and control groups does not necessarily mean that the difference is also of clinical importance. Clinical importance refers to the magnitude of the difference between the two treatments, whereas statistical significance indicates the probability of being able to reject the hypothesis of "no difference."¹

In recent years, recognition of the importance of measuring subjective health status has stimulated the development of a number of new questionnaires. Some questionnaires measure clinical status (often symptoms recorded in a diary), others measure functional status (e.g., generic health profiles such as the SF-36 and the SIP), and more recently, we have seen the advent of disease-specific quality of life questionnaires. Although a great deal of attention has been paid to the methods of instrument construction and the need for establishing the measurement properties and validity of these instruments,² little attention has been paid to the interpretation of the data they generate.

Repeated experience with a wide variety of physiologic measures has allowed clinicians to become familiar with these indices and to interpret the results in a meaningful manner. For instance, an experienced clinician will have little difficulty interpreting a 20% increase in FEV₁ percent predicted or a fall in peak expiratory flow from 450 to 300 ml/sec. In contrast, the meaning of a change in score of 0.5 from a symptom diary or health status questionnaire is less intuitively obvious. This is because each questionnaire has its own scoring system, there are no units, and these questionnaires are seldom used in routine clinical practice.

In the article by Mahajan et al. that appears in this issue of *The Journal* (*J Allergy Clin Immunol* 1998;102:19-23), three questionnaires were used to evaluate the effect of 1 year of treatment with fluticasone propionate (50 µg or 100 µg twice a day) on functional status and sleep disturbance in children with asthma 4 to 11 years of age and on the quality of life of their parents. When compared with placebo, most differences were statistically significant. Can they

Abbreviations used

MID: Minimal important difference

NNT: Number needed to treat

also be considered clinically important? How many children will have to be treated with fluticasone for one to have a clinically meaningful benefit in these outcomes? The authors have considered interpretation but acknowledge that the minimal clinically important difference for each of these three instruments is not known. However, such information is important because the mean differences *appear* to be small. For functional status over the 52 weeks, there was a mean improvement in the group receiving 100 µg fluticasone propionate twice a day from 92.3 to 95.3 (scale, 0 to 100; $p < 0.05$). In the figure we are provided with a mean normal value and a value for chronically ill children, but both the baseline and 52-week values fall between the normal and chronically ill values. Therefore we are still unable to determine whether the change of 3.0 is important. For the sleep scale, where change was from 89.2 to 90.6 ($p < 0.05$), and for the three of six domains of the parents' quality of life that reached statistical significance, even normal values are not available.

This study is far from unique. In a large number of respiratory and allergy clinical trials, clinical status, functional status, and quality of life data are presented only in terms of their statistical significance, with little or no attention paid to clinical relevance. In large clinical trials in which instruments with good responsiveness are used, even the most trivial differences between intervention will reach statistical significance.

For quality of life questionnaires, determination of clinical importance is already being addressed. Two approaches are used: one is *anchor-based* and the other is *distribution-based*.³ The former bases interpretation on clinically meaningful changes in other indices, and the latter depends entirely on statistical distribution. A limitation of the statistical approach is that even if one calculates the effect size, there is no evidence that this magnitude of change is of any importance to the patient. In addition, it has been suggested that the effect size is more an estimate of the responsiveness of the questionnaire and may be either an overestimate or an underestimate of the minimal clinically important difference.⁴

Several quality of life research groups have adopted the anchor-based approach. We use patient perception of a meaningful change in quality of life as the anchor to

From the Department of Clinical Epidemiology and Biostatistics, McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada.

Received for publication Feb. 11, 1998; revised Feb. 28, 1998; accepted for publication Mar. 2, 1998.

Reprint requests: Elizabeth Juniper, MCSP, MSc, Department of Clinical Epidemiology and Biostatistics, McMaster University Medical Centre, 1200 Main St. West, Hamilton, Ontario, L8N 3Z5, Canada.

J Allergy Clin Immunol 1998;102:16-7.

Copyright © 1998 by Mosby, Inc.

0091-6749/98 \$5.00 + 0 1/1/90088

calculate the minimal important difference (MID) for our questionnaires. We have defined the MID as "the smallest difference in score in the domain of interest that patients perceive as beneficial and would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management."⁵ A similar anchor-based approach has been used to establish the MID for the St. George's Respiratory Questionnaire.⁶ A slightly different technique of asking patients to compare their experiences with those of other patients has been used to establish the MID for rheumatoid arthritis instruments.⁷

If the MID for a clinical trial outcome (physiologic, clinical, or functional) is known, the interpretation may be taken a step further. For a long time, statisticians have cautioned against the sole use of mean data for the interpretation of clinical trial data. One of the reasons is that patients may be very heterogeneous in their responses to interventions. Wide heterogeneity (i.e., large standard deviation) may be well hidden if only the standard error of the mean or *p* values are presented. However, if the MID is known, the problem can be addressed by calculating the number needed to treat (NNT).⁸ The NNT is the number of patients who need to be treated with the new intervention for one patient to have a clinically important improvement over and above that which he or she would have experienced with the control intervention.⁸ It is calculated from the proportion of patients who show an improvement or deterioration (i.e., all those who show a change greater than the MID) on each of the treatments. Presentation of clinical trial results in this manner is meaningful for clinicians and easier to interpret than standard deviations and confidence intervals.

For example, in a recent comparison of salmeterol and salbutamol in which the Asthma Quality of Life Questionnaire was used as an outcome measure, the mean difference between the two interventions for the symptom domain was 0.49 (scale, 1 to 7; *p* < 0.0001).⁹ This value was on the borderline of being clinically important (MID = 0.5). Further analysis revealed an NNT of 4.⁹ In other words, four patients need to be treated with salmeterol for one to have a clinically important improvement over and above that which he or she would have experienced with salbutamol (i.e., each patient has a 25% chance of having a clinically important improvement). NNT results are often quite surprising to those who expect the majority of patients to respond close to the mean.

So far the establishment of MIDs and the calculation

of NNTs from clinical trial results have been restricted to those studies using quality of life questionnaires. The problem of interpreting change in generic functional status questionnaire scores is more difficult, but it is receiving attention.¹⁰ The problem of interpreting symptom scores has received markedly little attention even though they are used as primary outcomes in a large number of efficacy clinical trials. This is an area of research that requires attention.

All treatment decisions involve balancing a number of factors. These factors include the probability of patients receiving a clinically important benefit, the importance of that benefit (e.g., mortality risk reduction versus improved quality of life), the risk of adverse effects, and costs. It is important for clinicians to have the first of these factors available for this decision making process.

Elizabeth F. Juniper, MCSP, MSc

*Department of Clinical Epidemiology and Biostatistics
McMaster University Faculty of Health Sciences
Hamilton, Ontario, L8N 3Z5
Canada*

REFERENCES

1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little Brown; 1991.
2. Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spiker B, editor. Quality of life and pharmacoeconomics in clinical trials. Philadelphia: Lippencott-Raven; 1996. p. 49-56.
3. Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res 1993;2:221-6.
4. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987;40:171-8.
5. 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;47:81-7.
6. Jones PW, Lasserson D. Relationship between change in St. George's Respiratory Questionnaire score and patients' perception of treatment efficacy after one year of therapy with nedocromil sodium. Am J Respir Crit Care Med 1994;149:A211.
7. Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.
8. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldenstein RS. Interpreting treatment effects in randomised trials. Br Med J 1998; 316: 690-3.
9. Juniper EF. The value of quality of life in asthma. Eur Respir Rev 1997;49:333-7.
10. Ware JE, Keller SD. Interpreting general health measures. In: Spiker B, editor. Quality of life and pharmacoeconomics in clinical trials. Philadelphia: Lippencott-Raven; 1996. p. 445-60.