Assessment of outcome in dyspepsia: has progress been made?

S J O Veldhuyzen van Zanten

There is a lack of consensus among researchers on how to best measure outcome in functional dyspepsia trials and more importantly a lack of validated outcome measures. If symptoms resolve completely, treatment has been successful but with partial improvement interpretation is less straightforward. It is most likely that these issues will only be resolved if unequivocally efficacious treatments emerge to which the different outcome measures can be compared. Recently, a few validated outcome measures have been developed which look promising.

Correspondence to:
Dr S J O Veldhuyzen van
Zanten, Division of
Gastroenterology,
Dalhousie University,
Queen Elizabeth II Health
Sciences Centre, Victoria
General Site, Room 928,
Centennial Building, 5790
University Avenue, Halifax,
Nova Scotia, Canada
B3H 2YG;
zanten@is.dal.ca

SUMMARY

A systematic review of the design of functional dyspepsia trials has highlighted the problem of a lack of consensus over how to best measure outcome. Outcome measures can be broadly categorised as global scales, generic instruments, and disease specific instruments. An example of a global outcome measure is a Likert scale which is an interval scale that has graded definitions for the severity of symptoms, ranging from none to very severe. Before a trial is initiated, it is necessary to stipulate how much improvement on these scales is considered clinically meaningful. Complete disappearance of symptoms clearly is an acceptable outcome measure but it is less clear how partial improvement should be interpreted. A different global outcome measure is the "overall treatment effect". At the end of treatment, the patient is asked to decide whether he or she has remained the same, improved, or deteriorated, and improvement or worsening is rated on, for example, a one to seven point ordinal scale. Generic instruments are quality of life questionnaires that are applicable across different populations. An example of a generic instrument is the psychological general well being (PGWB) index. Disease specific instruments can be categorised as unidimensional or multidimensional. Unidimensional scales (for functional dyspepsia) focus mainly on gastrointestinal symptoms whereas multidimensional scales may also include domains such as emotional or social functioning and the impact that symptoms have on daily activities. An example of a unidimensional scale is the gastrointestinal symptom rating scale (GSRS) and an example of a multidimensional scale is the Glasgow dyspepsia severity score (GDSS).

INTRODUCTION

Functional dyspepsia is defined as persistent or recurrent pain or discomfort centred in the abdomen, without evidence of organic disease that is likely to explain the symptoms. It may be associated with other symptoms, such as upper abdominal fullness, excessive burping or bloating, nausea, retching, and early satiety. As no objective structural or pathophysiological measures exist to assess outcome, one has to rely on the subjective reporting of symptoms by the patients and their impact on normal daily activities to decide whether a treatment intervention is of benefit.

A systematic review of the design of functional dyspepsia trials has highlighted the problem of a lack of consensus among researchers as to how to best measure outcome.[1] More importantly, there is a lack of validated outcome measures. Only five of the 52 studies included in the review had used a validated scale. Furthermore, the placebo response was high, ranging from 13% to 73%. A high placebo response also makes it difficult to prove that a new intervention is superior to placebo. The main outcome measure should be reported as the proportion of patients who achieve a predetermined outcome, rather than an average response among the different treatment groups.

A detailed discussion on the requirements for validation of outcome measures for quality of life instruments is beyond the scope of this article.[2] In brief, four requirements need to be fulfilled. The first is that symptoms need to be representative of the disease under study. Secondly, the instrument has to be reproducible; that is, the same results are achieved in patients whose health status is unchanged. Thirdly, the instrument has to be able to detect a change. Fourthly, a detected change should correlate with a change in health status. The ability of an instrument to detect change is often referred to as responsiveness.

CLASSIFICATION OF OUTCOME MEASURES

Several types of outcome measures can be used. These can be broadly categorised as global scales, generic instruments, and disease specific instruments.[2] Disease specific instruments can be unidimensional, focusing mainly on gastrointestinal symptoms, or multidimensional. Multidimensional scales also evaluate other domains,

Abbreviations: GDSS, Glasgow dyspepsia severity score; GSRS, gastrointestinal symptom rating scale; PGWB, psychological general well being.

Table 1 Seven point Likert scale. An example of a global assessment question: "please rate how severe your upper abdominal pain and/or discomfort was"

0, None
1, Minimal
2, Mild
3, Moderate
4, Moderately severe
5, Severe
6, Very severe

Table 3 The gastrointestinal symptom rating scale. An example of an assessment question: "have you been bothered by stomach ache duing the past week?"

1, No discomfort at all
2, Minor discomfort
3, Mild discomfort
4, Moderate discomfort
5, Moderately severe discomfort
6, Severe discomfort
7, Very severe discomfort

Table 2 The psychological general well being index questionaire. An example of an assessment question: "have you been anxious, worried, or upset during the past week?"

1, Extremely so—to the point of being sick or almost sick
2, Very much so
3, Quite a bit
4, Somewhat, enough to bother me
5, A little bit
6, Not at all

such as emotional or social functioning, in addition to gastrointestinal functions or symptoms. Generic instruments are questionnaires which are applicable across populations whereas disease specific instruments are developed to focus on quality of life of a specific disease.

Global scales

An example of a global outcome measure is the "seven point Likert scale" shown in table 1. Likert scales are interval scales that have graded definitions for the severity of symptoms, ranging from none to very severe. This particular scale was used as one of the main outcome measures in both the ORCHID and OCAY studies.[3][4] The definition of a responder was a patient who during the last seven days before the final assessment rated the severity of their dyspepsia symptoms as either none or minimal. Other dyspepsia trials have used similar interval scales but have reported only the average improvement in the group of patients randomised to active treatment and compared this with the average score achieved by patients randomised to placebo. For example, Gilvarry *et al* used a summary score of four symptom clusters: ulcer-like, reflux-like, dysmotility-like, and unclassified dyspepsia.[5] They measured both the severity (rated as 0, 1, or 2) and frequency (rated as 0, 1, 2, or 3) of day pain, night pain, heartburn, and nausea, and added the symptoms together to a maximum score of 20. This randomised controlled trial of *Helicobacter pylori* eradication in patients with dyspepsia compared triple therapy using bismuth, metronidazole, and tetracycline with bismuth therapy plus placebo antibiotics. In patients in whom *H pylori* was successfully eradicated, the summary score improved from 14 to 9, whereas in those in whom eradication failed, symptoms changed from 14 to 12.[5] The difference between the two groups was statistically significant.

Before a trial is initiated, it is necessary that the protocol stipulates how much improvement is considered clinically meaningful. When interpreting the results of studies that use this type of global scale, complete disappearance of symptoms clearly is an acceptable outcome measure. It is less clear though how a partial improvement should be interpreted on such a scale.

A different method of using global outcome measures is the "overall treatment effect" approach. This method has been used successfully by Jaeschke and colleagues,[6] and an example of this method is shown in table 2. At the end of treatment, the patient is asked to decide whether he or she has remained the same, improved, or deteriorated. If the patient says that there has been either improvement or worsening, this can then be rated on, for example, a one to seven point ordinal scale. This method was used in the OCAY study.[4] In this study, there were no significant differences between patients randomised to either seven day anti-*H pylori* therapy compared with control treatment consisting of a proton pump inhibitor plus placebo antibiotics.

Generic instruments

Generic instruments are quality of life questionnaires that are applicable across different populations.[2] Examples of these are the sickness impact profile,[7] the short form-36,[8] and the psychological general well being (PGWB) index.[9] The PGWB questionnaire consists of six subscales that assess anxiety, depression, vitality, well being, health, and self control. It consists of 30 questions ranked on a six point ordinal scale. An example of a question is shown in table 2.

The PGWB questionnaire has been administered to normal controls, duodenal ulcer patients, patients suffering from functional dyspepsia,[10] and patients with gastro-oesophageal reflux disease.[11] The summary score was much lower in patients with functional dyspepsia (score 87) compared with healthy controls (score 103). Patients with duodenal ulcer also had lower scores (score 85). Following cure of the ulcer, the PGWB score improved significantly in patients with duodenal ulcers, from 87 to 109.

The PGWB index was used in the ORCHID and OCAY studies.[3][4] Although the score improved slightly, there were no differences over the 12 months of follow up. For example, in the OCAY study, the overall PGWB score changed from 93 to 98 in patients randomised to omeprazole, amoxycillin, and clarithromycin compared with a change from 94 to 100 in patients treated with omeprazole and placebo antibiotics.[4]

Disease specific instruments

Disease specific instruments for functional dyspepsia can be categorised as unidimensional or multidimensional. Unidimensional scales (for functional dyspepsia) focus mainly on gastrointestinal symptoms whereas multidimensional scales may also include other domains, such as emotional or social functioning and the impact that symptoms have on daily activities. An example of a unidimensional scale is the gastrointestinal symptom rating scale (GSRS).[12] This instrument consists of 15 questions graded on seven point Likert scales. An example of a scale is shown in table 3. GSRS has five domains: abdominal pain, reflux, indigestion, diarrhoea, and constipation. The results of the GSRS are expressed as the mean total score (that is, the response to all questions is added and then divided by 15). The GSRS has been used successfully in a variety of studies.

An example of a recently validated multidimensional disease specific scale for dyspepsia is the Glasgow dyspepsia

Table 4 Summary of the Glasgow dyspepsia severity score scale

Dyspeptic symptoms	
1, Frquency	0–5
2, Effect on normal activities	0–2
3, Time off work	0–2
Consultation with physician	
1, In physician's office	0–2
2, Home visits by GP	0–2
3, Tests for dyspepsia	0–2
Medication for dyspepsia	
1, Over the counter	0–2
2, On prescription	0–3
Total	0–20

GP, general practitioner.

severity score (GDSS) developed by El-Omar and colleagues.[13] A summary of the scale is given in table 4. It focuses on several aspects of dyspepsia: firstly, the frequency of dyspepsia symptoms and the effect that they have on normal activities and ability to work; secondly, the need for consultations with physicians for dyspepsia and the need for diagnostic investigations for dyspepsia; and thirdly, the need for over the counter and prescription medication for dyspepsia.

The GDSS scale was compared in healthy controls and patients with duodenal ulcers or functional dyspepsia.[14] The average score in healthy controls was 1.2 compared with 10.5 in patients with functional dyspepsia and 11.1 in patients with duodenal ulcer. Following eradication of *H pylori* in patients with duodenal ulcer, the score changed from 11.4 to 1.3, compared with an average change of 10.5 to 8.5 in patients in whom the infection was not eradicated. This scale was used by McColl *et al* in their UK Medical Research Council trial in which 315 patients with functional dyspepsia were randomised to anti-*H pylori* treatment or a proton pump inhibitor plus placebo antibiotics.[14] Patients were followed up for one year. In this trial, the main outcome was defined as the proportion of patients who scored 0 to 1 on the GDSS, indicating that the patients had to have either no or minimal symptoms. It showed a statistically significant effect in favour of omeprazole, amoxycillin, and metronidazole (21% response) compared with placebo (7%).

It is worth mentioning that for the GDSS, patients are asked to rate their symptoms over the last six months. Whether patients are able to accurately think back over a six month period is uncertain. However, it is also unclear whether there may be problems with recall if patients are asked to rank their symptoms, for example, over the preceding week.

CONCLUSION

We have briefly discussed methods by which outcome measures can be applied in functional dyspepsia trials. Recently, a few validated outcome measures have been developed and they look promising. However, further validation is necessary to confirm their operating characteristics. An important issue that has not yet been resolved is how one should interpret the different outcome measures and whether the interpretation may be different for generic and disease specific outcome measures. If symptoms resolve completely treatment definitely has been successful but with partial improvement interpretation will be less straightforward. It is most likely that these issues will only be resolved if unequivocally efficacious treatments emerge to which the different outcome measures can be compared.

Despite problems in the measurement of outcomes, some recently published treatment trials examining the effect of *H pylori* eradication on functional dyspepsia symptoms have been of high quality. Importantly, these trials used acceptable outcome measures. What is not clear is whether partial improvement of symptoms is a reasonable outcome and, if so, how much of an improvement is clinically meaningful. The latter will be ultimately important as this will determine whether such interventions are deemed cost effective.

REFERENCES

1 **Veldhuyzen van Zanten SJO**, Cleary C, Talley NJ, *et al*. Drug treatment of functional dyspepsia: a systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol* 1996;**91**:660–71.
2 **Guyatt GH**, Veldhuyzen van Zanten SJO, Feeney DH, *et al*. Measuring quality of life in clinical trials: a taxonomy and review. *Can Med J Assoc* 1989;**140**:1441–8.
3 **Talley NJ**, Janssens J, Lauritsen K, *et al*. Eradication of *Helicobacter pylori* in functional dyspepsia: randomized double blind placebo controlled trial with 12 months' follow up. The Optimal Regimen Cures *Helicobacter* Induced Dyspepsia (ORCHID) Study Group. *BMJ* 1999;**318**:833–7.
4 **Blum AL**, Talley NJ, O'Morain C, *et al*. Lack of effect of treating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. *N Engl J Med* 1998;**339**:1875–81.
5 **Gilvarry J**, Buckley MJM, Beattie S, *et al*. Eradication of *Helicobacter pylori* affects symptoms in non-ulcer dyspepsia. *Scand J Gastroenterol* 1997;**32**:535–40.
6 **Jaeschke R**, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;**10**:407–15.
7 **Gergner M**, Bobbitt RA, Carter WB. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;**19**:787–805.
8 **Stewart AL**, Hays RD, Ware JE. The MOS short-form general health survey. *Med Care* 1988;**26**:724–34.
9 **Wegner NK**, Mattson ME, Furberg CF, *et al*. Assessment of quality of life in clinical trials of cardiovascular therapies. *Am J Cardiol* 1984;**54**:908–13.
10 **Dimenäs E**, Glise H, Hallerbäck B, *et al*. Quality of life in patients with upper gastrointestinal symptoms: an improved evaluation of treatment regimens? *Scand J Gastroenterol* 1993;**28**:681–7.
11 **Wiklund I**, Halling K, Långström G, *et al*. Quality of life during acute and intermittent treatment of gastroesophageal reflux disease with omeprazole compared with ranitidine: results from a multicentre study. *Ital J Gastroenterol Hepatol* 1998;**30**:19–27.
12 **Svedlund J**, Sjödin I, Dotevall G. GSRS—a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci* 1988;**33**:129–34.
13 **El-Omar EM**, Banerjee S, Wirz A, *et al*. The Glasgow dyspepsia severity score—a tool for the global measurement of dyspepsia. *Eur J Gastroenterol Hepatol* 1996;**8**:967–71.
14 **McColl KEL**, Murray LS, El-Omar E, *et al*. Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with non-ulcer dyspepsia. *N Engl J Med* 1998;**339**:1869–74.