

## CLINICAL REVIEWS

# Symptom-Based Outcome Measures for Dyspepsia and GERD Trials: A Systematic Review

Adam Fraser, M.Sc., M.R.C.G.P., Brendan Delaney, M.D., F.R.C.P., M.R.C.G.P.,  
and Paul Moayyedi, Ph.D., F.R.C.P.

*Department of Primary Care and General Practice, Primary Care Clinical Sciences Building,  
The University of Birmingham, Edgbaston, Birmingham, United Kingdom*

- BACKGROUND:** Symptom assessment using questionnaires has been recommended as the primary outcome measure in clinical gastroesophageal reflux and dyspepsia trials. Questionnaires should have proven reliability, validity, and responsiveness, and may assess the frequency and/or severity of dyspepsia symptoms. Although a number of measures have been developed, it remains unclear which of these should be used in new trials.
- OBJECTIVE:** To describe existing questionnaire outcome measures that assess symptoms of gastroesophageal reflux dyspepsia for use in clinical trials.
- METHODS:** Studies were identified from Medline, Embase, the Cochrane library, and reference lists. The inclusion criterion was that the study assessed a questionnaire, which measured the frequency or severity of dyspepsia or gastroesophageal reflux symptoms, in a sample of patients.
- RESULTS:** No direct comparison between questionnaires was possible due to methodological heterogeneity. Thirty-seven studies describing 26 questionnaires met the inclusion criteria. Twelve were unidimensional (assessed symptoms only) and 14 were multidimensional (also assessed quality of life). Eleven questionnaires assessed both frequency and severity of dyspepsia, and 10 had proven reliability, validity, and responsiveness. No studies compared different questionnaires.
- CONCLUSIONS:** Future gastroesophageal reflux and dyspepsia clinical trials should use unidimensional or multidimensional outcome measures that assess both the frequency and severity of symptoms, and have proven reliability, validity, and responsiveness. Further research is necessary to compare existing outcome measures to determine which are the most reliable, valid, and responsive instruments.

(Am J Gastroenterol 2005;100:442–452)

## INTRODUCTION

Dyspepsia and gastroesophageal reflux disease (GERD) are common conditions, affecting around 28% of the population (1, 2). They cause significant impairment of quality of life (3). Only 20% of those with dyspepsia or GERD consult a doctor, but this accounts for 2–8% of all primary care consultations (1, 4, 5). Almost all of those who consult receive a prescribed medication, and 49% of all patients with dyspepsia take over-the-counter medications (6). In the United Kingdom, dyspepsia costs the National Health Service over 500 million every year, and the cost to society has been estimated at a further 1 billion per year (7).

“Dyspepsia” describes a symptom complex, which can be caused by several underlying conditions such as peptic ulceration, esophagitis, or gastric carcinoma. The cardinal feature of dyspepsia is a pain or discomfort in the upper abdomen, with or without reflux or dysmotility-like symptoms (8–10). There is no discernable biochemical or structural explanation

for the symptoms in around 60% of patients with dyspepsia, who are classified as having functional (or nonulcer) dyspepsia (10, 11).

Over the past 20 yr, there have been several attempts to redefine and subdivide “dyspepsia.” The 1988 working party definition (8) established the terms “ulcer-like,” “reflux-like,” and “dysmotility-like” dyspepsia. The Rome I (9) and Rome II (10) working parties defined dyspepsia as “epigastric pain” (9) and “predominant epigastric pain” (10), respectively, and excluded patients with sole or predominant reflux symptoms as GERD. GERD itself has been defined as symptoms of heartburn and/or acid regurgitation (12). It should be noted that definitions of GERD have not been refined to stipulate “predominant” heartburn, although this is implied by the Rome II dyspepsia definition. The Rome definitions were established to further research in patients with a diagnosis of functional dyspepsia following endoscopy. They have not proved sufficiently predictive of peptic ulcer or esophagitis in patients not having already undergone endoscopic

investigation, and for this reason trials in these patients need to consider effects on both epigastric pain and heartburn (13).

Despite its prevalence and impact, there is still a great deal of uncertainty regarding the management of gastroesophageal reflux disease (GERD) (14) and dyspepsia (4, 15). One difficulty for researchers is the lack of validated outcome measures (11, 14, 16–18). Since there are no structural or biochemical abnormalities in functional disease, response to treatment must be assessed primarily by measuring patients' symptoms. Assessment of symptoms by clinicians has been shown to be unreliable (19), and patient symptom diaries place a considerable burden on respondents and are difficult to analyze (19). As a result, questionnaires are the most appropriate and frequently used outcome measure for functional dyspepsia trials (11, 16–18).

Questionnaires may quantify characteristics of individual symptoms, such as their frequency, duration, and severity, or they may measure quality of life, by asking about social functioning, emotional well being, or impact of symptoms on daily activities (20). Assessment of patients' symptoms has been recommended as the primary outcome for clinical trials (17, 18), although there is a growing interest in the addition of quality-of-life assessments (21). Unidimensional questionnaires contain items concerning a single aspect of dyspepsia, such as symptom severity, whereas multidimensional instruments assess multiple aspects of the condition, usually symptoms and quality of life. Global scales use a single question to rate the overall severity of the condition.

Before they are used in clinical trials, questionnaires must undergo prior validation for three characteristics—reliability, validity, and responsiveness (11, 17, 18, 22). Reliability is the ability of an instrument to produce the same result in response to the same clinical state on different occasions. Validity refers to whether the questionnaire is actually measuring what it is designed to measure. Responsiveness describes whether the questionnaire is capable of detecting changes in a condition over time, which might reflect therapeutic effects.

A variety of GERD and dyspepsia outcome measures, based on the changing definitions of dyspepsia, have been used in clinical trials, often without prior validation. In 1994, a review of functional dyspepsia trials found that of 16 studies, none had used a validated outcome measure (18). Another review conducted in 1996 found that only 5 of 52 studies had used a previously tested measure (17), but even these outcome measures required further validation (11, 18). Research concerning the most appropriate measure has been recommended as a research priority (11, 16–18). The most recent review of outcome measures for dyspepsia was published in 1993 (11); since then a variety of new measures have been developed and tested. This systematic review summarizes published questionnaire outcome measures that assess the symptoms of GERD or dyspepsia to facilitate a debate as to the most appropriate measure for use in clinical trials.

## METHODS

Studies were included if they involved a questionnaire that quantified the frequency or severity of GERD or dyspepsia symptoms, which was tested in a sample of patients. Outcome measures for both GERD and dyspepsia were included to widen the scope of the review and make it relevant for research in both areas. This is particularly important for primary care research, where it is difficult to differentiate between diagnoses due to substantial overlap between symptoms of GERD and dyspepsia. The broader inclusion criteria reflect the contemporary debate about how best to define dyspepsia in primary care settings where patients have not been endoscoped (23).

Search strategies were completed in July 2004 using Medline (from 1966), Embase (from 1980), and the Cochrane library (issue 2, 2004). The search terms used were “dyspepsia,” “heartburn,” “questionnaires,” and “health surveys.” The title and abstract of each citation identified were reviewed to judge eligibility. Reference lists of eligible studies were searched for further citations.

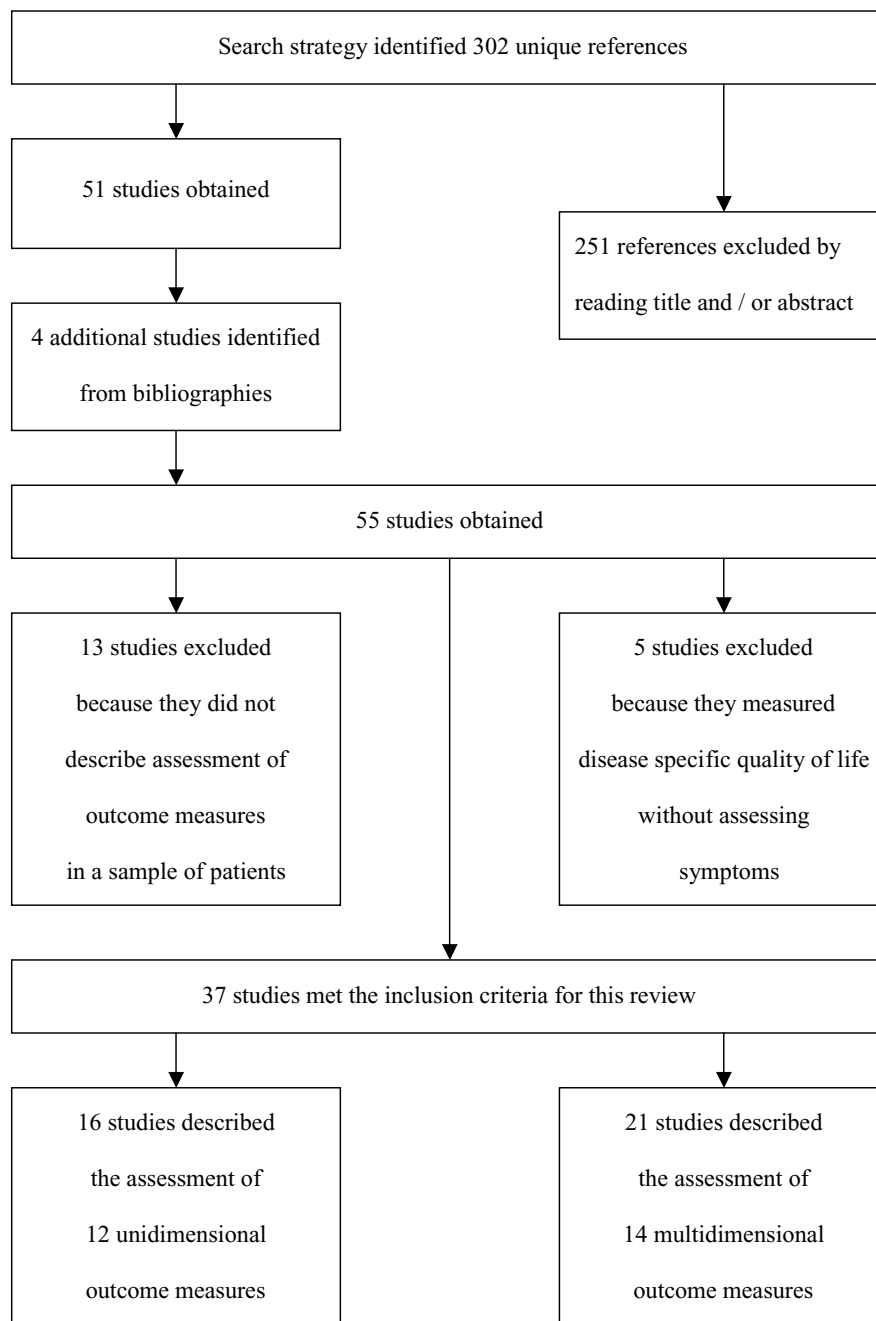
Patient diary cards and dyspepsia questionnaires that assessed global scores or quality of life without measuring symptoms were excluded. Uncertainties about the inclusion of an article were assessed by a second reviewer (BD).

A standard proforma was used by a single-trained reviewer (AF) to extract data on the study details, questionnaire characteristics (name, aims, language, method of administration, and questions asked), population studied, testing for reliability, validity and responsiveness (methods used, statistics, and significance), and conclusions. Studies were divided into those that assessed unidimensional scales (GERD or dyspepsia symptoms only) or multidimensional scales (several domains assessed, including GERD or dyspepsia symptoms). No statistical testing of the results was possible.

## RESULTS

The literature searches identified 302 citations. The title or abstract was used to exclude 251 citations and complete articles were obtained for the remaining 51 citations. A further four studies were identified from reference lists. Eighteen studies were excluded, leaving 37 eligible studies that described the assessment of 26 disease-specific questionnaires for GERD or dyspepsia (Fig. 1).

Thirteen studies were excluded because although they referred to outcome measures, they did not describe validation of these measures in a sample of patients (1, 2, 5, 6, 16, 24–30). Five studies were excluded because they examined three questionnaires that measured disease-specific quality of life alone without assessing the frequency or severity of symptoms. These questionnaires were the Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire (32, 33), Quality of life in Peptic Disease (QPD) questionnaire (34), and the Functional Digestive Disorders Quality of Life (FD-DQL) questionnaire (35, 36).



**Figure 1.** Flow diagram showing the studies included in the literature review.

No studies were found that compared outcome measures to determine which was the most reliable, valid, and responsive. No direct comparison between questionnaire characteristics was possible because the methods used to assess them varied considerably. Therefore, the questionnaires identified are described individually.

### UNIDIMENSIONAL QUESTIONNAIRES

Sixteen studies described the validation of 12 unidimensional outcome measures, which assess GERD or dyspepsia symp-

oms only (Table 1). Three of these (37–39) measured both frequency and severity of symptoms, which improves the accuracy of assessment compared to either frequency or severity alone (18, 19). Five outcome measures (40–44) measured symptom severity alone. The remaining four questionnaires (45–49) did not use standard methods to assess reliability, validity, and responsiveness.

### *Outcome Measures Assessing Frequency and Severity of Symptoms*

The Reflux Disease Diagnostic Questionnaire (RDQ) was designed to diagnose gastroesophageal reflux in primary care,

**Table 1.** Unidimensional Outcome Measures Assessing GERD or Dyspepsia Symptoms

First Author (Date)	Questionnaire	Reliability	Validity	Responsiveness	Completion	Comments
Hu <i>et al.</i> (2002) (43)	Hong Kong index of dyspepsia	Yes	Yes	Yes	Self	Validated in Chinese population Frequency of individual symptoms not assessed
Bisbal-Murrugarra <i>et al.</i> (2002) (45)	Questionnaire for the diagnosis of dyspepsia	No	No	No	Self	Promising items Poor study methodology
Shaw <i>et al.</i> (2001) (39)	Reflux Disease Diagnostic Questionnaire (RDQ)	Yes	Yes	Yes	Self	Presence and severity of reflux symptoms assessed
Talley <i>et al.</i> (2001) (40)	Gastrointestinal Symptom Rating Scale (GSRS)	Yes	Yes	Yes	Self	Dyspepsia questions not validated Used for IBS and dyspepsia
Junghard <i>et al.</i> (1998) (53)						Subjective response categories
Dimenas <i>et al.</i> (1993) (48)						Frequency of individual symptoms not assessed
Svedlund <i>et al.</i> (1988) (52)						Used for IBS, GERD, and dyspepsia
Shaw <i>et al.</i> (2001)(54)	Digestive Health Status Instrument (DHSI)	Yes	Yes	Yes	Self	Responsiveness assessed for GERD, not IBS or dyspepsia
Shaw <i>et al.</i> (1998) (41)						Subjective response categories
Leidy <i>et al.</i> (2000) (42)	Dyspepsia Symptom Severity Index (DSSI)	Yes	Yes	No	Self	Frequency of individual symptoms not assessed
Moayyedi <i>et al.</i> (1998) (37)	Leeds Dyspepsia Questionnaire (LDQ)	Yes	Yes	Yes	Res*	Presence and severity of dyspepsia symptoms assessed
Buckley <i>et al.</i> (1997) (38)	Dyspepsia Symptom Score (Ireland)	Yes	-	Yes	Res*	Presence and severity of dyspepsia symptoms assessed
Kennedy and Jones (1995) (46)	Postal health status questionnaire (UK)	Yes	No	No	Self	Small sample Validity not established
Hobbs <i>et al.</i> (1996) (47)	Adjectival scale symptom questionnaire (UK)	No	No	Yes	Self	Diagnostic instrument for dyspepsia Validity not established
Veldhuyzen van Zanten (1993) (44)	Adjectival scale dyspepsia questionnaire (Canada)	Yes	Yes	Yes	Self	Revised questionnaire not validated before use in clinical trial
Dimenas <i>et al.</i> (1993) (48)	Ulcer Esophagitis Subjective Symptoms Scale (UESS)	No	-	No	Self	Frequency of individual symptoms not assessed
Greatorex and Thorpe (1983) (49)	Esophageal Symptom Questionnaire	No	Yes	No	Res*	Visual analogue scale Validity not established Diagnostic instrument for GERD

Listed in date order, most recent first.

A dash (-) for reliability, validity, or responsiveness represents inadequate research methods or ambiguous results. \* Res = completed by researcher.

by assessing the frequency and severity of symptoms over 4 wk (39). Extensive piloting and scale development were used to minimize ambiguity and reduce the questionnaire to 12 items in three subscales. The regurgitation and heartburn subscales were shown to be reliable, valid, and responsive to change, but the validity of the dyspepsia subscale was not assessed. The RDQ was shown to be an excellent outcome measure for gastroesophageal reflux, but does not assess dyspepsia.

The Leeds Dyspepsia Questionnaire (LDQ) (37) was shown to be reliable, valid, and responsive to change in both primary and secondary care populations in the United Kingdom. It assessed the frequency and severity of eight symptoms over the previous 6 months using adjectival scale responses. The scoring system used the frequency stems of the first 5 items to determine the presence of dyspepsia, and the symptom severity of all 8 items to determine the severity of dyspepsia. The disadvantages were that the LDQ was administered by a researcher, assessed symptoms over a long time frame, and had a complex scoring system.

The questionnaire described by Buckley *et al.* assessed the frequency, severity, and duration of four common dyspepsia symptoms (38). Reliability and responsiveness were confirmed in a small secondary care sample, but assessment of validity compared with a gold standard was not performed. It took an average of only 3.6 min to complete, making it a convenient instrument (50). Disadvantages were that its validity was not established, it required administration by a researcher, and was tested in a small sample.

#### **Outcome Measures Assessing Severity of Symptoms Only**

The Gastrointestinal Symptom Rating Scale (GSRS) (40), Digestive Health Status Instrument (DHSI) (41), and the Dyspepsia Symptom Severity Index (DSSI) (42) used graded response categories from “none” to “very severe” without defining what these adjectives meant. This can produce subjective answers, reducing reliability and validity (51). The Hong Kong index of dyspepsia (43) and the questionnaire created by Veldhuyzen van Zanten *et al.* (44) both used descriptions for each response category.

The GSRS was developed in the early 1980s as an outcome measure for irritable bowel syndrome and peptic ulcer disease (52). It comprises 15 items addressing five symptom clusters (gastroesophageal reflux, abdominal pain, indigestion, diarrhea, and constipation). The GSRS was originally interview-based with a 4-point scale for each symptom (52), but was later modified to be self-completed with responses graded on a 7-point adjectival scale (48). It demonstrated reasonable validity and responsiveness, but relatively poor reliability (40), perhaps due to the absence of well-defined response categories. However, the GSRS was found to be more reliable, valid, and responsive than patient diary cards by Junghard *et al.* (53).

Shaw *et al.* reported the development of the 34-item DHSI to measure gastrointestinal symptoms based on the Rome criteria for dyspepsia, and the Manning and Rome criteria for

irritable bowel syndrome. Preliminary testing was performed and the DHSI was refined (41). The DHSI has demonstrated good internal consistency, reliability, and validity (54). Responsiveness to treatment for GERD was demonstrated during a clinical trial (54), but responsiveness to dyspepsia and irritable bowel syndrome has not yet been assessed.

The DSSI used twenty symptoms to assess subscales of dyspepsia (dysmotility-like, reflux-like, or ulcer-like) on a 5-point adjectival scale (42). It was shown to be reliable and valid, although responsiveness was not assessed.

The Hong Kong index of dyspepsia (43) originally contained 24 upper gastrointestinal symptoms, but logistic regression was used to select the 12 most discriminating symptoms. It was found to have good test–retest reliability and demonstrated validity compared with a generic quality-of-life instrument. Responsiveness was demonstrated by comparison with those who subjectively reported an improvement after 3 wk, which has been described as an unsatisfactory method of assessing symptom change (18). The Hong Kong index of dyspepsia was the only dyspepsia questionnaire validated in a Chinese population.

Veldhuyzen van Zanten *et al.* constructed a reliable, valid, and responsive questionnaire using adjectival responses to eight common symptoms (44). The researchers suggested that the four most important symptoms could be used by themselves as an outcome measure, rather than using the whole questionnaire.

#### **Studies that Did Not Adequately Establish Reliability, Validity, or Responsiveness**

Four other studies have described various outcome measures for GERD and dyspepsia symptoms, all of which did not adequately establish reliability, validity, or responsiveness. Bisbal-Murrugarra *et al.* (45) designed a questionnaire consisting of nine symptoms of dyspepsia, graded on a 4-point scale depending on the degree of “bothersomeness.” Although this scale included relevant symptoms, the study did not use any standard methods for assessing the characteristics of patient-based outcome measures.

Kennedy and Jones (46) developed a diagnostic postal questionnaire to identify the prevalence of dyspepsia in the general population, using the presence of eight symptoms over the past year, previous investigations, and consultation behavior. It was found to have good test–retest reliability. Validation against telephone interviews was conducted, but no statistical testing of validity was performed (46). Responsiveness was not assessed. Hobbs *et al.* (47) used 4 items from the questionnaire by Kennedy and Jones in a clinical trial of *Helicobacter pylori* eradication. This adjectival scale symptom questionnaire was found to be responsive, but reliability and validity were not assessed (47). Reliability of this questionnaire cannot be assumed, because modification will alter a questionnaire’s characteristics (51, 55, 56).

Dimenas *et al.* (48) developed the Ulcus Esophagitis Subjective Symptoms Scale to quantify nine symptoms using a 100-mm visual analogue scale. The reliability of the scale

was not assessed and there was only a moderate correlation with a generic quality-of-life instrument to assess validity. The use of visual analogue scales is unlikely to confer any benefits over adjectival scales (18, 24).

Greatorax and Thorpe (49) developed the esophageal symptom questionnaire to assist the diagnosis of gastroesophageal reflux disease. It contained 6 items, which used adjectival scales. They found that the questionnaire was valid, but reliability and responsiveness were not assessed.

## MULTIDIMENSIONAL QUESTIONNAIRES

Twenty-one studies were identified describing the validation of 14 questionnaires, which assessed the frequency or severity of GERD or dyspepsia symptoms as well as other dimensions of symptom-related quality of life (Table 2). Four questionnaires (57–60) assessed symptoms specifically relating to gastroesophageal reflux disease (GERD).

Two questionnaires were developed and validated exclusively in non-English-speaking populations—the Spanish language dyspepsia questionnaire (SLDQ) (61) and the German gastrointestinal quality-of-life index (GIQLI) (62). Two were designed as diagnostic instruments for dyspepsia rather than outcome measures—the domestic/international gastroenterology surveillance study (DIGEST) (3, 63) and the bowel disease questionnaire (64, 65). The remaining six questionnaires were designed as outcome measures for dyspepsia in an English-speaking population—the Nepean Dyspepsia Index (66, 67), Short-Form Nepean Dyspepsia Index (68), Severity of Dyspepsia Assessment (SODA) (69), Clinical Dyspepsia Questionnaire (70), questions from the UCLA questionnaire (71), and the Glasgow Dyspepsia Severity Score (31). The Nepean Dyspepsia Index has been translated into French, Dutch, Italian, German, Spanish, and American English (66).

### *Dyspepsia Outcome Measures Tested in English-Speaking Populations*

The Nepean Dyspepsia Index was developed primarily as a disease-specific quality-of-life measure (66–68). It also contains a symptoms checklist, which measures the frequency, intensity, and bothersomeness of 15 upper gastrointestinal symptoms during a 2-wk period. The 42 item quality-of-life component has undergone extensive development, and has demonstrated good internal consistency (66, 67), reliability (66), validity (66–68), and responsiveness (67, 68). The reliability, discriminant validity, and responsiveness of the symptom checklist have also been established, although its validity has not been assessed by comparison with another instrument (66–68). The Short-Form Nepean Dyspepsia Index (67, 68) has been developed from this instrument, and contains 10 quality-of-life items in five domains without a symptom component (67).

The Severity of Dyspepsia Assessment (SODA) (69) assessed symptoms in three dimensions—pain intensity,

nonpain symptoms, and satisfaction with dyspepsia-related health. Each dimension had good internal consistency, but test–retest reliability was low, probably due to a long test–retest interval of 4 wk (69). The validity and responsiveness of this instrument were assessed (72). SODA has undergone extensive evaluation during development, including analysis of floor and ceiling effects (73), the scoring of the questionnaire (74), and the number of items and response categories (73). The scoring of the SODA questionnaire has been transformed into an equal interval scale using statistical techniques (73), which may improve its responsiveness to change (75). However, SODA assesses the severity of symptoms but not their frequency.

The Clinical Dyspepsia Questionnaire was based on questions used during a clinical history of dyspepsia (70). It assesses frequency and severity of symptoms as well as their impact on quality of life. Test–retest reliability was relatively low, perhaps because a long test–retest interval of 2 wk was used. Validity was examined by correlation with a generic quality-of-life scale (the SF-36), but yielded only modest correlations. Further testing of the reliability, validity, and responsiveness of this tool would be necessary before its use as a clinical outcome measure.

Poitras *et al.* (71) described the use of symptom severity and quality-of-life items from the previously unreported UCLA questionnaire as an outcome measure in a clinical trial. This study was a clinical trial with a secondary outcome to test the questionnaire, so the utility of this measure was not rigorously assessed.

The Glasgow Dyspepsia Severity Score (GDSS) (31) is a simple, short dyspepsia outcome measure that has proven reliability, validity, and responsiveness and is often referred to in the literature (16, 20, 76). It has also been validated for use over the phone (77), making it a feasible and acceptable tool. However, five of the seven questions concern health-care utilization rather than symptoms or quality of life (31). This is a problem if a symptom status measure is needed for cost effectiveness trials, since health-care utilization cannot be used in both the “cost” and “effect” dimensions. The primary outcome in this case should be assessment of patients’ symptoms alone (17, 18), but the GDSS does not allow measurement of symptoms without collecting health-care utilization data.

### *Dyspepsia Outcome Measures Tested in Non-English-Speaking Populations*

Goldman *et al.* (61) developed and validated a Spanish Language Dyspepsia Questionnaire (SLDQ) for use in Mexican populations. Due to low literacy levels, the SLDQ was administered by a researcher. It included several items to check the understanding of common symptoms due to cultural differences. It included questions on symptom frequency and severity as well as quality of life, grouped into seven different areas, making it a long questionnaire. It was found to be reliable and valid, but responsiveness was not assessed and it has not been tested in an English-speaking population.

**Table 2.** Multidimensional Outcome Measures Assessing GERD and Dyspepsia Symptoms

First Author (Date)	Questionnaire	Reliability	Validity	Responsiveness	Completion	Comments
Rabeneck <i>et al.</i> (2002) (69)	Severity of Dyspepsia Assessment (SODA)	-	Yes	Yes	Self	Assessed symptom severity only Equal interval scales created Reliability not adequately established
Rabeneck <i>et al.</i> (2001) (72)						
Cook <i>et al.</i> (1999) (73)						
Kuykendall <i>et al.</i> (1998) (74)	Spanish Language Dyspepsia Questionnaire (SLDQ)	Yes	Yes	No	Res*	Validated in Spanish Included checking understanding of common dyspepsia symptoms
Goldman <i>et al.</i> (2002) (61)						
Poitras <i>et al.</i> (2002) (71)	Questions selected from the UCLA symptom questionnaire	No	No	No	Self	Symptom severity and quality of life assessed
Talley <i>et al.</i> (1999) (66)	Nepean Dyspepsia Index	Yes	Yes	Yes	Self	Poor study methodology Separate symptom and quality of life components of questionnaire
Talley <i>et al.</i> (1999) (67)						
Talley <i>et al.</i> (2001) (68)	Short-form Nepean Dyspepsia Index	No	Yes	Yes	Self	Short-form assesses quality of life only
Rothman <i>et al.</i> (2001) (57)	Gastroesophageal reflux disease Symptom Assessment Scale (GSAS)	Yes	Yes	Yes	Self	Postal questionnaire Highly selected study population
Calvet <i>et al.</i> (2000) (77)	Glasgow Dyspepsia Severity Score (GDSS)	Yes	Yes	Yes	Res*	Assessed reflux only Global ratings of symptom used
El-Omar <i>et al.</i> (1996) (31)						Questions on health-care costs Validated by phone
Eggleston <i>et al.</i> (1999) (63)	Domestic/International Gastroenterology Surveillance Study	Yes	Yes	No	Res*	Designed to diagnose dyspepsia
Enck <i>et al.</i> (1999) (3)	(DIGEST) questionnaire					Collected information on social and economic impact
Carlsson <i>et al.</i> (1998) (58)	"Patient Questionnaire" for Gastroesophageal Reflux Disease	No	-	No	Self	Designed to diagnose reflux Poor specificity
Mathias <i>et al.</i> (1996) (59)	Health-Related Quality-of-Life (HRQoL) questionnaire	No	No	Yes	Self	Used "word picture" of symptoms Used in clinical trial without prior validation
Garratt <i>et al.</i> (1996) (70)	The Clinical Dyspepsia Questionnaire	Yes	Yes	No	Self	Assessed reflux only Frequency, severity, and quality of life assessed
Eypasch <i>et al.</i> (1995) (62)	Gastrointestinal Quality-of-Life Index (GIQLI)	Yes	Yes	Yes	Self	Validated in German Applicable to any GI disease
Locke <i>et al.</i> (1994) (60)	Questionnaire for Gastroesophageal Reflux Disease	Yes	Yes	No	Self	Designed for responsiveness to change Designed to diagnose reflux
Talley <i>et al.</i> (1990) (64)	The Bowel Disease Questionnaire (BDQ)	-	Yes	No	Self	Long questionnaire Applicable to any GI disease
Talley <i>et al.</i> (1989) (65)						Aimed to distinguish pathology, functional GI diseases, and psychosomatic conditions

Listed in date order, most recent first.

A dash (-) for reliability, validity, or responsiveness represents inadequate research methods or ambiguous results. \*Res = completed by researcher.

The Gastrointestinal Quality-of-Life Index (GIQLI) was designed by surgeons for use in any gastrointestinal condition to assess changes in symptoms over time (62). It was envisaged that the GIQLI would have clinical and research applications. Reliability, validity, and responsiveness were demonstrated in a German population, but it would need further evaluation in order to assess symptoms in an English-speaking population (62).

#### **Outcome Measures for Gastroesophageal Reflux Disease**

Two questionnaires, the “Patient Questionnaire” (58) and the “Questionnaire for Gastroesophageal Reflux Disease” (60) were designed and tested as diagnostic tools for GERD, making them less useful as outcome measures for clinical trials. The “Gastroesophageal reflux disease Symptom Assessment Scale” (GSAS) (57) and the Health-Related Quality-of-Life (HRQoL) questionnaire (59) were designed primarily as outcome measures.

The GSAS measured the frequency, severity, and level of distress caused by fifteen symptoms (57). Some of these symptoms (bloating, belching, and early satiety) were more predictive of dysmotility-like dyspepsia rather than reflux disease. The results of the validation study may have limited generalizability, since only 185 patients of 9,182 health survey participants were included. However, Rothman *et al.* (57) demonstrated that the GSAS was reliable, valid, and responsive to change when administered as a postal questionnaire.

The HRQoL questionnaire (59) was primarily a disease-specific quality-of-life instrument for GERD, which included several items to assess symptoms. It was tested in a clinical trial without prior validation, so it has been shown to be responsive without assessment of reliability and validity.

The “Patient Questionnaire” (58) had good face validity, since it incorporated “word pictures” using simple English to describe symptoms of GERD. However, it was validated by comparison with diagnosis by endoscopy and esophageal pH monitoring, and so had poor specificity because many patients with GERD have neither esophagitis nor abnormal esophageal pH. Reliability and responsiveness were not assessed, since it was designed as a diagnostic instrument for GERD.

The “Questionnaire for Gastroesophageal Reflux Disease” (60) was reliable and valid and included items on a range of symptoms of dyspepsia as well as GERD. Responsiveness was not assessed, since it was intended to be a diagnostic instrument. It was a long questionnaire containing 76 items, making it inconvenient for use in clinical trials.

#### **Questionnaires Designed as Diagnostic Instruments for Dyspepsia**

The Domestic/International Gastroenterology Surveillance Study (DIGEST) (2, 3, 6, 26, 29, 30, 63) examined the prevalence of upper GI symptoms and the impact of these symptoms on health-care utilization and quality of life in 5,581 people from 10 countries. DIGEST used a questionnaire that was specifically created for this purpose. The questionnaire

was shown to be reliable and valid, but responsiveness was not assessed (63). The questionnaire was too long for routine use in clinical trials, and included questions about demographics, socio-economic status, quality of life, and health-care utilization.

The Bowel Disease Questionnaire (64, 65) was designed as a diagnostic tool to distinguish between gastrointestinal pathology, functional gastrointestinal disorders, and psychosomatic conditions. It included a variety of different questions concerning symptoms, past history, health habits, and mental health and was compared with diagnostic evaluation by experienced clinicians. It showed reasonable validity, but reliability was only assessed in a small sample. Responsiveness was not assessed.

## **DISCUSSION**

Symptom assessment using questionnaires has been recommended as the primary outcome measure in clinical GERD (12) and dyspepsia trials (17, 18). Assessment of quality of life is also important, since this reflects the degree of distress experienced by patients (21). However, the impact on quality of life that a patient experiences as a result of their disease is more variable and less reliable than symptom assessment.

This review identified 26 disease-specific questionnaires for GERD or dyspepsia, which could be used as clinical trial outcome measures. Twelve were unidimensional outcome measures that assessed symptoms only, and 14 were multidimensional outcome measures that also assessed aspects of quality of life. This reflects the lack of consensus about which type of instrument is the most appropriate for clinical trials. Rothman *et al.* found that patients did not differentiate well between symptom severity and quality of life (57), indicating that there may be some redundancy if both of these aspects are measured. Multidimensional instruments tend to be longer, which reduces their acceptability and feasibility (22). However, multidimensional instruments may provide a more complete assessment of a patient’s condition, thereby increasing validity (51).

Symptoms can be assessed by measuring either frequency or severity. Frequency of symptoms correlates more closely with diagnosis of dyspepsia than severity, indicating that frequency may be more valid (78). In gastroesophageal reflux disease, severity of symptoms correlates more closely with esophagitis cure than frequency, indicating that severity may be more responsive to change (12). Measuring both frequency and severity of dyspepsia symptoms may improve both validity and responsiveness to change of an instrument compared with measuring either alone (12, 18, 19, 78).

Three of 12 unidimensional questionnaires assessed frequency and severity of symptoms—the Leeds dyspepsia questionnaire (LDQ) (37), Buckley’s dyspepsia symptom score (38), and the reflux disease diagnostic questionnaire (RDQ) (39). Eight of 12 multidimensional questionnaires assessed frequency and severity of symptoms (57, 59–61, 63,

64, 66–68, 70). Of these, two were diagnostic instruments (63, 64), and one was not tested in an English-speaking population (61). Of the remaining questionnaires, the clinical dyspepsia questionnaire (70) and the Nepean dyspepsia index (66–68) measured dyspepsia symptoms, and the “Gastroesophageal Reflux Disease Symptom Assessment Scale” (GSAS) (57), the Health-Related Quality-of-Life (HRQoL) questionnaire (59), and the “Questionnaire for Gastroesophageal Reflux Disease” (60) measured GERD symptoms.

Outcome measures for symptom assessment should be tested for reliability, validity, and responsiveness before they are used in dyspepsia trials (11, 17–19). However, there was evidence to support the reliability, validity, and responsiveness of only 10 of 26 questionnaires (31, 37, 40, 43, 44, 54, 57, 62, 66–68). This was partially because eight instruments (45, 46, 49, 58, 60, 62, 63, 64) were designed as diagnostic instruments rather than outcome measures, and so responsiveness to change was not tested. However, methodological problems were a factor in the remaining studies.

Most questionnaires were self-completed by patients, although six required a researcher to administer the instrument (31, 37, 38, 49, 61, 63). Patient self-completion increases the feasibility of a questionnaire and reduces social desirability bias, compared to those that require a researcher to administer them (11, 18, 22, 55).

The period of time that questions in an outcome measure refer to is the time frame. Determination of the correct time frame is a balance between reducing recall bias (requiring a shorter time frame) and maximizing the amount of time for which data is available (requiring a longer time frame). McColl found that 1 month was the maximum period over which patients could provide reliable data due to recall errors (19). The time frame of the questionnaires in this review ranged from 1 wk to 6 months.

A variety of outcome measures for GERD and dyspepsia clinical trials have been developed. It was not possible to compare these outcome measures directly due to methodological heterogeneity between studies. Further research is necessary to compare existing outcome measures using the same population to determine which are the most reliable, valid, and responsive instruments.

## ACKNOWLEDGMENTS

This work was undertaken as part of a Masters in Primary Care degree at the University of Birmingham.

Prof. Delaney is supported by an NHS R&D Primary Care Career Scientist Award.

---

**Reprint requests and correspondence:** Brendan C. Delaney, M.D., F.R.C.P., M.R.C.G.P., Department of Primary Care and General Practice, Primary Care Clinical Sciences Building, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

*Received January 26, 2004; accepted August 12, 2004.*

---

## REFERENCES

- Heading RC. Prevalence of upper gastrointestinal symptoms in the general population: A systematic review. *Scand J Gastroenterol Suppl* 1999;231:3–8.
- Stanghellini V. Three-month prevalence rates of gastrointestinal symptoms and the influence of demographic factors: Results from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol Suppl* 1999;231:20–8.
- Enck P, Dubois D, Marquis P. Quality of life in patients with upper gastrointestinal symptoms: Results from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol Suppl* 1999;231:48–54.
- Talley NJ. Dyspepsia: Management guidelines for the millennium. *Gut* 2002;50(suppl 4):iv72–iv78.
- Maconi G, Tosetti C, Stanghellini V, et al. Dyspeptic symptoms in primary care. An observational study in general practice. *Eur J Gastroenterol Hepatol* 2002;14:985–90.
- Haycox A, Einarson T, Eggleston A. The health economic impact of upper gastrointestinal symptoms in the general population: Results from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol Suppl* 1999;231:38–47.
- Moayyedi P, Mason J. Clinical and economic consequences of dyspepsia in the community. *Gut* 2002;50:10–12.
- Management of dyspepsia: Report of a working party. *Lancet* 1988;1:576–9.
- Talley Colin-Jones NJD, Koch KL, Koch M, et al. Functional dyspepsia: A classification with guidelines for diagnosis and management. *Gastroenterol Int* 1991;4:145–60.
- Talley NJ, Stanghellini V, Heading RC, et al. Functional gastroduodenal disorders. *Gut* 1999;45(Suppl 2):II37–II42.
- Veldhuyzen van Zanten SJ. A systematic overview (meta-analysis) of outcome measures in Helicobacter pylori gastritis trials and functional dyspepsia. *Scand J Gastroenterol Suppl* 1993;199:40–3.
- Sharma N, Donnellan C, Preston C, et al. A systematic review of symptomatic outcomes used in oesophagitis drug therapy trials. *Gut* 2004;53(Suppl 4):iv58–iv65.
- Chiba N, Thompson ABR, Barkun AN, et al. The Rome II definition of dyspepsia does not exclude patients with GERD in Primary Care. *Gastroenterology* 2003;124(S2):A1572.
- Bytzer P. Assessment of reflux symptom severity: Methodological options and their attributes. *Gut* 2004;53(Suppl 4):iv28–iv34.
- Delaney BC, Moayyedi P, Deeks J, et al. The management of dyspepsia: A systematic review. *Health Technol Assess* 2000;4:1–189.
- Sandha GS, Hunt RH, Veldhuyzen van Zanten SJ. A systematic overview of the use of diary cards, quality-of-life questionnaires, and psychometric tests in treatment trials of Helicobacter pylori-positive and -negative non-ulcer dyspepsia. *Scand J Gastroenterol* 1999;34:244–9.
- Veldhuyzen van Zanten SJ, 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;91:660–73.
- Talley NJ. A critique of therapeutic trials in Helicobacter pylori-positive functional dyspepsia. *Gastroenterology* 1994;106:1174–83.
- McColl E. Best practice in symptom assessment: A review. *Gut* 2004;53(Suppl 4):iv49–iv54.
- Veldhuyzen van Zanten SJ. Assessment of outcome in dyspepsia: Has progress been made? *Gut* 2002;50(Suppl 4):iv23–iv25.

21. Glise H, Hallerback B, Wiklund I. Quality of life: A reflection of symptoms and concerns. *Scand J Gastroenterol Suppl* 1996;221:14–7.
22. Fitzpatrick R, Davey C, Buxton MJ, et al. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:1–74.
23. Chiba N. Treat the patients' main dyspepsia complaint, not the ROME criteria. *Am J Gastroenterol* 2004;99:1059–62.
24. Johannessen T, Petersen H, Kristensen P, et al. The intensity and variability of symptoms in dyspepsia. *Scand J Prim Health Care* 1993;11:50–5.
25. Xuan J, Kirchdoerfer LJ, Boyer JG, et al. Effects of comorbidity on health-related quality-of-life scores: An analysis of clinical trial data. *Clin Ther* 1999;21:383–403.
26. Tougas G, Chen Y, Hwang P, et al. Prevalence and impact of upper gastrointestinal symptoms in the Canadian population: Findings from the DIGEST study. *Domestic/International Gastroenterology Surveillance Study*. *Am J Gastroenterol* 1999;94:2845–54.
27. Valle C, Broglia F, Pistorio A, et al. Prevalence and impact of symptoms suggestive of gastroesophageal reflux disease. *Dig Dis Sci* 1999;44:1848–52.
28. Talley NJ, Holtmann G, Agreus L, et al. Gastrointestinal symptoms and subjects cluster into distinct upper and lower groupings in the community: A four nations study. *Am J Gastroenterol* 2000;95:1439–47.
29. Stanghellini V. Relationship between upper gastrointestinal symptoms and lifestyle, psychosocial factors and comorbidity in the general population: Results from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol Suppl* 1999; 231:29–37.
30. Tijssen JG. Representativeness and response rates from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol Suppl* 1999;231:15–9.
31. 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.
32. Kleinman L, Leidy NK, Crawley J, et al. A comparative trial of paper-and-pencil versus computer administration of the Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire. *Med Care* 2001;39:181–9.
33. Wiklund IK, Junghard O, Grace E, et al. Quality of Life in Reflux and Dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD). *Eur J Surg Suppl* 1998;583:41–9.
34. Bamfi F, Olivieri A, Arpinelli F, et al. Measuring quality of life in dyspeptic patients: Development and validation of a new specific health status questionnaire: Final report from the Italian QPD project involving 4000 patients. *Am J Gastroenterol* 1999;94:730–8.
35. Chassany O, Marquis P, Scherrer B, et al. Validation of a specific quality of life questionnaire for functional digestive disorders. *Gut* 1999;44:527–33.
36. Irvine EJ. The quality of quality of life. *Gut* 1999;44:450–1.
37. Moayyedi P, Duffett S, Braunholtz D, et al. The Leeds Dyspepsia Questionnaire: A valid tool for measuring the presence and severity of dyspepsia. *Aliment Pharmacol Ther* 1998;12:1257–62.
38. Buckley MJ, Scanlon C, McGurgan P, et al. A validated dyspepsia symptom score. *Ital J Gastroenterol Hepatol* 1997;29:495–500.
39. Shaw MJ, Talley NJ, Beebe TJ, et al. Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. *Am J Gastroenterol* 2001;96:52–7.
40. Talley NJ, Fullerton S, Junghard O, et al. Quality of life in patients with endoscopy-negative heartburn: Reliability and sensitivity of disease-specific instruments. *Am J Gastroenterol* 2001;96:1998–2004.
41. Shaw M, Talley NJ, Adlis S, et al. Development of a digestive health status instrument: Tests of scaling assumptions, structure and reliability in a primary care population. *Aliment Pharmacol Ther* 1998;12:1067–78.
42. Leidy NK, Farup C, Rentz AM, et al. Patient-based assessment in dyspepsia: Development and validation of Dyspepsia Symptom Severity Index (DSSI). *Dig Dis Sci* 2000;45:1172–9.
43. Hu WH, Lam KF, Wong YH, et al. The Hong Kong index of dyspepsia: A validated symptom severity questionnaire for patients with dyspepsia. *J Gastroenterol Hepatol* 2002;17:545–51.
44. Veldhuyzen van Zanten SJ, Tytgat KM, Pollak PT, et al. Can severity of symptoms be used as an outcome measure in trials of non-ulcer dyspepsia and Helicobacter pylori associated gastritis? *J Clin Epidemiol* 1993;46:273–9.
45. Bisbal-Murrugarra O, Leon-Barua R, Berendson-Seminario R, et al. A new questionnaire for the diagnosis of dyspepsia. *Acta Gastroenterol Latinoam* 2002;32:25–8.
46. Kennedy T, Jones R. Development of a postal health status questionnaire to identify people with dyspepsia in the general population. *Scand J Prim Health Care* 1995;13:243–9.
47. Hobbs FD, Delaney BC, Rowsby M, et al. Effect of Helicobacter pylori eradication therapy on dyspeptic symptoms in primary care. *Fam Pract* 1996;13:225–8.
48. Dimenas E, Glise H, Hallerback 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.
49. Greatorex R, Thorpe JA. Clinical assessment of gastroesophageal reflux by questionnaire. *Br J Clin Pract* 1983;37:133–5.
50. Thomson AB. Symptom evaluation for dyspepsia—Promises and illusions. *Ital J Gastroenterol Hepatol* 1997;29:487–90.
51. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press, 1995.
52. Svedlund J, Sjodin 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.
53. Junghard O, Lauritsen K, Talley NJ, et al. Validation of seven graded diary cards for severity of dyspeptic symptoms in patients with non ulcer dyspepsia. *Eur J Surg Suppl* 1998;583:106–11.
54. Shaw MJ, Beebe TJ, Adlis SA, et al. Reliability and validity of the digestive health status instrument in samples of community, primary care, and gastroenterology patients. *Aliment Pharmacol Ther* 2001;15:981–7.
55. McColl E, Jacoby A, Thomas L, et al. Design and use of questionnaires: A review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess* 2001;5:1–256.
56. McColl E, Thomas R. The use and design of questionnaires. London: The Royal College of General Practitioners, 2000.
57. Rothman M, Farup C, Stewart W, et al. Symptoms associated with gastroesophageal reflux disease: Development of a questionnaire for use in clinical trials. *Dig Dis Sci* 2001;46:1540–9.
58. Carlsson R, Dent J, Bolling-Sternevald E, et al. The usefulness of a structured questionnaire in the assessment of symptomatic gastroesophageal reflux disease. *Scand J Gastroenterol* 1998;33:1023–9.

59. Mathias SD, Castell DO, Elkin EP, et al. Health-related quality of life of patients with acute erosive reflux esophagitis. *Dig Dis Sci* 1996;41:2123-9.
60. Locke GR, Talley NJ, Weaver AL, et al. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;69:539-47.
61. Goldman J, Conrad DF, Ley C, et al. Validation of Spanish language dyspepsia questionnaire. *Dig Dis Sci* 2002;47:624-40.
62. Eypasch E, Williams JI, Wood-Dauphinee S, et al. Gastrointestinal Quality of Life Index: Development, validation and application of a new instrument. *Br J Surg* 1995;82:216-22.
63. Eggleston A, Farup C, Meier R. The Domestic/International Gastroenterology Surveillance Study (DIGEST): Design, subjects and methods. *Scand J Gastroenterol Suppl* 1999;231:9-14.
64. Talley NJ, Phillips SF, Wiltgen CM, et al. Assessment of functional gastrointestinal disease: The bowel disease questionnaire. *Mayo Clin Proc* 1990;65:1456-79.
65. Talley NJ, Phillips SF, Melton J, III, et al. A patient questionnaire to identify bowel disease. *Ann Intern Med* 1989;111:671-4.
66. Talley NJ, Haque M, Wyeth JW, et al. Development of a new dyspepsia impact scale: The Nepean Dyspepsia Index. *Aliment Pharmacol Ther* 1999;13:225-35.
67. Talley NJ, Verlinden M, Jones M. Validity of a new quality of life scale for functional dyspepsia: A United States multicenter trial of the Nepean Dyspepsia Index. *Am J Gastroenterol* 1999;94:2390-7.
68. Talley NJ, Verlinden M, Jones M. Quality of life in functional dyspepsia: Responsiveness of the Nepean Dyspepsia Index and development of a new 10-item short form. *Aliment Pharmacol Ther* 2001;15:207-16.
69. Rabeneck L, Wristers K, Goldstein JL, et al. Reliability, validity, and responsiveness of severity of dyspepsia assessment (SODA) in a randomized clinical trial of a COX-2-specific inhibitor and traditional NSAID therapy. *Am J Gastroenterol* 2002;97:32-9.
70. Garratt AM, Ruta DA, Russell I, et al. Developing a condition-specific measure of health for patients with dyspepsia and ulcer-related symptoms. *J Clin Epidemiol* 1996;49:565-71.
71. Poitras MR, Verrier P, So C, et al. Group counseling psychotherapy for patients with functional gastrointestinal disorders: Development of new measures for symptom severity and quality of life. *Dig Dis Sci* 2002;47:1297-307.
72. Rabeneck L, Cook KF, Wristers K, et al. SODA (severity of dyspepsia assessment): A new effective outcome measure for dyspepsia-related health. *J Clin Epidemiol* 2001;54:755-65.
73. Cook KF, Rabeneck L, Campbell CJ, et al. Evaluation of a multidimensional measure of dyspepsia-related health for use in a randomized clinical trial. *J Clin Epidemiol* 1999;52:381-92.
74. Kuykendall DH, Rabeneck L, Campbell CJ, et al. Dyspepsia: How should we measure it? *J Clin Epidemiol* 1998;51:99-106.
75. Rabeneck L, Cook KF, Wray NP. Planning a randomized clinical trial in dyspepsia: Do we have a valid outcome measure? (abstract). *Gastroenterology* 1998;114:A35.
76. Talley NJ, Agreus I, Veldhuyzen vZ. Is the Glasgow Dyspepsia Severity Score a tool for the global measurement of dyspepsia? *Eur J Gastroenterol Hepatol* 1997;9:413-4.
77. Calvet X, Bustamante E, Montserrat A, et al. Validation of phone interview for follow-up in clinical trials on dyspepsia: Evaluation of the Glasgow Dyspepsia Severity Score and a Likert-scale symptoms test. *Eur J Gastroenterol Hepatol* 2000;12:949-53.
78. Agreus L. Natural history of dyspepsia. *Gut* 2002;50 (suppl 4):iv2-iv9.