Symptom evaluation in reflux disease: workshop background, processes, terminology, recommendations, and discussion outputs

J Dent, D Armstrong, B Delaney, P Moayyedi, N J Talley, N Vakil
....

There has been no published indepth systematic evaluation of the best approaches to symptom evaluation in gastro-oesophageal reflux disease (GORD). A two day international multidisciplinary workshop was therefore held in Marrakech, Morocco, in September 2002 to address these issues. The aim of the workshop was to critically review the data regarding the reliability, processes, and priorities for symptom evaluation in GORD patients. The workshop was designed to give outputs that could be readily reported and to arrive at specific recommendations on best practice in symptom evaluation in reflux disease.

....

SUMMARY

To date, there has been no published indepth systematic evaluation of the best approaches to symptom evaluation in gastro-oesophageal reflux disease (GORD). A two day international multidisciplinary workshop was therefore held in Marrakech, Morocco, in September 2002, to address this. The workshop focused on four key topics and the outcomes are reported here.

(1) Diagnostic use of symptoms.
(2) Assessment of reflux symptom severity.
(3) Quality of life.
(4) Patient expectations and satisfaction.

In addition, recommendations were made on the terminology to be used in this area.

GORD should be defined by the presence of reflux oesophagitis (Los Angeles grades A–D) and/or when it causes reflux symptoms that are sufficient to impair quality of life and/or when it is associated with a risk of long term complications. Moderate symptoms that occur more than once per week impair quality of life and can therefore be considered as GORD. The best method of eliciting the predominant symptom is with a technically adequate clinical interview. Consensus was reached that dysphagia should be investigated if the pattern and duration were appropriately characterised.

For short term therapy, absence of heartburn, self-assessed by the patient using a seven point modified Likert scale over a one week period, is considered to be an optimal objective end point. Regurgitation should also be monitored. Absence of symptoms is predictive of healed oesophagitis during short and long term therapy but there are few data on predictors of a symptomatic response to therapy.

Disease specific measures of quality of life, such as the quality of life in reflux and dyspepsia (QOLRAD) scale, are more responsive to change than generic measures such as SF-36 and EuroQol. However, generic measures do allow comparison with other diseases. Quality of life should at least be measured at the beginning and end of a trial, with at least annual measurement in long term trials.

Patient satisfaction depends on several factors, including the outcome of treatment, process of care, and the patient-doctor relationship. There are no validated instruments to measure patient satisfaction in this disease, and there is little information on patient expectations.

BACKGROUND TO THE WORKSHOP

The last 15–20 years have seen major advances in the understanding of GORD and major development of pharmacological, surgical, and luminally delivered physical therapies. There has been a commensurate burgeoning of clinical trials into reflux disease therapies, driven mainly by the very high prevalence of this problem in developed countries, and the rapid evolution of therapeutic options.

The overall quality of clinical trials into reflux disease has improved greatly in the last 15–20 years but there is still considerable potential for making such trials more authoritative and comparable. This potential now resides predominantly in approaches to symptom evaluation.

Symptom evaluation is a crucial aspect of both the routine clinical management and the clinical trialling of reflux disease. Diagnosis and pretreatment severity assessment rest heavily on symptom evaluation, as does assessment of the outcomes of therapy. With some therapies, effective screening for side effect symptoms is an important part of the measurement of therapeutic outcomes.

Despite the importance of symptom evaluation in reflux disease, there has been no published indepth systematic evaluation of the best
....

See end of article for authors' affiliations
....

Correspondence to:
Professor J Dent, Department of Gastroenterology, Hepatology, and General Medicine, Royal Adelaide Hospital, Adelaide, Australia; jdent@ mail.rah.sa.gov.au
....

Table 1 Grading of the nature of the evidence for each workshop proposition

Nature of evidence	Study design	Study execution	Consistency	Directness of evidence
A	Meta-analysis of RCTs* (for interventions) RCTs (for interventions) Non-randomised studies (for diagnosis and prognosis)	No important flaws	Consistent	Direct or strong indirect
B	Meta-analysis of RCTs or RCTs (for interventions)	Important flaw or inconsistent or weak indirect		
	Non-randomised studies (for diagnosis or prognosis)	Important flaw or inconsistent or weak indirect		
	Non-randomised controlled studies (for interventions)	No important flaws	Consistent	Direct or strong indirect
C	Non-randomised controlled studies (for interventions)	Important flaw or inconsistent or weak indirect		
	Meta-analyses or RCTs with a combination of important flaws and inconsistency and/or indirect evidence			
D	Other evidence (not expert opinion)			
E	Expert opinion			

*RCTs, randomised controlled trials.
Adapted from: Mason J, Eccles M. *Guideline Recommendation and Evidence Grading (GREG): a new grading method for Clinical Guideline Development Groups*. University of Newcastle upon Tyne: Centre for Health Services Research, 2003; report 109.
Exceptions that can alter the quality of grading: sparse data (few events); use of data not in its initial randomisation, or apparent publication bias, can lower the quality; a very strong association can raise the quality.
Coding notes: important flaws occur when the highest standards of research that could be achieved by a study are not applied; *consistency* occurs at two levels—design (consistent methods, patients, outcomes) and statistical (a test of homogeneity of a summary estimate when the level of design consistency is acceptable and meta-analysis appropriate); *directness*: ''direct evidence'': relevant patient benefits and harms are measured in studies; ''strong indirect'': the surrogate end point is strongly related to desirable end points, or that direct evidence is available for a sufficiently related patient group; ''weak indirect'': the relationship between the study outcomes and patient benefits or harms is insufficient.

approaches to this for routine practice and clinical trials. This lack of considered guidance is evident in the design of published clinical trials that use a wide range of terminologies and approaches to symptom evaluation. Not all of the methods used can be optimal for identification of suitable patients for enrolment in clinical studies and for assessment of therapeutic outcomes. The diversity of approaches used also makes it difficult or impossible to make detailed comparisons among trials, or to safely pool symptom data from different clinical trials.

The potential for further enhancement of the quality of clinical trials in reflux disease led to the holding of a two day international multidisciplinary workshop in Marrakech, Morocco, in September 2002, that is the subject of this report. This evaluated both the general principles and the particulars of symptom evaluation in reflux disease for routine clinical care and clinical trials. The aim of the workshop was to critically review the data regarding the reliability, processes, and priorities for symptom evaluation in GORD patients. The workshop was designed to give outputs that could be readily reported and to arrive at specific recommendations on best practice in symptom evaluation in reflux disease.

Workshop processes
Structure of the workshop
The workshop involved 28 participants from 10 countries, who were specialist gastroenterologists, primary care physicians, surgeons, and researchers who have a major involvement in managing and/or researching reflux disease, and/or researching more general fields of methodology relevant to the topic.

Following reviews of generic methodological issues, the workshop was divided into four sequential sessions, focusing on the following topics:

(1) diagnostic use of symptoms;
(2) assessment of reflux symptom severity;
(3) quality of life;
(4) patient expectations and satisfaction.

Each session opened with an overview of the clinical practicalities and a methodological review relevant to the session, followed by division of the participants into four concurrent discussion subgroups, and then a plenary discussion and voting session.

Propositions were prepared focusing on specific issues, and were assigned to individual participants to research in advance of the workshop. Participants presented a review of the data relevant to their assigned proposition in a discussion subgroup where the proposition was then discussed and voted on (see below). The conclusions of the subgroup on each of their assigned propositions were presented to the full workshop in the plenary session, further discussion occurred as necessary, and all participants then voted anonymously on the proposition, this time electronically (see below).

After these four main workshop sessions, a final session was devoted to discussion of the major workshop conclusions and outcomes, with voting on further propositions as necessary to clarify outstanding questions and issues.

Process, format, and reporting of voting
For each proposition, participants in the discussion subgroup agreed on the nature of evidence for the proposition, after discussion of the study design and execution, consistency of the findings, and directness of the evidence (table 1). They then voted on the strength of their recommendation of the proposition (table 2). Both the nature of evidence that had been agreed and the strength of the recommendation from the discussion subgroup were presented to the full workshop in the plenary session. After discussion in the plenary session, all participants voted on the strength of the recommendation (table 2).

The outcome of voting from the plenary session is provided in this manuscript for each proposition in each of the four different topic sessions. The proposition is given in bold italics, followed by the nature of evidence agreed by the subgroup, which is given in italics. The strength of recommendation is then given, expressed as the percentage of participants voting at each level. The level of recommendation receiving the largest vote is highlighted in bold.

Table 2 Levels of strength of recommendation used in voting on the workshop propositions
Strength of recommendation
Agree strongly
Agree with reservation
Disagree with reservation
Disagree strongly

Each proposition is followed by a discussion of the major points raised in both the subgroups and the plenary session, together with the relevant references. Additionally, editorial commentary and opinion is given from the editors of this report who were also the Core Group responsible for the planning of the workshop, including preparation of the propositions. In addition to the reporting of the voting, discussion, and editorial comment on the workshop propositions given in this manuscript, individual manuscripts authored by the presenters of methodological reviews within the workshop are given in the rest of the workshop report.

Considerations involved in judging the nature of evidence and strength of recommendation

In any evidence based guideline, there needs to be an objective and transparent process by which the strength of evidence and the strength of a recommendation are graded. The two are not directly linked, as there may be other powerful factors, such as affordability, lack of evidence of long term efficacy, or lack of international generalisability that may influence the interpretation of evidence. Workshop participants were reminded that, in general, the more explicit and detailed the questions asked about the validity of the evidence presented, the more that evidence should weigh in the decision. The evidence grading used in the workshop is an example of this more explicit testing, incorporating domains for design and execution of the study, the consistency of the effect, and allowing for non-randomised designs where these are the appropriate method.

A properly conducted randomised controlled trial is the best method of avoiding bias when comparing two interventions, but where prognosis or diagnosis is concerned, a well designed cohort study is the choice. Furthermore, where good evidence of an indirect nature (for example, for a proxy outcome and mortality) can be linked, grade A was also given (table 1). Flaws, lack of consistency, inappropriate outcome measures or settings, and lack of direct linkage drop the level of evidence. In general, sample size was explored using confidence intervals as a measure of precision but large numbers of small studies are more prone to publication bias and bias due to poor design; participants were cautioned against this.

Additional data analyses for the workshop

As part of their preparation for the workshop, participants were given access to the AstraZeneca reflux disease clinical trial database. Further exploratory analyses were made on this large database where relevant. Where the results of such analyses have been used as evidence, this is noted in the report. The findings of such secondary analyses are identified and referenced to the original study publication.

TERMINOLOGY ISSUES AND RECOMMENDATIONS

One output of the workshop was a recommendation for consistent use of simplified terminology for symptom evaluation in GORD. The terminology used in the area of symptom evaluation in reflux disease is highly varied, and in many cases is vague. A systematic review has been conducted of randomised clinical trials of medical therapies in reflux disease, which includes over 200 publications (see Sharma and colleagues[1] in this supplement *(page iv58–iv65)*). Analysis of these trials highlights some 20 different terms used to describe the absence of GORD symptoms, and several terms to describe a reduction in symptoms (table 3). Moreover, there is significant overlap in what these terms are considered to mean, ''symptom relief'', for example, being used variously to describe absence of symptoms as well as reduction in symptoms. There is also considerable heterogeneity in how terms are defined. Some of the terminology used involves poor use of the English language. For instance, some terms are used that are internally contradictory when their pure meanings are considered (for example, ''complete relief''). This is often because qualifiers are attached to words that are intended to convey absolute measures. Similarly, there is also variation in the terminology used to describe regurgitation (table 4).

The survey of the use of terminology highlights the need for consistent use of unambiguous terms that describe symptom status in reflux disease. There is a clear need for greater consistency and definition of terms, not least to facilitate comparisons among different studies. Use of the same terms, with the same meanings, would be a significant step forward. The time available for the workshop did not allow detailed consideration of these issues of terminology. The authors of this section of the report, who were also the Core Group responsible for its planning, therefore met after the workshop to consider issues of terminology and make recommendations that simplify, and define, the terminology to be used. These recommendations have been applied throughout this workshop report.

Recommendations for improved terminology
When a symptom is not present

''Absence'' is the recommended term to state that a symptom is not present. It is interchangeable with ''symptom free/free of symptoms''. These terms were selected on the basis of simplicity, minimal ambiguity, and being most readily understood in other languages. This terminology needs to be used in conjunction with specification of the symptom in question (for example, heartburn) and a qualifier of the timescale or duration (for example, ''absence of heartburn for

Table 3 Terms used to describe the absence or reduction of gastro-oesophageal reflux disease symptoms in studies assessed for the systematic review by Sharma and colleagues[1] in this supplement *(see page iv58–iv65)*	
Absence of symptoms	**Reduction of symptoms**
Symptom free	Symptoms improved
Completely symptom free	Symptom relief
Persistently symptom free	Clinical improvement
Totally symptom free	
Symptom relief	
Profound symptom relief	
Complete symptom relief	
Symptom resolution	
Complete symptom resolution	
Sustained symptom resolution	
Symptom remission	
Symptoms ceased	
Symptoms absent	
Symptoms eliminated	
Persistent absence of symptoms	
Disappearance of symptoms	
Asymptomatic	
Complete disappearance of symptoms	
No symptoms	
Completely gone	

Table 4 Terms used to describe the symptom of regurgitation in studies assessed for the systematic review by Sharma and colleagues[1] in this supplement *(see page iv58–iv65)*

Regurgitation
Regurgitation
Acid regurgitation
Gastro-oesophageal regurgitation
Acid eructation

the last seven days'', or ''heartburn free for the last seven days'', or ''free of heartburn for the last seven days''.

When a symptom persists to some degree

''Reduction/reduced'' is recommended to describe when a symptom has decreased but is still present to some extent. A baseline comparator is implicit in this. The term ''relief'' was rejected as it was seen to be too open to interpretation (see above). When ''relief'' is used to describe a reduction of a symptom, it is a term that implies a judgement by the patient as to whether the reduction reaches a threshold. This could vary among patients according to the patient's expectations and level of satisfaction with therapy, potentially suggesting that a patient is satisfied with their symptom status when, in fact, they are still suffering significant residual symptoms.

When a symptom worsens

''Increase'' is the recommended term to describe worsening of a symptom. This may occur, for example, when there is an increase in a specific symptom or symptoms as a side effect of therapy.

Specification of the symptom(s)

The symptom(s) in question should always be specified, with the four main symptoms being ''heartburn'', ''regurgitation'', ''epigastric pain'', and ''dysphagia'', along with the descriptors that describe presence/absence and intensity. Terms to describe regurgitation were considered, as there is some heterogeneity in this. ''Acid regurgitation'', ''gastro-oesophageal regurgitation'', and ''acid eructation'' were all rejected, not least because the content of what is regurgitated cannot be defined based on symptoms. The recommended term is thus ''regurgitation''. This needs to be clearly distinguished from ''water brash''.

The authors recognise the difficulties in communicating specific symptoms to patients and, for example, the value of ''word pictures'' in describing symptoms. This is addressed by Shaw[2] in this supplement *(see page iv25–iv27)*, and the workshop also identified the need for local definitions of specific symptoms, taking into account variations in the translation of words into other languages.

Timescale and duration

This is addressed elsewhere in the workshop report, and the actual timescale and duration measured will depend on the aims of the symptom assessment. However, timescale of symptom evaluation and/or duration of symptoms should always be clearly stated in any reporting of symptom assessment. For more discussion of the appropriate duration for assessment of symptom status, see McColl[3] in this supplement *(page iv49–iv54)*.

Use of response scales

Response scales—for example, in the measurement of symptom severity—are frequently referred to as Likert scales, and this term was used in workshop discussions. ''Likert scale'' is commonly used, for example, to describe symptom scales, such as ''none, mild, moderate, severe''. The original Likert scale (see Wyrwich and Staebler Tardino,[4] in this supplement *(page iv45–iv48)*, for more on Likert scales) was a five point scale, from ''strongly approve'' to ''strongly disapprove'' with a defined (''no opinion'') zero point. In other words, a strict definition of a Likert scale could include a neutral midpoint with the scale symmetrical about this point with positive and negative loadings. However, in practice, the issue of a zero midpoint (and thus an odd number of responses) comes down to the intention of the author, with arguments for and against the use of a neutral alternative. To accommodate this, while at the same time encompassing common usage of Likert scales, the term ''modified Likert scale'' is used throughout the workshop report.

1. DIAGNOSTIC USE OF SYMPTOMS: PROPOSITIONS, VOTING, DISCUSSION, AND COMMENTARY

Introduction

Diagnostic symptom assessment is especially important in GORD where objective tests, such as endoscopy and oesophageal pH studies, are relatively insensitive. It is not enough however to document symptoms that may be important in GORD. Surveys suggest that approximately one third of the population experience heartburn,[5] and it is unlikely that all of these subjects have a disease. It is therefore important to determine the severity of symptoms and pathology that defines GORD. The next problem is establishing what symptom or symptoms are necessary for the diagnosis of GORD. Rome II focused on predominant heartburn as the feature that best identifies GORD, based on trial evidence that this group of patients with a normal endoscopy responds better to proton pump inhibitor (PPI) therapy than those with predominant dyspepsia or other symptoms.[6] It is important to establish the likely accuracy of this approach, how easy it is for patients to describe their predominant symptom, and what other features may enrich the diagnosis. This will be useful for clinicians diagnosing GORD and for identifying patients that are suitable for GORD trials.

The observation that the risk of subjects with severe reflux symptoms developing oesophageal adenocarcinoma was increased 40-fold[7] has raised the profile of GORD as a serious disease. The majority of patients with oesophageal adenocarcinoma have dysphagia as a symptom and most guidelines recommend that all patients with this symptom have urgent endoscopy. Dysphagia is also a common symptom of GORD and therefore the utility of investigating all patients with this problem needs to be questioned. These questions were addressed by the workshop under the following six topic headings:

- When does gastro-oesophageal reflux become GORD?
- What is the positive predictive value of predominant heartburn in diagnosing GORD?
- What is the negative predictive value of absence of predominant heartburn in excluding GORD?
- Does the practice setting in which the diagnostic assessment is made influence the positive predictive value of GORD symptoms?
- Are there any other symptoms that help make the diagnosis of GORD?
- What symptoms and characteristics identify reflux disease patients who are at relatively high risk of having or developing serious complications from this?

Propositions, voting, and discussion

When does gastro-oesophageal reflux become GORD?

(1.1) *GORD is defined by the presence of reflux oesophagitis (Los Angeles grades A–D) and/or when it causes reflux symptoms that are sufficient to impair quality of life and/or when it is associated with a risk of long term complications.*

Strength of recommendation: **agree strongly, 62%**; *agree, reservation, 27%; disagree, reservation, 12%; disagree strongly, 0%.*

This is consistent with the definition of GORD in the Genval Workshop Report.[8] When symptoms significantly impact on patient quality of life, this fits with holistic definitions of disease. The question of what level of reflux symptom load impacts on quality of life is addressed in proposition 2.2. The workshop recommended that this should now be adopted as the working definition of GORD.

Editorial comment. "Risk of long term complications" includes any severity of mucosal breakage due to reflux oesophagitis, as well as stricture and development of Barrett's oesophagus. However, to clarify this, during the workshop, the definition was modified to include explicit mention of reflux oesophagitis, as defined by the Los Angeles classification system.[9] This is the final definition (above), which was strongly supported.

(1.2) *In patients consulting with GORD, moderate symptoms and/or symptoms occurring two or more days per week significantly impair quality of life.* (nature of evidence: B).

Strength of recommendation: agree strongly, 7%; **agree, reservation, 63%**; *disagree, reservation, 30%; disagree strongly, 0%.*

The question of the threshold level of symptoms that causes clinically relevant impairment of quality of life is influenced to some extent by how quality of life is measured, and evidence from studies in patients with more severe symptoms is limited by significant selection bias. The Genval report proposed that reflux disease is likely to be present when heartburn occurs on 2 or more days a week, on the basis of the negative impact of this symptom frequency on quality of life.[8] It was noted in the report however that the evidence at that time may have been insufficient to define severity of reflux induced symptoms by frequency alone. New data now support this hypothesis, with a marked fall in patients' willingness to accept two or more days of mild heartburn per week (fig 1). Over 90% of patients accepted up to one day of mild heartburn during treatment as sufficient control of their heartburn but this fell to 32% when they experienced mild heartburn on 2–4 days of the week.[10] These data seem to justify the "two or more days a week" cutoff for symptoms that impair quality of life. The concept that frequency of symptoms is correlated with severity is supported by new analyses of data from patients with
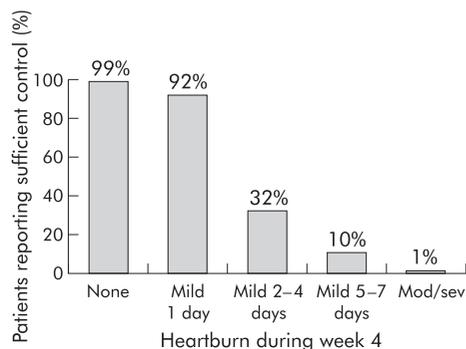

Figure 1 Most patients accept up to one day with mild heartburn per week during treatment. Few accept more frequent heartburn or moderate/severe heartburn.[10]
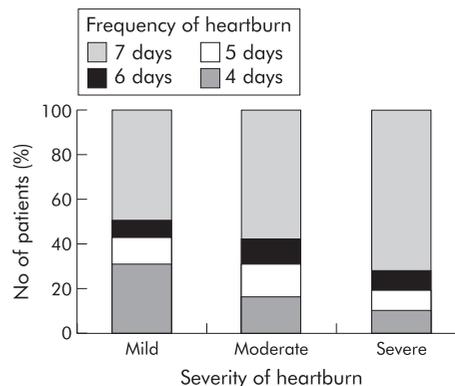

Figure 2 Patients with severe heartburn are likely to have more frequent heartburn than those with mild heartburn (AstraZeneca, data on file).

endoscopy negative GORD. These data show that patients with severe heartburn are more likely to experience daily heartburn than those with mild heartburn (fig 2) (AstraZeneca, data on file). Further support comes from the ProGERD study in over 5000 patients presenting with symptoms of GORD. In this study, both the SF-36 and QOLRAD scales show a decrease in quality of life dimensions with increasing frequency and severity of heartburn and, more particularly, indicate that the drop in quality of life is apparent when symptoms occur on more than one day of the week, or when they are of moderate or greater severity (AstraZeneca, data on file).

Editorial comment. While in general two or more symptoms per week is associated with impaired quality of life, clinical experience suggests there is the occasional patient that complains of infrequent but moderate or severe symptoms that are sufficiently troublesome to affect quality of life. It should also be noted that a minority of patients with more than twice weekly heartburn does not have impaired quality of life. This appears to be why 30% of workshop participants disagreed with the proposition.

What is the positive predictive value of predominant heartburn in diagnosing GORD?

(1.3) *In at least 90% of people with heartburn as their sole symptom, this will be caused by gastro-oesophageal reflux* (nature of evidence: D).

Strength of recommendation: agree strongly, 0%; agree, reservation, 17%; disagree, reservation, 35%; **disagree strongly, 48%**.

This proposition was rejected because only a very small subset of patients has a sole symptom, and there are no data available for such patients. In a study of patients with predominant heartburn whose GORD was defined by pH monitoring,[11] the specificity of heartburn was 89%, but sensitivity was 38%. Patient numbers were small, heartburn was not the sole symptom, and pH monitoring is not sufficiently accurate.

Editorial comment. There are propositions throughout this workshop that the group rejected or accepted where there was little or no evidence. This decision is therefore driven by clinical opinion and, for this proposition, the implication is that the group felt heartburn alone does not have a superior positive predictive value to heartburn as a predominant symptom, which is addressed in proposition 1.4.

(1.4) *In at least 80% of patients consulting with heartburn as their predominant symptom, this is induced by gastro-oesophageal reflux* (nature of evidence: C).

*Strength of recommendation: agree strongly, 37%; **agree, reservation, 59%**; disagree, reservation, 4%; disagree strongly, 0%.*

Acceptance of this proposition relied heavily on clinical experience, as there are no direct data to support it. Indirect data are available by extrapolation from studies of empirical PPI therapy and GORD diagnosis in patients with heartburn as their predominant symptom.[12–15] Sensitivity was 75–83% and specificity 55–63%. The Genval workshop concluded that when heartburn is a major or sole symptom, gastro-oesophageal reflux is the cause in at least 75% of individuals, and this was, again, based on consistency of indirect evidence and clinical experience.[8] It should be noted that the value of 80% accepted in the proposition here relates to patients consulting in secondary care with heartburn, and may be expected to be lower in people with heartburn in the general population.

(1.5) *Heartburn is associated with epigastric pain in at least two thirds of patients with upper gastrointestinal symptoms* (nature of evidence: C).

*Strength of recommendation: agree strongly, 4%; **agree, reservation, 62%**; disagree, reservation, 15%; disagree strongly, 19%.*

The quite wide scatter of votes reflects the lack of studies of adequate design which specifically address the proposition, the fact that the relevant data available are largely secondary unpublished analyses, and that the definition, frequency, and severity scorings of heartburn and epigastric pain are generally not reported in detail. Consequently, there was debate over the precise proportion of patients with both symptoms. Support for the proposition comes from unselected subjects surveyed randomly in UK general practices.[16] Re-analysis of these data indicate that of the 3177/8350 individuals (38%) that had upper gastrointestinal symptoms, 2403/3177 (76%) had heartburn more than once per month, and 1518/2403 (63%) of these individuals had coexisting epigastric pain. The Canadian Cadet-PE study in patients with uninvestigated dyspepsia also supports the proposition. Of 84% of patients with upper gastrointestinal symptoms found to have heartburn or regurgitation, 75% also had ulcer-like dyspepsia (AstraZeneca, data on file). Reanalysis of the AstraZeneca database from a study comparing healing of reflux oesophagitis with esomeprazole and lansoprazole in over 5000 patients in the USA with reflux oesophagitis and heartburn[17] shows that 66% of these patients also had epigastric pain at baseline. In general practice in Denmark,[18] 32% of consecutive consulting dyspepsia patients had dominant heartburn and/or regurgitation, 37% had dominant epigastric pain, and 66% had both. In conclusion, the available data indicate that there is significant overlap between heartburn and epigastric pain, but controversy remains over the precise extent of this overlap.

(1.6) *The epigastric pain that occurs in patients with reflux disease is generated predominantly by oesophageal contact with refluxate* (nature of evidence: E).

*Strength of recommendation: agree strongly, 4%; **agree, reservation, 46%**; disagree, reservation, 42%; disagree strongly, 8%.*

The evenly split vote reflects the absence of data showing a temporal association between epigastric pain and reflux episodes. Evidence to support the proposition is limited and circumstantial. Epigastric pain is part of the symptom complex for many GORD patients, and improves with PPI therapy, although to a lesser extent than heartburn. Some functional dyspepsia patients with non-dominant heartburn have increased oesophageal acid exposure[19] but there are no prospective data to show that epigastric pain in GORD is triggered predominantly by gastro-oesophageal reflux. A single study has shown an association between oesophageal acidification by acid perfusion and epigastric pain, although this was in duodenal ulcer patients.[20]

(1.7) *Patients with heartburn and epigastric pain find it difficult to describe their predominant symptom* (nature of evidence: C).

*Strength of recommendation: agree strongly, 21%; **agree, reservation, 71%**; disagree, reservation, 8%; disagree strongly, 0%.*

No studies have directly addressed this although it is supported by indirect evidence.[11 21–23] Epigastric pain and heartburn frequently coexist in various populations with dyspepsia, and a significant proportion cannot select their predominant symptom. When primary care patients with dyspepsia were asked to select their predominant symptom from heartburn, regurgitation, or epigastric pain, 19% were unable to choose, 10% said that it was none of these, and 10% failed to respond.[21] This highlights two elements to the question: firstly, the patient's ability to differentiate epigastric pain from heartburn and, secondly, their ability to select a predominant symptom.

(1.8) *A word description helps patients decide whether heartburn or epigastric pain is the predominant symptom* (nature of evidence: C).

*Strength of recommendation: agree strongly, 17%; **agree, reservation, 75%**; disagree, reservation, 8%; disagree strongly, 0%.*

Evidence to support this comes from two studies that have documented the value of using word descriptions for heartburn that included upward movement of pain, discomfort, or a burning feeling starting in the epigastrium and rising towards the neck.[24 25]

Editorial comment. Phrasing of the word picture is important because if more than two concepts are combined in the same sentence this can be confusing for respondents.[21]

What is the negative predictive value of absence of predominant heartburn in excluding GORD?

(1.9) *In people with upper abdominal pain in whom heartburn occurs as a secondary symptom, GORD is present in approximately 30%* (nature of evidence: C).

*Strength of recommendation: agree strongly, 15%; **agree, reservation, 70%**; disagree, reservation, 10%; disagree strongly, 5%.*

Direct evidence is not available for the significance of heartburn as a secondary (that is, non-dominant) symptom. Using likelihood ratios on the available data,[26] the probability of GORD in the absence of predominant heartburn can be estimated to be 12%,[27] 30%,[25] and 34%,[11] based on 50% probability of having GORD in all patients attending a secondary care dyspepsia clinic. A study of patients with a primary symptom of dyspepsia and secondary heartburn suggests that at least 13% of patients had endoscopically confirmed reflux oesophagitis.[28] Assuming that patients with GORD have oesophagitis or endoscopy negative reflux disease in roughly equal proportions, then these data also support a value of 25–30% as an approximate estimate.

Does the practice setting in which the diagnostic assessment is made influence the positive predictive value of GORD symptoms?

(1.10) *Of patients who seek advice in primary care about predominant heartburn, less than 70% will have reflux disease* (nature of evidence: E).

*Strength of recommendation: agree strongly, 0%; **agree, reservation, 52%**; disagree, reservation, 32%; disagree strongly, 16%.*

Data in primary care on this proposition are lacking, and the lack of direct evidence is compounded by variation in how reflux disease is defined, with most data based on endoscopic detection of oesophagitis, on pH monitoring, or response to PPI therapy. Additionally, the term ''heartburn'' is not used consistently in different settings.

Editorial comment. The majority felt that the prevalence of GORD was likely to be lower in the primary care setting than

in a secondary care dyspepsia clinic. The positive predictive value of predominant heartburn will therefore also fall compared with that seen in secondary care studies. The value of 70% is not supported by direct data and is estimated using likelihood ratios and extrapolating from secondary care studies.[26]

Looking at the propositions above collectively raises some questions. The workshop participants agreed that in at least 80% of patients consulting with heartburn as their predominant symptom, this is induced by GORD (proposition 1.4). However, there is significant overlap between heartburn and epigastric pain (proposition 1.5), and participants agreed that patients with heartburn and epigastric pain find it difficult to describe their predominant symptom (proposition 1.7), although this can be helped by use of a word description (proposition 1.8). Opinion was divided on the proposition above (proposition 1.10) that of patients who seek advice in primary care about predominant heartburn, less than 70% will have reflux disease. The responses to these propositions raise the issue as to from what perspective "predominant" heartburn is defined. Is a patient's self-reporting of heartburn as the predominant symptom adequate, or are appropriately trained clinicians more accurate? To help clarify this, a further proposition (proposition 1.11) was developed during the plenary session, which seeks to define predominant heartburn.

(1.11) *Predominant heartburn is defined as the most bothersome symptom based on a physician interview*.

*Strength of recommendation: agree strongly, 31%; **agree, reservation, 58%**; disagree, reservation, 12%; disagree strongly, 0%.*

Taken together with propositions 3.13 and 3.14, a key recommendation from the workshop is that global clinical opinion, based on a technically adequate clinician interview, is the most accurate approach to the diagnosis of GORD, rather than relying on the patient's description of their predominant symptom. The positive predictive value of the symptom of predominant heartburn to detect GORD will still fall as the prevalence of GORD falls, even if a trained clinician assesses symptoms.

(1.12) *Less than 50% of people found to have heartburn of any severity by a population survey will have reflux disease* (nature of evidence: E).

*Strength of recommendation: agree strongly, 4%; **agree, reservation, 75%**; disagree, reservation, 17%; disagree strongly, 4%.*

There are few primary data to support this hypothesis and this is compounded by a lack of a gold standard test to diagnose GORD. It was felt that the prevalence of GORD was likely to be lower than seen in either primary or secondary care and that the positive predictive value of heartburn (particularly of any severity) will fall.

Editorial comment. The implication of this proposition is that population surveys reporting the prevalence of "heartburn" in a general population may not be identifying GORD as accurately as studies in secondary care patient populations. This does not invalidate these studies but suggests the results should be interpreted with caution. Many people may report having heartburn but it may not be related to gastro-oesophageal reflux or if it is, it may not reach the threshold of severity required to define "disease".

Are there any other symptoms that help make the diagnosis of GORD?
(1.13) *In people with predominant heartburn, this is more likely to be due to gastro-oesophageal reflux if regurgitation has also been noted* (nature of evidence: D).

*Strength of recommendation: agree strongly, 13%; agree, reservation, 35%; **disagree, reservation, 52%**; disagree strongly, 0%.*

The even split in the vote reflects the paucity of studies in this area. Acceptance was based on clinical experience, while disagreement was due to a lack of evidence. A single study has reported the positive predictive value of symptoms versus pH testing in patients with symptoms suggestive of reflux disease.[29] The positive predictive value of heartburn was 59%, rising slightly to 66% when regurgitation was also present, although it was also 66% for regurgitation alone. Similarly, the positive predictive values were 70% for both heartburn and regurgitation using oesophageal pH monitoring as the reference diagnostic test in patients with suspected reflux disease although this study does not provide data on the value of heartburn and regurgitation together.[11] Consideration of this question is confounded by the need for a universal definition of regurgitation, and there are anecdotal reports that interpretation of the term may not be the same in different languages. This may be enhanced by use of a word description, as is the case for more reliable recognition of heartburn.

Editorial comment. The uncertainty in this and other related questions emphasises the need for high quality, prospective, cross sectional surveys that carefully detail patients' symptoms and correlate this with the final diagnosis reached.

(1.14) *Occurrence of reflux symptoms for more than six months is a confirmatory feature of GORD* (nature of evidence: E).

*Strength of recommendation: agree strongly, 17%; **agree, reservation, 61%**; disagree, reservation, 17%; disagree strongly, 4%.*

Although there are no direct data to support duration of symptoms as being helpful in the diagnosis of GORD, acceptance by the majority of participants was based on the view that it makes clinical sense. Most clinical trials report the duration of GORD symptoms in terms of years rather than months, which is not helpful in assessing the proposition.

Editorial comment. Two case control studies have shown that increasing duration of reflux symptoms increases the risk of developing oesophageal adenocarcinoma.[7][30] This was taken as very indirect evidence supporting the proposition, under the assumption that part of the reason for this may be that chronic symptoms are more likely to be due to more severe GORD. Additionally, although no studies have directly addressed the proposition, it is in accordance with patients' experience, with most reporting symptoms for six months or more. The strength and consistency of the supporting evidence is probably underestimated.

(1.15) *In approximately 30% of patients with recurrent non-cardiac chest pain, this is caused by gastro-oesophageal reflux* (nature of evidence: B).

*Strength of recommendation: agree strongly, 21%; **agree, reservation, 75%**; disagree, reservation, 4%; disagree strongly, 0%.*

Recurrent angina-like chest pain is a symptom of GORD although GORD and chest pain are linked through intermediary mechanisms that interfere with establishing a cause-effect relationship. However, there was broad acceptance of gastro-oesophageal reflux as a cause of non-cardiac chest pain, and debate centred around the proportion of patients in which this is the case. Updating an analysis of cross sectional surveys in patients with non-cardiac chest pain[31] with two subsequent studies[32][33] shows that 200/947 non-cardiac chest pain patients (21%—from 14 studies) had endoscopically confirmed oesophagitis, 423/1002 (42%—from 14 studies) had abnormal acid exposure time, and 278/787 (39%—from 16 studies) had a positive association of chest pain with reflux episodes during pH monitoring. Evidence for an association between non-cardiac chest pain and GORD also comes from studies of the response of non-cardiac chest pain

to PPI therapy[33–35] and laparoscopic surgery,[36] although the size of response varies between studies. The available data indicate that the prevalence of reflux induced, provoked, or otherwise related pain in the non-cardiac chest pain population is substantial, possibly representing 30–50% of patients.

Editorial comment. A label of non-cardiac chest pain suggests the patient has had extensive cardiac investigations to exclude ischaemic heart disease. These patients, by definition therefore, are highly selected and there is likely to be further bias in those patients that are included in studies. While selection bias would lead to an overestimate of the proportion of patients in which GORD is the cause of non-cardiac chest pain, the data are likely to be valid for patients attending secondary care clinics.

What symptoms and characteristics identify reflux disease patients who are at relatively high risk of having or developing serious complications from this?

(1.16) *Presence of dysphagia of any pattern should not be considered an alarm symptom* (nature of evidence: C).

*Strength of recommendation: agree strongly, 8%; agree, reservation, 31%; disagree, reservation, 15%; **disagree strongly, 46%**.*

The particularly broad spread in voting highlights differences in interpretation of how to act on dysphagia of any severity, duration, or pattern of occurrence given the loose definitions of dysphagia used in the community. Some 78% of oesophageal cancer cases have dysphagia[37] but conversely, dysphagia is common in the community. A pooled analysis of six community surveys[16 27 32 38–40] done for the purpose of this workshop, involving 12 700 subjects, indicates a point prevalence of dysphagia in the community of 14%. Dysphagia is particularly common in patients with heartburn, increasing in incidence with increasing frequency of heartburn. In five large randomised controlled esomeprazole trials, involving approximately 12 000 reflux oesophagitis patients, 37% had dysphagia but there were no cases of oesophageal malignancy.[41] Thus while dysphagia, when correctly evaluated, may indicate a risk of oesophageal malignancy (see the next proposition, 1.17), its presence per se may not necessarily be regarded as an alarm symptom.

Editorial comment. The voting reflects concern that physicians may be too dismissive of dysphagia, and not sufficiently diligent regarding the duration and pattern of the symptom to decide whether or not it is an alarm symptom (see the next proposition 1.17). There is also a lack of consensus on how to elicit a symptom of dysphagia and how to assess its severity in clinical practice or research. When a patient reports dysphagia, either spontaneously or in response to a direct question from the physician, the physician should then go through a number of steps to filter those that require investigation. In other words, given the high level of self-reporting of dysphagia, it is appropriate that dysphagia of any pattern should not be considered an alarm symptom. Despite the vote of the workshop, the editorial group for the report believe that it is not appropriate to endoscope everyone who reports dysphagia, as it is very common, responds rapidly to treatment, and is not associated with a level of risk to justify screening. Lack of adequately researched evidence on how to identify patients in whom dysphagia is of concern warrants study but conventional clinical wisdom is that in patients with newly appearing dysphagia, increasing severity of this symptom, or persistence of dysphagia despite therapy, this demands investigation.

(1.17) *Dysphagia is a useful indicator of risk for oesophageal malignancy, provided its duration and pattern of occurrence are also evaluated* (nature of evidence: D).

*Strength of recommendation: agree strongly, 21%; **agree, reservation, 71%**; disagree, reservation, 8%; disagree strongly, 0%.*

Broad acceptance of this puts the previous proposition in perspective. Three cross-sectional surveys have evaluated the role of dysphagia as an alarm symptom, and although they have limitations, the data indicate that dysphagia is not a good predictor of the presence of cancer.[42–45] Despite this lack of evidence, and although there are no reports detailing the impact of dysphagia pattern and duration on risk of cancer, the proposition was accepted on the grounds of clinical experience. This assumes that the time of onset, progression, and associated features, such as weight loss and family history, are evaluated.

(1.18) *Patients with reflux disease for more than five years are at an increased risk of long segment Barrett's oesophagus compared with a control population* (nature of evidence: C).

*Strength of recommendation: agree strongly, 13%; **agree, reservation, 71%**; disagree, reservation, 17%; disagree strongly, 0%.*

The ideal study to support this proposition has not been done. Indirect supportive evidence comes from studies in patients with reflux symptoms and frequent antacid users,[46–49] and from studies reporting long duration of symptoms to be a risk factor for Barrett's oesophagus.[50 51] These data do not specifically highlight the five year timeframe but do suggest Barrett's oesophagus is associated with chronic reflux symptoms.

(1.19) *People who have had heartburn severe enough to be defined as causing reflux disease for more than five years are at an increased risk of oesophageal adenocarcinoma compared with a control population* (nature of evidence: B).

*Strength of recommendation: agree strongly, 37%; **agree, reservation, 56%**; disagree, reservation, 7%; disagree strongly, 0%.*

Three case control studies have demonstrated an association between oesophageal adenocarcinoma and increasing duration of GORD symptoms,[7 30 52] as well as increasing frequency of symptoms.[7 52] One study[7] used squamous oesophageal cancers as a second control group, so recall bias is unlikely to explain the association.

(1.20) *Classical Barrett's oesophagus is present in less than 1% of reflux disease patients younger than 50 years of age* (nature of evidence: C).

*Strength of recommendation: agree strongly, 12%; **agree, reservation, 76%**; disagree, reservation, 12%; disagree strongly, 0%.*

"Classical" indicates long segment Barrett's oesophagus that is at least 3 cm in length. Supportive data for the proposition are limited. A prevalence of Barrett's oesophagus of 5% has been reported in patients aged 40–49 years, rising to 10% in those aged 50–69 years, although the study population of relatives of patients with Barrett's oesophagus was highly selected.[51] Data from patients in the Mayo Clinic between 1976 and 1989 indicate that the prevalence of Barrett's oesophagus is 0.41% in patients aged 30–49 years and 1.61% in those aged 50–69 years.[53]

Editorial comment. The implication from the voting on this proposition is that young patients with reflux symptoms do not need endoscopy to exclude Barrett's oesophagus. It must be noted however that most of the data is from all patients endoscoped rather than patients specifically with reflux disease.

(1.21) *Of patients with oesophageal adenocarcinoma in developed countries, more than 95% are older than 50 years of age*. (nature of evidence: A).

*Strength of recommendation: **agree strongly, 92%**; agree, reservation , 8%; disagree, reservation, 0%; disagree strongly, 0%.*

There is geographical variation in the prevalence of oesophageal carcinoma but published data strongly support the proposition.[54–60]

(1.22) *Compared to females, males with Barrett's oesophagus have greater than twice the risk of developing adenocarcinoma* (nature of evidence: B).

Strength of recommendation: agree strongly, 48%; **agree, reservation, 52%**; *disagree, reservation, 0%; disagree strongly, 0%.*

There is clear evidence to support this[7] [54] [61–64] although there are caveats to the interpretation of the epidemiology of oesophageal adenocarcinoma. There are some uncertainties about the nature of adenocarcinoma of the gastro-oesophageal junction. Some reports combine junctional and oesophageal body cancer, and coding conventions for junctional and oesophageal adenocarcinoma were altered in the 1990s.[60] [65] The proposition relates to Barrett's oesophagus although it may understate the differential in risk between males and females in so far as the differential is probably greater for adenocarcinoma per se, and most cases of adenocarcinoma are not preceded by Barrett's oesophagus diagnosed at endoscopy.

(1.23) *Among patients over 50 years of age in primary care presenting with reflux symptoms for over five years, the yield of endoscopy in any year for detecting oesophageal adenocarcinoma is less than 1 in 1000* (nature of evidence: B).

Strength of recommendation: agree strongly, 46%; **agree, reservation, 50%**; *disagree, reservation, 4%; disagree strongly, 0%.*

This is supported by publications reporting a risk of oesophageal adenocarcinoma in Barrett's oesophagus of approximately 0.5% per annum,[66] coupled with a prevalence of Barrett's oesophagus in patients with reflux symptoms of approximately 10%. The data are limited by differing definitions of Barrett's oesophagus, inclusion and exclusion of short segment Barrett's oesophagus, and different definitions of landmark features, such as the gastro-oesophageal junction. A prevalence of long segment Barrett's of 7% was seen in one study of asymptomatic patients attending colorectal cancer screening but this was a study in US veterans and was not felt to be generalisable.[67]

Future directions for research

Statistical techniques have been developed that overcome many of the problems of not having a gold standard to diagnose reflux disease. Future studies should use these techniques to assess the accuracy of heartburn to diagnose GORD in secondary and primary care populations. A comprehensive history should be taken in these studies so that the additional value of regurgitation and duration of symptoms can be evaluated as well as assessing the overlap with epigastric pain.

Most guidelines recommend that patients with dysphagia should have endoscopy. This would entail endoscoping 14% of the entire population given the high prevalences of reflux disease and dysphagia. Direct assessment of the additional value of the duration and pattern of occurrence of dysphagia would be helpful in refining our perception of dysphagia and risk of neoplasia.

The risk of Barrett's oesophagus and oesophageal adenocarcinoma with increasing duration and severity of heartburn has now become established but this important finding is based on only a few studies. Further studies evaluating these associations in different populations would therefore be useful.

2. ASSESSMENT OF REFLUX SYMPTOM SEVERITY: PROPOSITIONS, VOTING, DISCUSSION, AND COMMENTARY
Introduction

Translation of the results of clinical research studies into clinical practice is a significant challenge. For the patient presenting in clinical practice, several questions need to be addressed in order to determine optimal therapy. Are symptoms reflux related and, if so, are they typical or atypical? Are symptoms mild or severe, and are they associated with reduced quality of life or oesophageal damage? If treatment is warranted, will it be effective in the short and long term? Should the response to treatment be complete or satisfactory to the patient? These may not be the same thing and, furthermore, it is not clear that the measure of response should be equivalent for all symptoms.

For initial therapy, there are two major strategies—start high with daily PPI or at a lower level with daily H_2 receptor antagonists. Experience with testing these options in the CADET-HR study underlines the difficulties of translating outcomes from clinical trials such as this into clinical practice. The different measures of symptom response used in this study gave different estimates of the extent of superiority of omeprazole over ranitidine.[68] Greatest differentiation was seen if absence of reflux induced symptoms was used but the absolute response rate was lower.

The same uncertainties about the most relevant outcome measures also apply to the evaluation of reflux symptom relapse during long term management. For instance, the CADET-HR study found that the number of heartburn free days gave the greatest differentiation between on-demand omeprazole or ranitidine therapy compared with slightly different versions of "unwillingness to continue".[69]

Data such as these underline the difficulties experienced when translating the results of clinical trials into clinical practice, and the need for guidance when assessing reflux symptom severity. However, identification of these difficulties raises many questions. For example, should it be symptom severity, duration, frequency, or "density" that is assessed? Should the response to therapy be documented as complete abolition of symptoms or adequate control of symptoms, and should the response be assessed by the clinician or by the patient? Is heartburn the only symptom that should be assessed and, if not, what other symptoms are relevant? These and other issues were addressed in this session of the workshop, and the outcome of the deliberations will be presented under five topic categories.

- How should treatment response be measured in reflux disease?
- Which of the typical reflux symptoms should be measured to assess treatment response?
- What are the outcome variables for long term therapy?
- Are there symptom patterns that predict outcome for therapy of reflux disease?
- How should extra-oesophageal symptoms, ascribed to reflux, be monitored during therapy?

Propositions, voting, and discussion
How should treatment response be measured in reflux disease?

(2.1) *In clinical trials, the proportion of patients who have been free of heartburn for one week prior to assessment is the optimal end point for assessment of symptom response* (nature of evidence: B).

Strength of recommendation: agree strongly, 22%; **agree, reservation, 44%**; *disagree, reservation, 33%; disagree strongly, 0%.*

Placebo controlled clinical trials of antisecretory therapy in patients with endoscopy negative GORD have shown that the differential between active treatment and placebo increases from an end point of "did study medication give sufficient control of your symptoms",[22] [70] to "adequate control of heartburn (one day with episodes of mild heartburn in the last seven days)",[70] [71] to "absence of heartburn at four weeks (no heartburn in the last seven days)".[70] [72] In other words,

the placebo response decreased with increasing stringency of the end point.

However, because more than 90% of patients accept up to one day of mild heartburn during treatment as sufficient control of their heartburn (see fig 1),[6] a significant number of workshop participants voted to ''disagree with the proposition with reservation''. Counter to this, it was noted that in the USA at least, patients are willing to pay more for absence of symptoms,[73] suggesting that classification of a response as sufficient does not mean that patients do not want, or should be denied, more effective therapy.

This poses a fundamental question: should the end point in clinical trials be optimal for the patient or, for example, optimal for discriminating between treatments? The majority view was that what patients accept is not necessarily optimal as a clinical trial end point, in part because acceptance may depend on other factors in addition to the frequency and severity of symptoms. It was felt that an end point, defined as ''no episodes of heartburn during the last seven days of study'' is attractive as it is rigorous, unambiguous and, therefore, methodologically sound. A one week timeframe was considered to be reasonable for standard clinical study durations although it may not be appropriate for shorter studies intended, for example, to assess the rapidity of symptom improvement. The advantages of using a one week timeframe for an end point are that it results in low placebo response rates and that it provides the patient with an internal standard of the best possible care. In addition, an end point of complete absence of heartburn at four weeks predicted healing in patients with reflux oesophagitis,[17 74 75] and predicts subsequent symptom status while on PPI therapy.[76]

Editorial comment. In clinical trials, complete absence of symptoms for a predefined time period provides a clear reproducible end point that allows comparison between studies. There were however concerns that this end point is too stringent and too far removed from clinical practice. Consequent discussion in the workshop identified that a less stringent end point (for example, less than two mild symptom episodes in the prior week) may be an acceptable measure of symptom response in clinical practice.

The discussion of this proposition did not address a precise definition of ''absence of symptoms'' but it is worth noting that terms used in recent oesophagitis healing studies,[17 74 75] ''complete resolution of heartburn'' (investigator assessment of symptoms over the previous week) and ''sustained resolution of heartburn'' (patient diary card record of symptoms on the previous seven days), led to somewhat different estimates of treatment efficacy within the same studies. Both measures are consistent with the above proposition and, although the difference in reported outcome may reflect discrepancies arising from patient self-assessment compared with investigator assessment (discussed in proposition 2.8, below), it is necessary to ensure that all clinical trial end points are defined as precisely as possible.

(2.2) *In clinical practice, patient satisfaction with improvement in reflux symptoms is the optimal measure of response to therapy* (nature of evidence: D).

Strength of recommendation: agree strongly, 8%; ***agree, reservation, 32%****; disagree, reservation, 32%; disagree strongly, 28%.*

Patient satisfaction with improvement in reflux symptoms was seen to be an intuitively meaningful measure of treatment response. However, the very limited relevant data available from clinical practice indicate that there may be little difference in patient satisfaction with respect to the extent of symptom reduction (see section 4). Additionally, there are measurement issues with the data available, such as use of single item measures, response bias, and acquiescence bias. The majority of participants therefore questioned

whether patient satisfaction with improvement in reflux symptoms has actually been shown to be the optimal measure, especially given the lack of validated tools for measurement of satisfaction. In addition, a confounding factor is that patient satisfaction is related to their expectations prior to therapy. The challenges associated with measurement of patient expectation and satisfaction are addressed in more detail in a later section.

(2.3) *In clinical practice, assessment of symptom response using daily diaries is feasible* (nature of evidence: C).

Strength of recommendation: agree strongly, 8%; agree, reservation, 0%; disagree, reservation, 12%; ***disagree strongly, 81%****.*

(2.4) *In clinical practice, it is useful to assess symptom response with daily diaries* (nature of evidence: C).

Strength of recommendation: agree strongly, 8%; agree, reservation, 4%; disagree, reservation, 23%; ***disagree strongly, 65%****.*

Although the value of daily diaries was recognised in the workshop, their use was not considered to be practicable in clinical practice. Daily diaries have been used extensively in the clinical trial setting, providing valuable data and high diary response rates have been reported (see McColl[3] in this supplement *(page iv49–iv54)*). In clinical practice however, accuracy and compliance are likely to be poor. Diary card records of peak flow measurements by asthma patients have been reported to contain at least one discrepancy in 75% of cases[77] although this may reflect poor compliance with peak flow measurements as much as poor compliance with the process of completing a daily diary. In addition, actual compliance with paper diaries has been shown to be only 11% compared with the 90% compliance that was reported by patients, and ''hoarding'' (when the patient fills in the diary at the end of the week, for example) was common, although the study which generated these data employed a rigorous protocol, requiring four diary entries per day.[78] Actual compliance was 94% when electronic diaries were used but this is not currently practicable for broad use in clinical practice. While the use of daily diaries in clinical practice was not seen to be practicable, the second proposal was not rejected as strongly.

Editorial comment. This discussion reflects the view that daily diaries may still be qualitatively useful in clinical practice in helping assess efficacy of therapy in selected patients, and as a tool to facilitate clinician/patient communication. (See also McColl[3] in this supplement *(page iv49–iv54)* for discussion of diary cards.)

(2.5) *In clinical trials, a modified Likert scale is superior to a visual analogue scale for measurement of symptom status* (nature of evidence: C).

Strength of recommendation: agree strongly, 31%; ***agree, reservation, 65%****; disagree, reservation, 4%; disagree strongly, 0%.*

Comparison of modified Likert scales and visual analogue scales (VAS) has shown that it is time consuming to train patients to use a VAS, and that it makes more sense to patients to discuss changes on a 1–7 point modified Likert scale than in terms of a 10–20 mm change on a 100 mm VAS.[79–81] A VAS is also more difficult to complete for the illiterate and the elderly.[80 81]

Editorial comment. The level of evidence was perhaps underestimated in the discussion. A detailed review of the literature (see also Wyrwich and Staebler Tardino,[4] in this supplement *(page iv45–iv48)*, for discussion of VAS versus modified Likert scales) suggests that there is reasonable evidence to support the proposition.

(2.6) *In clinical trials, seven is the optimal number of response options in a modified Likert scale for measurement of symptom status* (nature of evidence: C).

Strength of recommendation: agree strongly, 30%; **agree, reservation, 63%**; *disagree, reservation, 7%; disagree strongly, 0%.*

A seven point adjectival scale allows identification of small but clinically relevant changes and is suitable from a psychometric point of view. A change of 0.5 points on a seven point scale has been shown to be clinically relevant in GORD using the gastrointestinal symptoms rating scale (GSRS) reflux dimension.[82] Five point scales were discussed as a simpler alternative[83] but these have not been validated in GORD, and have the drawback that more patients are likely to choose the midpoint than with a seven point scale. Seven point scales are probably optimal but they need more extensive validation in reflux disease, particularly when translated into other languages.

Editorial comment. Taking propositions 2.5 and 2.6 together, the conclusion is that validated outcome measures with established responsiveness should be applied in clinical trials, and moreover that a seven point modified Likert scale should be used to assess symptom outcomes, rather than dichotomous ''yes/no'' scales, other scale gradings, or VAS. Most clinical trials to date have used four point scales, which is probably suboptimal.

(2.7) *In clinical trials, a symptom improvement score of 0.5 on a seven point modified Likert scale over placebo is the minimally important difference* (nature of evidence: C).

Strength of recommendation: agree strongly, 0%; agree, reservation, 23%; **disagree, reservation, 46%**; *disagree strongly, 31%.*

There are limited data to support the proposition as it relates to global symptom improvement but existing data support the notion that a change of 0.5 is a minimal clinically important difference with respect to the reflux dimension of the GSRS. Furthermore, there are data to support an improvement score of 0.5 on a seven point modified Likert scale as a minimally important difference using specific quality of life scales. Mean changes in the QOLRAD scale have been shown to correlate with overall treatment effect classifications, according to a seven point modified Likert scale.[82]

Editorial comment. The level of evidence is probably underestimated in that the cited studies do provide evidence in support of the clinical relevance of an improvement of 0.5 points. However, despite initial validation studies,[84] data are not available to confirm that a change of this magnitude is a minimal clinically important difference for global symptom scales.

(2.8) *Patient self-report of reflux symptoms is more appropriate than clinician assessment in measuring treatment effect* (nature of evidence: C).

Strength of recommendation: **agree strongly, 44%**; *agree, reservation, 41%; disagree, reservation, 11%; disagree strongly, 4%.*

In general, there is only weak correlation between patient and clinician assessment of symptom severity.[83] Analysis of the AstraZeneca clinical trial database in GORD[85] shows fair to moderate agreement between investigators and patients, with better agreement at the lower end of the symptom severity continuum. However, clinicians tended to underestimate symptom severity and, although this may be partially due to interpretation of heartburn, the same pattern was seen across the range of GORD symptoms and has been reported generally for other conditions.[83] As it is the patient, not the clinician, who experiences symptoms, it was agreed that more weight should be assigned to the patient's assessment. This proposition relates only to the assessment of symptom severity and treatment effect for defined symptoms. The assessment of symptoms for diagnostic purposes may require greater input from the clinician.

Editorial comment. See McColl[3] in this supplement **(page iv49–iv54)** for a more detailed discussion of clinician versus patient assessments. The recommendation that self-reported measures of reflux symptoms are preferable to physician based measurement in clinical trials is an important concept. It reflects the reality that the physician's assessment is, of necessity, based on the patient's self assessment and that there is, therefore, no a priori reason to accept the physician's assessment preferentially. However, this recommendation is distinct from the recommendation that global clinical opinion, based on a technically adequate clinician interview, is the most accurate approach to the diagnosis of GORD (see propositions 1.11, 3.14, and 3.13). Diagnosis is more complex (see proposition 1.11 which specifies the physician interview as the means of diagnosing predominant heartburn) than assessment of therapy. While patients may find it difficult to describe or define their predominant symptom (see proposition 1.7), self-reporting of symptoms following therapy is much simpler, as it involves a predefined symptom scale or a dichotomous ''yes/no'' response and a baseline comparator.

(2.9) *Measurement of heartburn severity does not provide any additional information to the measurement of heartburn frequency in assessing response to therapy* (nature of evidence: C).

Strength of recommendation: agree strongly, 0%; agree, reservation, 8%; **disagree, reservation, 46%; disagree strongly, 46%**.

Both symptom frequency and severity are important in assessing response to therapy. An analysis by Sharma and colleagues[1] in this supplement **(page iv58–iv65)**, indicates that frequency is a more sensitive and conservative measure than severity of symptoms but that severity correlates better with healing of oesophagitis, although the data are sparse. Measurement of frequency of heartburn alone risks underestimating the impact on the patient of infrequent but severe episodes exemplified by nocturnal heartburn with choking or severe non-cardiac chest pain. Given that more than one episode of mild heartburn per week is not acceptable to patients (see fig 1), both severity and frequency are important to the patient.

Editorial comment. It is important to emphasise that this proposition addressed the relationship between symptom characteristics and the response of symptoms to treatment. It did not address the relationship between symptom characteristics and the presence or persistence of oesophagitis in response to therapy.

(2.10) *Both frequency and severity of heartburn should be measured on therapy, using validated scales, in clinical trials where heartburn is the primary entry criterion*.

Strength of recommendation: **agree strongly, 59%**; *agree, reservation, 37%; disagree, reservation, 4%; disagree strongly, 0%.*

Editorial comment. Acceptance of this proposition is a corollary of the rejection of the previous proposition. As frequency and severity may vary independently in some, if not all, patients, it is important to measure changes in both when assessing a patient's response to therapy.

Which of the typical reflux symptoms should be measured to assess treatment response?

(2.11) *In clinical trials, there is no need to monitor all reflux symptoms in patients with typical symptoms since heartburn response is associated with response of other symptoms* (nature of evidence: C).

Strength of recommendation: agree strongly, 4%; agree, reservation, 19%; **disagree, reservation, 63%**; *disagree strongly, 15%.*

An analysis by Sharma and colleagues[1] in this supplement **(see page iv58–iv65)** indicates that absence of heartburn correlates with absence of regurgitation, and with

absence of dysphagia. However, this is based on patient groups, and it is not known if this applies in individual patients. Indeed, the severity of regurgitation and heartburn does not correlate in all patients. Different reflux related non-heartburn symptoms may be present in different patients. For example, in a recent study, acid regurgitation (72.6%) was significantly more prevalent than epigastric pain (50.0%), retrosternal pain (47.1%), retrosternal tightness (33.2%), or nausea (36.5%), and these symptoms responded differently to therapy.[85] Thus monitoring heartburn alone risks missing improvement or worsening of other symptoms attributable either to the disease process or to therapy.

Editorial comment. Symptoms other than heartburn should be monitored in clinical trials. One difficulty is that although patients may have reflux symptoms other than heartburn, it is heartburn that is the enrolment criterion for most studies of therapy in GORD, and change in heartburn severity or frequency is the primary symptomatic outcome. Thus studies are not generally designed or powered to examine the effect of therapy on other symptoms or to correlate changes in heartburn with changes in other symptoms. In addition, most data are from acid suppression trials in which regurgitation and dysphagia both respond to therapy. However, symptoms that may respond to or develop as a result of other medical, surgical, and endoscopic treatments should also be monitored in clinical trials.

(2.12) **In clinical practice, regurgitation should be evaluated routinely** (nature of evidence: E).

Strength of recommendation: agree strongly, 31%; **agree, reservation, 69%**; disagree, reservation, 0%; disagree strongly, 0%.

(2.13) **In clinical trials, regurgitation should be evaluated routinely** (nature of evidence: C).

Strength of recommendation: agree strongly, 30%; **agree, reservation, 44%**; disagree, reservation, 19%; disagree strongly, 7%.

Despite an absence of data, routine evaluation of regurgitation was recommended in both clinical practice and clinical trials as it does not necessarily occur in all patients with heartburn, and vice versa.[86] Further analyses of data from Belgium,[87] undertaken specifically for the workshop, show that in patients with no or mild heartburn, moderate or severe regurgitation is present in some 5% of patients in primary care and 16% in the specialist setting. Regurgitation is an important symptom of reflux disease that should be measured.

Editorial comment. Assessment of regurgitation may also be hampered by the lack of standardised description, akin to the ''word picture'' of ''retrosternal burning rising towards the throat'' that was developed to standardise the description of heartburn.[25] As indicated above, different therapies may have different effects on these symptoms but the spread of voting in proposition 2.13 reflects recognition that the focus of the trial may not require monitoring of regurgitation.

What are the outcome variables for long term therapy?

(2.14) **In clinical trials, a validated measure of patient satisfaction with heartburn control is an important outcome measure for evaluation of long term treatment** (nature of evidence: D).

Strength of recommendation: agree strongly, 15%; agree, reservation, 38%; **disagree, reservation, 42%**; disagree strongly, 4%.

Subjective measures of symptom response are essential, and if a validated measure of patient satisfaction with heartburn control were available, it could be a valuable outcome measure. However, the proposition is difficult to support as no validated instrument exists. A systematic review of comparative studies of surgical and medical therapy

for GORD highlights the numerous outcome measures that have been assessed, including patient satisfaction, but no unifying outcome was expressed, and the results were too heterogeneous for meta-analysis.[88] A study of post-surgical symptoms following open and laparoscopic antireflux surgery showed discrepancies between whether patients would recommend the surgery (similar for both procedures), and their reported satisfaction (lower for laparoscopy) and failure rates (higher for laparoscopy).[89] ''Willingness to continue''[69 90 91] is probably not a valid or reliable measure of patient satisfaction, and global measures of efficacy do not reflect satisfaction accurately. Patients' expectations influence their satisfaction with treatment and, as expectations may change with ongoing therapy, the assessment of satisfaction, on its own, is a poor measure of efficacy. In conclusion, a validated global measure of patient satisfaction and dissatisfaction is needed to compare outcomes of different therapeutic interventions.

Editorial comment. A validated measure of satisfaction could provide a scale by which different treatment modalities could be compared, including surgery, endoscopic therapy, and different medical treatment strategies. This would be of value even though patient satisfaction is dependent, not only on symptom control, but also, among other things, on the knowledge and expectations of patients as well as treatment complications and costs. This opinion was confirmed by broad acceptance of a subsequent proposition to this effect (see proposition 4.12).

(2.15) **Unwillingness to continue treatment due to inadequate control of heartburn should be the primary outcome for clinical trials of on-demand therapy** (nature of evidence: C).

Strength of recommendation: agree strongly, 15%; agree, reservation, 30%; **disagree, reservation, 52%**; disagree strongly, 4%.

Disagreement with this proposition was based on the fact that willingness to continue is influenced by factors other than efficacy, a timescale is not specified, and that evidence is limited for this outcome measure. However, data from six month placebo controlled studies of on-demand therapy in patients with endoscopy negative GORD support the view that discontinuation in these studies is due to inadequate control of heartburn.[69 90 91] The primary end point was willingness to continue but separate evaluation of heartburn status shows that discontinuation was virtually entirely due to insufficient control of heartburn. Although not relevant to this proposition, which is specific to clinical trials of on-demand therapy, willingness to continue might be applicable to studies of regular maintenance therapy but it cannot be applied to surgical studies unless the concept could be tested and validated as ''unwillingness to continue'' after surgery without the use of supplementary therapy (medical, surgical, or endoscopic). A measure is needed that focuses more adequately on the control of reflux symptoms and is thus more broadly applicable, such as ''satisfaction with control of heartburn'', but see proposition 2.14 above. Further studies are required, preferably over time periods longer than six months, to define more precisely what are the reasons for and implications of a patient's ''willingness to continue'' therapy. That said, unwillingness to continue is currently a useful outcome measure for clinical trials of on-demand therapy.

Editorial comment. One of the difficulties with the concept of ''willingness to continue'' is that it is not known how it relates to symptom characteristics before, during, and after therapy. However, although there were concerns with ''willingness to continue'' as an end point, there are difficulties with alternative end points. Pragmatically, although it may not be the optimal outcome measure, willingness to continue is the main measure used currently for studies of

''on-demand'' therapy, and it is also relevant to patient management in clinical practice.

Disagreement with the proposition was, in part, because willingness to continue was seen to be influenced by factors other than sufficient control of heartburn. When this concept was reviewed, and inadequate control of heartburn was not specified as the reason for discontinuation, acceptance of the proposition increased (see proposition 2.16).

(2.16) *In long term on-demand trials, unwillingness to continue should be the primary outcome*.

Strength of recommendation: agree strongly, 4%; **agree, reservation, 54%**; *disagree, reservation, 29%; disagree strongly, 13%.*

Despite the removal of heartburn as a qualifier to describe the reason for discontinuation, over 40% of participants still disagreed with the proposition. This may have been due to use of the phrase ''unwillingness to continue'' rather than, for example, ''happy to continue''. However, there was also concern that patients may discontinue therapy because they feel better and not because therapy has failed. ''Unwillingness to continue'' is a difficult concept which needs further study to define what patients understand by willingness and unwillingness to continue a treatment strategy, as well as the specific treatment related and treatment unrelated reasons why they might discontinue therapy. Outside of the context of clinical studies, it is also important to note that costs and ''willingness to pay'' on the part of the patient or a third party payer may be determinants of the patient's willingness to continue.

(2.17) *Assessment of treatment efficacy in long term therapeutic trials should include a record of the number of symptom free days* (nature of evidence: D).

Strength of recommendation: agree strongly, 12%; **agree, reservation, 60%**; *disagree, reservation, 28%; disagree strongly, 0%.*

Heartburn free days is a sensitive measure of efficacy, albeit labour intensive, and although it is unlikely to be the primary outcome measure of a trial, it may provide a more patient centred measure than a change in symptom score. This has been demonstrated for the CADET-HR study[69] in which mean percentage of days spent heartburn free over six months was a greater differentiator of efficacy between omeprazole and ranitidine than willingness to continue, although the study was not designed to compare the two outcomes. ''Symptom free days'' is sensitive to the cumulative effects of treatment and a more sensitive measure of change than survival curves or symptoms at end point. Numerous practical questions remain however concerning the measurement of symptom free days. Should this be assessed by diary cards, retrospective assessment, telephone contact, or some other technology, such as an electronic diary? Should measurements be conducted daily throughout the trial, recognising that this is labour intensive, or should they be taken for a period prior to the end of the trial or rather at various time points during the trial, recognising that there are no guidelines as to when these assessments should be carried out? There was agreement that further research is needed to answer these questions.

Editorial comment. Reservations were based in part on the fact that it may be impracticable, although not impossible,[69] to use daily diary cards in long term studies, and that it is difficult to relate this back to clinical practice. However, if ''willingness to continue'' is similar between two treatments but there is a large difference in symptom free days, it suggests that ''willingness to continue'' is an insensitive index of treatment response or that it may be measuring something other than symptom control. Under these circumstances, there is an incentive for the development of a methodology to monitor symptoms on a regular basis,

partly to monitor symptom response to therapy and partly to identify appropriate patients for study.

(2.18) *When on-demand long term drug therapy is being studied in clinical trials, the consumption of medication during the trial should be recorded* (nature of evidence: B).

Strength of recommendation: **agree strongly, 91%**; *agree, reservation, 9%; disagree, reservation, 0%; disagree strongly, 0%.*

Although not a primary measure of efficacy, consumption of trial medication should be recorded both for information on efficacy and for health economic evaluation. Use of rescue medication (for example, antacid consumption) should also be recorded as a measure of efficacy. Support for this comes from a considerable number of clinical trials of on-demand therapy.[69 90 91] The dose of trial medication taken, as well as frequency and timing of dose, may provide useful information, including pathophysiological insights into patterns of relapse.

Editorial comment. It was recognised that medication intake monitoring is impracticable in clinical practice but that it is useful in clinical research, despite the difficulties inherent in acquiring the data. The development of new technologies, such as ''MEMS'' (Medical Event Monitoring System) containers,[91] will facilitate monitoring of medication usage, and this should also provide important data for health economic studies.

(2.19) *The same outcome measures of symptom status should be used for trials of drug therapy, antireflux surgery, and other therapeutic interventions* (nature of evidence: D).

Strength of recommendation: **agree strongly, 56%**; *agree, reservation, 33%; disagree, reservation, 11%; disagree strongly, 0%.*

The overriding argument is that there are increasing numbers of therapies, all for the same disease and, with a broader spectrum of GORD patients now being treated with surgery, it is very important that the same outcome measures be used to assess these different interventions. There is now experience of using the psychological general well being index (PGWBI) and the GSRS to assess the outcomes of surgery. Pretreatment symptoms need to be addressed more effectively in surgery trials, in particular to distinguish these background side effects from true surgery related side effects.

Editorial comment. Reservations regarding the proposition were, firstly, that some additional outcome measures would be needed for surgical trials but not necessarily for medical therapy trials and, secondly, that the predictive values of these measures might vary between primary and tertiary care centres. As blinding is virtually impossible in trials of surgical therapy, it is particularly important that measures of symptom status be validated in both medical and surgical treatment populations. All different therapies should be assessed in the same manner to provide comparable data.

(2.20) *The proportions of patients taking drug therapies and the volume of their use following antireflux surgery or other therapeutic interventions are too imprecise for use as a primary efficacy measure in clinical trials* (nature of evidence: D).

Strength of recommendation: **agree strongly, 71%**; *agree, reservation, 25%; disagree, reservation, 4%; disagree strongly, 0%.*

High rates of antisecretory drug use have been reported following antireflux surgery[92] but this may be inappropriate use.[93]

Editorial comment. Although the proposition was accepted as written, medication use remains an important secondary outcome measure as it reflects an intention to treat outcome and it is important for health economic studies, particularly when comparing medical and surgical therapies.

Furthermore, there are few data on the proportions of patients in long term medical studies who may take their medication for reasons other than heartburn control.

(2.21) *Absence of heartburn after an initial course of therapy is a good predictor of freedom from oesophagitis* (nature of evidence: A).

Strength of recommendation: **agree strongly, 85%**; *agree, reservation, 15%; disagree, reservation, 0%; disagree strongly, 0%.*

Recent comparative studies in patients with reflux oesophagitis have shown that absence of heartburn with esomeprazole corresponds with absence of oesophagitis in at least 80% of patients.[17 74–76] Based on the analysis by Sharma and colleagues[1] in this supplement **(see page iv58– iv65)**, correlation of absence of heartburn with healing of oesophagitis is excellent. Overestimation of healing by the absence of heartburn is approximately 5% but this overestimate rises to 28% when reduction in heartburn is used as a predictor of oesophagitis healing. Absence of heartburn thus seems to be a suitable surrogate marker for healing of oesophagitis during short term (4–8 week) therapy although this needs further investigation, particularly documenting clinician versus patient self-assessment of the absence of heartburn.

Editorial comment. A qualifier to the conclusion from the analysis by Sharma and colleagues[1] in this supplement **(see page iv58–iv65)** is that absence of heartburn and healing of oesophagitis do not necessarily occur concurrently in the same patients.[94] Additionally, the nature of the evidence may be an overestimate because although the data are derived from randomised controlled trials, studies were not designed to assess the relationship between symptoms and healing in a randomised fashion.

(2.22) *Absence of heartburn during continuous long term therapy is a good predictor of freedom from oesophagitis* (nature of evidence: A).

Strength of recommendation: **agree strongly, 87%**; *agree, reservation, 13%; disagree, reservation, 0%; disagree strongly, 0%.*

The systematic review by Sharma and colleagues[1] in this supplement **(see page iv58–iv65)**, based on seven trials of antisecretory maintenance therapy, shows that absence of heartburn and absence of oesophagitis are well correlated. Absence of moderate to severe symptoms overestimated oesophagitis remission by approximately 9%. Similarly, in a meta-analysis of five randomised long term trials with omeprazole, asymptomatic relapse of oesophagitis was only found in 8.6% of patients.[94]

Editorial comment. Again, as for proposition 2.21, the nature of the evidence may be an overestimate as studies were not designed to assess the relationship between recurrent symptoms and recurrent oesophagitis in a randomised fashion, even though these data were derived from randomised controlled trials.

Are there symptom patterns that predict outcome for therapy of reflux disease?
(2.23) *Absence of heartburn after one week of PPI therapy predicts sustained symptom reduction after four weeks of therapy* (nature of evidence: B).

Strength of recommendation: agree strongly, 12%; **agree, reservation, 88%**; *disagree, reservation, 0%; disagree strongly, 0%.*

Pooled data from studies of esomeprazole in endoscopy negative reflux disease have shown that heartburn response during days 5–7 of the first week of therapy is the most discriminating predictor of treatment outcome although it was a secondary objective of the trials from which the data were derived.[95] Of patients who were heartburn free for days 5–7 of treatment, 85% were heartburn free at week 4 while of patients with moderate or severe heartburn every day for days 5–7, only 22% were heartburn free at week 4. Comparable data however are not available for patients with reflux oesophagitis. Also, the symptom response at four weeks may not be the gold standard because a proportion of patients who are symptomatic at four weeks may still become symptom free with more prolonged therapy. Knowledge that a patient responding at one week will continue to respond at four weeks is of value in clinical practice although the converse does not apply; a lack of response at one week does not necessarily mean that the patient will not respond at four weeks.

Editorial comment. Although early abolition of heartburn symptoms is predictive of a more sustained response, assessment of symptoms after one week of treatment probably has low specificity for the diagnosis of reflux related symptoms. This presumption should be tested prospectively, particularly because of widespread interest in the clinical potential of a ''PPI test'' or acid suppression test for the diagnosis of GORD and acid related disorders.[13 96 97]

(2.24) *Nocturnal heartburn at baseline is an important predictor of failure of PPI therapy* (nature of evidence: C).

Strength of recommendation: agree strongly, 0%; agree, reservation, 12%; **disagree, reservation, 52%**; *disagree strongly, 36%.*

Indirect evidence indicates that nocturnal heartburn, although common, is not a predictor of relapse or PPI treatment failure. Studies with both rabeprazole[98] and esomeprazole[99] have shown that improvement in daytime heartburn with PPI therapy is paralleled by improvement in nocturnal heartburn although this was a secondary study objective. An analysis of pooled studies with esomeprazole in a total of approximately 12 000 reflux oesophagitis patients shows that 42% had night-time symptoms at baseline. After four weeks of treatment, only 15% still had nocturnal heartburn. Although these data were part of a secondary analysis, they suggest that nocturnal heartburn improves in as many patients as does daytime heartburn. However, it may be that persistent nocturnal heartburn after initial therapy is more difficult to treat (or more troublesome to the patient) than persistent daytime heartburn, but this does not necessarily mean that nocturnal heartburn is a predictor of treatment failure.

(2.25) *Patients with multiple symptom patterns at baseline have a lesser response to PPI therapy* (nature of evidence: D).

Strength of recommendation: agree strongly, 23%; **agree, reservation, 73%**; *disagree, reservation, 4%; disagree strongly, 0%.*

Evidence in support of this proposition is limited. There are unpublished post hoc analyses of studies in endoscopy negative GORD patients which show that reflux symptoms respond less well to PPI therapy in patients who have more non-heartburn symptoms, as assessed by the GSRS (AstraZeneca, data on file). The percentage of patients with absence of heartburn at four weeks is lower in patients with over 13 GSRS items, including, for example, diarrhoea, than in those with only one or two items. A minority of GORD patients have multiple unexplained symptoms which may be associated with other psychological distress and, in general, medical and surgical treatments have been shown to be less effective in somatising patients.[100] In addition, patients with uninvestigated heartburn dominant dyspepsia are less likely to respond to initial therapy if they have concomitant symptoms of irritable bowel syndrome.[68 101]

Editorial comment. This proposition raises a matter that is particularly important for the treatment of endoscopy negative reflux disease patients or patients with uninvestigated reflux symptoms. Proposition 2.11 addressed the need to monitor symptoms other than heartburn and, in clinical trials, it may also be necessary to consider a prospective study

of the role of other symptoms as predictors of treatment response. Again, as with proposition 2.24, there is no indication that the impact of multiple symptoms on outcome is specific to PPI therapy, or that they are predictive, specifically, of PPI treatment failure

How should extra-oesophageal symptoms, ascribed to reflux, be monitored during therapy?

(2.26) ***The response to treatment of extra-oesophageal symptoms caused by reflux occurs typically over weeks, rather than days*** (nature of evidence: E).

Strength of recommendation: agree strongly, 0%; **agree, reservation, 60%**; *disagree, reservation, 32%; disagree strongly, 8%.*

The spread of voting reflects the lack of evidence, and support for the proposition was based largely on empirical clinical experience that prolonged therapy is of value. A single study has shown that asthma symptom scores continue to decline over three months during omeprazole therapy[102] but the data are based on low patient numbers and are confounded by the fact that responders had more severe baseline symptoms than non-responders. Thus the symptoms of "responders" could just have been regressing to the mean with time, rather than therapy. Studies of treatment of suspected reflux laryngitis,[103] GORD related asthma,[104] and GORD and cough[105] have typically involved treatment of at least four weeks, and frequently longer, but the results are quite heterogeneous and confounded by the use of different, frequently high dose, therapies.

Editorial comment. Ear, nose, and throat (ENT) symptoms in laryngitis patients, measured by the reflux symptom index (RSI) symptom scale, have been shown to respond to PPI therapy within two months while those assessed by the reflux finding score (RFS) took 4–6 months to respond, although these measures are not validated.[106] Thus the above data should be qualified by the comment that a four month difference in the time to response may reflect the measurement instrument rather than the disease. It should also be noted that even typical reflux symptoms do not necessarily respond rapidly, and that there is an increase in the proportion of erosive oesophagitis patients who achieve symptom reduction as treatment is continued up to four weeks,[16 73 74] and beyond, to eight weeks.[107] Thus it is quite reasonable to suppose that reflux related respiratory tract symptoms may take many weeks to resolve.

(2.27) ***The measurement tools needed to assess the treatment response of extra-oesophageal symptoms are different from those needed to assess the response of heartburn to treatment*** (nature of evidence: E).

Strength of recommendation: **agree strongly, 83%**; *agree, reservation, 17%; disagree, reservation, 0%; disagree strongly, 0%.*

There is little documentation for this proposition but extra-oesophageal GORD is quite different from traditional GORD. Heartburn, regurgitation, and oesophagitis are often absent in patients with extra-oesophageal GORD who may have multiple aetiologies for their extra-oesophageal symptoms and signs. Consequently, the measurement tools for these symptoms clearly need to be different from those used in traditional GORD patients. Potential objective parameters include peak expiratory flow rates, spirometry, ENT examination, and cough meters. However, spirometry has not been shown to be useful, and interobserver variability is very poor for ENT examinations of the mucosa.[108] Potential subjective parameters include questionnaires for asthma, cough,[109] and ENT complaints (such as the RSI and RFS),[106] but these require validation. This is an area requiring considerable further research, including the development of new validated measurement tools and a better understanding of the

pathogenesis of oral, ENT, and respiratory conditions that are ascribed to gastro-oesophageal reflux.

Editorial comment. Tools designed to measure the severity of symptoms, such as heartburn, are very unlikely to be valid in the assessment of dyspnoea, cough, wheezing, or dysphonia.

Future directions

Consideration of the practicalities of reflux symptom severity assessment summarised above defined many areas that need further study.

One overriding dilemma relates to the definition of relevant reflux symptoms. Future research is needed into this. Many patients experience other symptoms, in addition to heartburn, and these symptoms may respond differently to therapy. To date, the majority of studies have concentrated on heartburn as the primary outcome variable. This is the most prevalent symptom and the one that responds most predictably to acid suppression therapy but other symptoms should also be assessed in conjunction with heartburn. One approach, to use a global outcome score, has the advantage that it would encompass the overall response to therapy. However, the disadvantage is that inclusion of symptoms that are less likely to respond to therapy may render the score less sensitive as a measure of treatment outcome. The use of a global score is complicated further by the fact that heartburn is a common, and possibly incidental, symptom in patients who may have many other symptoms of functional bowel disorders, including dyspepsia and irritable bowel syndrome. Prediction of a poorer response to PPI therapy by the presence of multiple symptoms (see proposition 2.25) indicates the need to determine whether patients with dominant heartburn respond differently from those with non-dominant heartburn. Similarly, regurgitation may not respond as well to therapy as heartburn, and it will be important to conduct prospective studies of therapy in patients who have regurgitation as their dominant or only symptom.

There is a continuing need to improve the translation of clinical trial outcomes into clinical practice. Abolition of symptoms may be an important outcome in clinical research but it is an unrealistic expectation for many patients in day to day practice. In consequence, it will be important to define better the relationship between abolition of symptoms and clinically acceptable outcomes, including measures of patient satisfaction.

As reflux symptoms vary considerably in severity and frequency between and within individuals, better techniques are needed (for example, using a personal digital assistant, mobile phone, or two way pager) for recording symptoms on a daily basis without the difficulties attributable to recall bias, compliance, or hoarding that hamper the use of a daily diary card. Modified Likert seven point scales should also be validated across the spectrum of reflux related symptoms, including patients with heartburn dominant, heartburn non-dominant, and extra-oesophageal symptoms. Additionally, changes in symptom severity should be correlated with clinically relevant outcomes to define minimal clinically important treatment related changes for these other symptoms. Related to this, "word pictures", akin to that developed to describe heartburn, are likely to help patients understand and report less typical symptoms of reflux disease more objectively and reliably.

Assessment of outcomes in long term therapy of reflux disease is particularly difficult, partly because expectations appear to change during its therapy, and partly because of the range of different treatment options and strategies. "On-demand" medical therapy is a useful option for the management of milder or less frequent symptoms. There is a need for further research into the temporal pattern of symptom occurrences during such therapy and the factors that drive

''on-demand'' use of medication. More needs to be known about the natural history of symptom recurrence in GORD, and the factors that determine a patient's willingness to continue an established management strategy, whether it be ''on-demand'' medical therapy or use/rejection of rescue therapy to treat recurrent symptoms after a surgical or endoscopic antireflux procedure. It seems reasonable to assess symptom free days and medication usage (including active therapy and rescue therapy) during long term therapy to determine treatment efficacy and the health economic implications of different management strategies. The practicalities of making these measurements reliably are challenging, as these measures are subject, like daily diaries, to confounding by recall bias, compliance, and hoarding. Again, new technologies may avoid some of the difficulties experienced to date in acquiring these data.

The pathophysiological mechanisms responsible for the generation of reflux symptoms, whether they be typical oesophageal symptoms or atypical extra-oesophageal symptoms, are poorly understood. A better understanding of these mechanisms may help determine which patients will respond to therapy and will facilitate prospective studies to identify predictors of symptom response across the spectrum of reflux related diseases.

In conclusion, assessment of symptom severity is fundamental to the management of reflux disease but there is much still to be done to optimise the treatment of patients with this very common condition.

3. QUALITY OF LIFE: PROPOSITIONS, VOTING, DISCUSSION, AND COMMENTARY
Introduction
Quality of life is a dynamic construct and therefore inherently difficult to assess as time of assessment will affect the responses given. Moreover, quality of life may not be adequately represented in the impact of the disease on a patient's daily activities, such as sleep and work, as assessed by quality of life questionnaires, but can be defined more broadly as the gap between a patient's expectation and experience. Furthermore, quality of life measures need to be patient centred rather than reflecting what clinicians think is important.

Although there has been considerable discussion of the importance of patient quality of life as an outcome of therapy, surprisingly, it has rarely been assessed in clinical trials. The systematic review by Sharma and colleagues[1] in this supplement *(page iv58–iv65)* found that of 157 publications on long term medical therapy for GORD, 48 were eligible and from these, data were extractable from only 37. Of these, only three assessed patient quality of life, two using the PGWBI and one using SF-36. For short term therapy, 126 publications were eligible, and data were extractable from 108, of which six measured quality of life. One publication used the PGWBI, one used SF-36, two used other generic measures, and two used disease specific measures. Thus of 174 eligible randomised controlled trials in GORD, only nine assessed patient quality of life as an outcome.

If patient quality of life is to be a key outcome in clinical trials in GORD, guidance is needed on how best to measure and interpret changes in quality of life. Should measurement be based on utilities or domains, and should measures be generic or disease specific? If disease specific measures are used, are they really just symptom impact scores? In clinical practice, adaptation is often observed whereby patients adjust to the quality of life they have, so how practical is measurement of patient quality of life in clinical practice, particularly as it could be time consuming? Can quality of life measures be a substitute for symptom measures and will they serve to raise patient expectations?

These questions, relating to quality of life and symptom assessment in GORD, were addressed in the workshop under five topic areas.

- Should generic or disease specific measures of quality of life be used in determining response to therapy in clinical trials?
- How frequently should quality of life be measured in trials in reflux disease?
- How should changes in quality of life be reported in trials?
- Is ''symptoms sufficient to impair quality of life'' a meaningful concept for defining presence of reflux disease in clinical trials or practice?
- Do quality of life measures correlate with other outcome measures?

Should generic or disease specific measures of quality of life be used in determining response to therapy in clinical trials?
(3.1) **Disease specific measures of quality of life are more responsive to changes in the impact of reflux symptoms in response to therapy** (nature of evidence: C).

*Strength of recommendation: agree strongly, 25%; **agree, reservation, 71%**; disagree, reservation, 4%; disagree strongly, 0%.*

Generic measures, such as the SF-36, are less responsive to symptom improvement in GORD than disease specific questionnaires.[110] In the ProGERD study of esomeprazole therapy, the effect sizes (standardised means) in components of the SF-36 were approximately 0.3–0.5 compared with over 1.0 with the disease specific QOLRAD instrument,[111] while the physical and mental components of the SF-36 were unable to detect changes in GORD patients treated with lansoprazole or ranitidine.[112] A disease specific measure of quality of life should therefore be used to assess the impact of GORD symptoms in response to therapy. Addition of generic measures would serve to increase clinical trial burden considerably.

(3.2) **Generic measures of quality of life are appropriate for making comparisons of disease impacts across different diseases** (nature of evidence: B).

*Strength of recommendation: **agree strongly, 100%**; agree, reservation, 0%; disagree, reservation, 0%; disagree strongly, 0%.*

The relative benefits and shortcomings of generic versus disease specific quality of life measures are well recognised,[113] and while generic measures are appropriate for making comparisons across diseases, disease specific measures are, by definition, inappropriate for this purpose. There are numerous examples of the use of generic measures, such as the SF-36 questionnaire, to compare the impact of diseases on quality of life—for example, comparing GORD with heart failure and clinical depression.[114 115]

Editorial comment. There is clearly a trade off implied by the last two propositions. On the one hand, efficient trial design demands a responsive and, therefore, disease specific measure. On the other hand, if any comparison with other disease states is likely, a generic population validated measure, such as the EuroQol, should be used. These are particularly appropriate when health economic outcomes are being considered.

(3.3) **In clinical trials of reflux disease, measurement instruments must be validated for both the language and culture of participating patients** (nature of evidence: E).

*Strength of recommendation: agree strongly, 37%; **agree, reservation, 59%**; disagree, reservation, 4%; disagree strongly, 0%.*

Some work has been done on multiple translations of the SF-36 questionnaire[116] and the development of cross cultural

questionnaires.[117] Anecdotal evidence suggests that the QOLRAD instrument performs similarly in different countries. However, validation of measurement instruments for different cultures is needed. Although expensive, formal translation may not be enough, given that wording is interpreted differently between countries, and language, responsiveness, and reliability all need validating. This may not be practical in every language and culture. Further research should determine a core set of items that could be used in cross cultural studies in the area of GORD.

How frequently should quality of life be measured in trials in reflux disease?

(3.4) **Measurement of quality of life at baseline and at the end of initial therapy or at dropout is sufficient for clinical trials of initial drug therapy** (nature of evidence: C).

Strength of recommendation: agree strongly, 19%; **agree, reservation, 58%**; *disagree, reservation, 23%; disagree strongly, 0%.*

The proposition would not be valid if patient quality of life showed significant variations during the course of therapy, rather than a progressive improvement. There are no data on daily quality of life in GORD patients treated with anti-secretory therapy but the few studies which have made more than one quality of life assessment after baseline support the view that there is a progressive improvement in patient quality of life during therapy.[118–120] In the study by Talley *et al* in patients with endoscopy negative GORD treated with esomeprazole or omeprazole, GSRS symptom scores improved in parallel with QOLRAD scores, with no clinically significant differences at two and four weeks. In reflux oesophagitis patients treated with esomeprazole,[120] heartburn severity and QOLRAD scores improved dramatically at four weeks, with a small additional improvement at eight weeks. Measurement of quality of life at baseline and at the end of initial therapy or at dropout is probably therefore sufficient for clinical trials of initial drug therapy although measurement at dropout is an important qualification.

Editorial comment. The dissenting opinion on this proposition relates partly to the wording of ''sufficient'', and the lack of evidence from multiple time points rather than two time points. Dissenting opinion also relates to the use of disease specific or generic measures of quality of life and the nature of the intervention. While the evidence presented supports the proposition that quality of life improves with treatment, the evidence is limited to disease specific measures and continuous therapy. Generic measures may be subject to competing influences on quality of life during the course of the trial and adaptation may also occur, reducing the impact of health status change. In addition, intermittent therapies, such as on-demand therapy, may result in fluctuating quality of life. This concept is addressed by the following proposition.

(3.5) **For clinical trials of continuous long term therapy of any type, time based measurement (for example, at yearly intervals) of quality of life is the most appropriate indicator** (nature of evidence: E).

Strength of recommendation: agree strongly, 18%; **agree, reservation, 64%**; *disagree, reservation, 14%; disagree strongly, 4%.*

(3.6) **For clinical trials of intermittent long term therapy, event based measurement of quality of life is most appropriate** (nature of evidence: E).

Strength of recommendation: agree strongly, 0%; agree, reservation, 22%; **disagree, reservation, 70%**; *disagree strongly, 7%.*

There is very little evidence with which to address these propositions, and recommendations must be based on expert opinion. Time based evaluations are administratively convenient and appropriate where quality of life is expected to be relatively stable over time although they may not capture the variation in relapsing-remitting conditions or on-demand therapies, particularly with less frequent assessments or smaller sample sizes. Time based evaluations are appropriate for comparison of two or more continuous long term therapies, with the reservation that the choice of assessment interval will depend on the nature of therapy, including surgery, and yearly intervals may not be adequate.

Event based evaluations are less convenient to administer as knowledge is needed of when the event occurs, be it a clinical event or change in therapy. If appropriately timed, event based measurements are likely to be more responsive when quality of life is expected to fluctuate over time and, in theory, they are appropriate for comparison of intermittent therapies. However, the proposition was rejected because of problems with its practicability, and the potential for bias if event related measurement in one trial group led to a significant difference in the timing of measurements between groups. A further issue is what form of measurement to use for comparison of continuous and intermittent therapy. In this case, regular, and more frequent, time based evaluations, with a large sample size, may be more appropriate than event based measurement.

Editorial comment. The group quite consistently rejected the notion of event based measurement, preferring the alternative option of more frequent measurement intervals, and accepting that some events might be missed.

How should changes in quality of life be reported in trials?

(3.7) **Reporting of subscales in quality of life measures is the most responsive measure of change in trials provided that adjustment is made for multiple testing and end points are prespecified** (nature of evidence: C).

Strength of recommendation: agree strongly, 24%; **agree, reservation, 68%**; *disagree, reservation, 8%; disagree strongly, 0%.*

Data from a single study of endoscopy negative reflux disease patients treated with esomeprazole or omeprazole indicate that subscales are more responsive than global scores.[82] Subscales of the QOLRAD were highly responsive, and effect sizes were impressive.

Editorial comment. This is to be expected, as GORD principally affects dimensions relating to pain, emotion, and physical function, having less effect on other dimensions. The effects on global scores are therefore somewhat diluted. This may be less so with disease specific measures, which focus more on aspects of quality of life relevant to particular diseases.

(3.8) **Reporting population derived QALYs (quality adjusted life years) is most appropriate for cost utility studies from the third party payer's perspective** (nature of evidence: D).

Strength of recommendation: agree strongly, 4%; agree, reservation, 35%; **disagree, reservation, 54%**; *disagree strongly, 8%.*

(3.9) **Reporting patient derived QALYs is most appropriate for cost utility studies from the patient's perspective** (nature of evidence: D).

Strength of recommendation: agree strongly, 0%; agree, reservation, 35%; **disagree, reservation, 54%**; *disagree strongly, 12%.*

These were largely rejected because of the lack of adequate utility measures in GORD. Without these, the propositions cannot be recommended as ''the most appropriate''. However, the workshop recognised that there will be increasing pressure from health care and research funding bodies to incorporate utility measures in GORD studies, and various public health bodies are promoting QALYs for the measurement of health care. This cannot be ignored, and the lack of patient centred end points in studies is a drawback.

Further work is needed to define the appropriate means of eliciting utility measures in GORD.

Editorial comment. QALYs are years of life multiplied by the utility of each year of life on a scale of 0 to 1, where 1 is perfect health and 0 is death. Thus 10 years of life at a utility of 0.5 would be 5 QALYs.[121] The whole concept of ''population derived'' QALYs can be criticised in that the methods used to assess them are based on presenting scenarios to healthy individuals and ascertaining their expectation of the health state described, rather than their experience of the actual health state. Critique from the perspectives of cognitive psychology and sociology[122] suggests that this is likely to be unreliable. The principal criticism is that subjects not directly experiencing the disease state fail to take adaptation into account.

Is ''symptoms sufficient to impair quality of life'' a meaningful concept for defining presence of reflux disease in clinical trials or practice?

(3.10) *For clinical trials, symptoms are a more appropriate entry and outcome measure than quality of life measures* (nature of evidence: C).

*Strength of recommendation: **agree strongly, 65%***; *agree, reservation, 23%; disagree, reservation, 8%; disagree strongly, 4%.*

Although changes in some dimensions of quality of life measures, such as food and drink problems, are very important to patients, the magnitude of change of quality of life measures is not as great as change in heartburn symptoms.[82] Thus while quality of life measures provide useful information, heartburn, and possibly regurgitation, are the most important entry and outcome measures for clinical trials in GORD.

Editorial comment. Although symptoms were seen to be a more appropriate entry and outcome measure than quality of life measures in clinical trials, quality of life measures are of potential value if used to assess secondary outcomes. They may also provide information on any adverse impact of intervention. This was recognised in a further proposition (3.11) that was fully accepted.

(3.11) *A quality of life measure that is responsive and measures multiple dimensions validly should be used to assess secondary outcomes in clinical trials*.

*Strength of recommendation: **agree strongly, 69%***; *agree, reservation, 31%; disagree, reservation, 0%; disagree strongly, 0%.*

(3.12) *For clinical practice, exploration of the impact of symptoms on patient quality of life is an important part of the assessment of adequate therapy* (nature of evidence: E).

*Strength of recommendation: **agree strongly, 38%; agree, reservation, 38%***; *disagree, reservation, 23%; disagree strongly, 0%.*

There is no evidence in the area of GORD to support this proposition but it has a strong theoretical basis in the ideas of holistic care, patient centeredness, and shared decision making. The only indirect evidence comes from a systematic review of the effect of formal decision aids on patient outcomes which found that they increased patient knowledge and reduced decisional conflict but did not increase satisfaction.[123] The main reservation with the proposition was that quality of life parallels symptom assessment. However, quality of life may not be concordant with symptom control. Patients with good symptom control may have impaired quality of life (for example, due to dietary restrictions) while, conversely, patients may have a good quality of life despite residual symptoms because they have adjusted their expectations. The marginal cost effectiveness of moving from minimal but tolerated symptoms to complete abolition will be very low, emphasising the value of exploration of the impact of symptoms on patient quality of life, which should be routine clinical practice.

(3.13) *A self-administered quality of life questionnaire can be used in conjunction with a symptom score to assess the presence or absence of reflux disease* (nature of evidence: E).

*Strength of recommendation: agree strongly, 0%; agree, reservation, 4%; **disagree, reservation, 85%***; *disagree strongly, 11%.*

No data exist to support this proposition. Quality of life is impaired in proportion to symptom severity[124] while GORD treatment improves symptom severity and quality of life proportionately.[82]

Editorial comment. In the absence of data demonstrating value of quality of life measurement on top of symptom assessment in the diagnosis of GORD, the additional burden cannot be justified, and the proposition was therefore rejected. Note that this is distinct from assessing the outcome of therapy where quality of life measurements do have a useful role.

(3.14) *A self-administered symptom scoring system used in conjunction with a quality of life questionnaire is more reliable than a global clinical opinion to assess the presence or absence of reflux disease* (nature of evidence: E).

*Strength of recommendation: agree strongly, 0%; agree, reservation, 0%; disagree, reservation, 19%; **disagree strongly, 81%**.*

The proposition implies that a self-administered symptom scoring system and quality of life questionnaire is better than clinical assessment, for which there are no data. Rejection of the proposition, though, assumes that the clinical interview is technically adequate in assessing the patient.

Editorial comment. Taken together, propositions 3.13 and 3.14 indicate that global clinical opinion, based on a technically adequate clinician interview, is the most accurate approach to the diagnosis of GORD, rather than a self-administered symptom scoring system for the patient, coupled with a quality of life questionnaire. This is a key recommendation and is consistent with proposition 1.11, which specifies the physician interview as the means of diagnosing predominant heartburn.

Do quality of life measures correlate with other outcome measures?

(3.15) *Quality of life measures correlate well with frequency of heartburn* (nature of evidence: C).

*Strength of recommendation: agree strongly, 25%; **agree, reservation, 71%***; *disagree, reservation, 4%; disagree strongly, 0%.*

Unpublished data from the ProGERD study (AstraZeneca, data on file) support the proposition, showing a good correlation between the frequency of heartburn and the dimensions of both the QOLRAD and SF-36 scales.

(3.16) *Quality of life measures correlate well with severity of heartburn* (nature of evidence: A).

*Strength of recommendation: agree strongly, 48%; **agree, reservation, 52%***; *disagree, reservation, 0%; disagree strongly, 0%.*

Quality of life measures have been shown to correlate well with the severity of heartburn for the QOLRAD,[68] PGWBI,[124] and SF-36 scales.[125] The majority of data relate to the response to short term therapy, and data are required on long term changes in quality of life in relation to symptom severity, as well as symptom frequency, overall treatment effect, and patient satisfaction.

Editorial comment. Although both these propositions (3.15 and 3.16) were supported, the evidence was limited to one trial and one measure for heartburn frequency, and three trials for severity. Early reports of effects frequently suggest a stronger effect than is subsequently confirmed by later studies. Much more evidence is needed before it can be

implied that symptom response can be taken as a proxy for improvement in quality of life.

(3.17) *Changes in quality of life correlate well with patient satisfaction with treatment* (nature of evidence: C).

*Strength of recommendation: agree strongly, 0%; agree, reservation, 18%; **disagree, reservation, 71%**; disagree strongly, 11%.*

Data are available from one study to support a weak correlation between quality of life and patient satisfaction in GORD[82] but quality of life scores have been reported not to predict patient satisfaction in reflux oesophagitis patients treated with on-demand therapy.[126] In other conditions, a lack of correlation has been reported for psychiatric care,[127] diabetes care,[128] and cancer nursing.[129] Patient satisfaction is determined by age, anxiety, self-perceived health status, and expectations. It is also influenced by the process and overall quality of healthcare, and so a patient may have a good quality of life but poor satisfaction if, for example, they have had a long wait for endoscopy. Thus a complete correlation may not be expected between quality of life and patient satisfaction. Patient expectation and satisfaction is addressed in more detail in the next paper.

Future directions for research

Many more randomised clinical trials need to include quality of life measures as an outcome. This would provide more data on the following areas.

- Associations between symptom improvement, changes in quality of life, and patient satisfaction. Proposition 3.13 was rejected on the basis that there is insufficient evidence to support incorporation of quality of life measurement into the disease definition of GORD. Support could come from well designed studies examining the relationship of symptom changes (both frequency and severity) with quality of life and patient satisfaction. It needs to be shown whether the addition of quality of life measurement adds anything to symptom measures.

- Relationships between changes in disease specific measures and changes in commonly used generic quality of life measures. This is needed to help resolve the tensions between responsiveness, generalisability, and the burden on research subjects of being given multiple measures.

- Role of quality of life measurement in the diagnosis of reflux disease.

- Change in quality of life over time with different therapies, and without intervention. Propositions 3.5 and 3.6 showed that there was a great need to determine when is the most appropriate time to assess quality of life in long term trials, particularly of intermittent therapy.

- Effects of long term treatment on quality of life.

In addition, the methodology of quality of life measurement itself faces challenges, particularly in the following areas.

- More research in the translation of generic quality of life measures into utility based measures and QALYs, taking into account response shift.

- Simpler measures that might be used to audit the quality of clinical care. Proposition 3.14 was strongly rejected on the basis of an absence of a simple score that might augment a clinical interview.

- More research on cross cultural and cross language validation of quality of life measures. Proposition 3.3 was accepted with reservation owing to the lack of evidence as to the degree of validation required.

4. PATIENT EXPECTATIONS AND SATISFACTION: PROPOSITIONS, VOTING, DISCUSSION, AND COMMENTARY

Introduction

The three main determinants of a patient's satisfaction with treatment are clinical outcome, interpersonal care and relationship with the physician, and the physical environment of the health care process.[130] In parallel with this, patient expectations can vary according to knowledge and prior experience, and can be changed by the physician, and may also change with accumulating experience. Moreover, patients new to a disease and the therapies available may not really know what their expectations are. Expectations are also affected by culture, socioeconomic status, personal values, attitudes, education, and knowledge.

Given this dynamic, multifactorial nature of patient expectation and satisfaction, it is not surprising that presentation of patient satisfaction results alone, without the context of efficacy and safety results, can be misleading. Patients may be happy with the process of care (for example, access to physicians, the physical environment of the facility, etc), even though their state of health is no better following treatment. Conversely, a patient's health may be greatly improved by treatment but their satisfaction is low because they are unhappy with the process of care that they experienced. This is exemplified by data from 80 GORD patients following laparoscopic fundoplication.[131] Most patients responded to a global question regarding satisfaction by indicating they were satisfied despite persistent GORD symptoms or the development of new symptoms, such as dysphagia, which were severe enough to impair quality of life measured by a disease specific instrument.

This underlines the potential lack of correlation between global patient satisfaction data and clinical outcomes. It also highlights the fact that patient satisfaction alone is an imperfect measure of disease management, particularly when patient expectations are low. In a cross sectional survey of chronic heartburn sufferers,[132] 45% of patients treated with H_2 receptor antagonists and 58% of patients receiving PPIs were totally satisfied with treatment. The low rate of satisfaction with PPI therapy is surprising, given their high efficacy rates. The study did not address the causes of satisfaction and dissatisfaction. Did patients experience inadequate symptom control or were they not happy with the process or cost of care, and what were their expectations in the first place? This raises numerous questions about the role of patient expectation and satisfaction in clinical practice and clinical trials in GORD, and the validity of the instruments currently available in this area.

Questions related to patient expectation and satisfaction, and symptom assessment in GORD, were addressed in the workshop under four topic areas.

- Why is measurement of patient satisfaction important in reflux disease?
- How should patient satisfaction be measured?
- Are patients satisfied with current therapy for reflux disease?
- How important are patient expectations in determining patient satisfaction?

Propositions, voting, and discussion
Why is measurement of patient satisfaction important in reflux disease?

(4.1) *Patient satisfaction is an important outcome measure in the treatment of reflux disease* (nature of evidence: D).

*Strength of recommendation: agree strongly, 32%; **agree, reservation, 50%**; disagree, reservation, 14%; disagree strongly, 4%.*

The proposition relates to routine clinical practice, rather than clinical trials, and recommends patient satisfaction as one of several outcome measures, not the sole measure. In this context, although there are no direct data to support it, the proposition is correct, since evaluation by the clinician of patient satisfaction, and definition of their expectations, is a routine clinical skill. It is intuitive to measure satisfaction with therapy, and to determine if reasonable expectations are being met. However, this has not been formalised into a validated tool with which to measure patient satisfaction in GORD. Such a formal measure needs to accurately reflect impacts of therapy, be referenced to appropriate patient expectations, and not be unduly influenced by psychological and functional variables that are not directly linked to reflux disease symptoms. It needs to be a validated standardised scale that can be compared across groups and studies.

Editorial comment. It is clearly desirable to have a satisfied patient in clinical practice. This vote reflects the view that while it is a useful outcome measure it cannot be important until proper tools to evaluate it become available. This view is reflected also in the rejection of the propositions related to the use of patient satisfaction as outcome measures in assessing treatment response (propositions 2.2 and 2.14).

(4.2) *Systematic use of patient satisfaction data enables choices to be made between alternatives in the organisation or provision of health care for reflux disease* (nature of evidence: E).

Strength of recommendation: agree strongly, 0%; agree, reservation, 12%; **disagree, reservation, 77%***; disagree strongly, 12%.*

Conceptually, the proposition is acceptable, but practically it is not, and it was therefore rejected. Systematic use of patient satisfaction data is not possible in the absence of validated tools with which to measure it.

Editorial comment. Patient satisfaction data are widely used by managed care organisations to provide comparisons between plans and physician groups. These may be useful measures of the process of care. There are no validated tools in GORD that can separate the process of care from clinical outcomes. This vote reflects the lack of confidence that patient satisfaction can replace clinical outcome data in making determinations about treatment choices for reflux disease.

How should patient satisfaction be measured?

(4.3) *Patient satisfaction surveys require appropriate methodology and validated instruments* (nature of evidence: B).

Strength of recommendation: **agree strongly, 96%***; agree, reservation, 0%; disagree, reservation, 4%; disagree strongly, 0%.*

Current measures of patient satisfaction are inappropriate, as they have no conceptual model, few include qualitative patient data, they are mostly simple single item scales, there are problems with response bias, and psychometric data are limited. Consequently, they result in biased measurement and decreased ability to detect small but meaningful differences in outcome. The rationale for supporting the proposition is the potential value of a validated measure of patient satisfaction, which can provide an overall assessment of health care delivery from the patient's standpoint.

Editorial comment. The nature of evidence may be an overestimate, as support of this proposition is largely based on expert opinion. Global estimates are useful to identify deficiencies in the process of care (scheduling of appointments, wait times, responsiveness of staff, etc).

(4.4) *There are several important dimensions of patient satisfaction relevant to reflux disease treatment* (nature of evidence: E).

Strength of recommendation: agree strongly, 30%; **agree, reservation, 63%***; disagree, reservation, 7%; disagree strongly, 0%.*

Treatment satisfaction in GORD decreases with increasing disease severity but there is a large disconnect between patient satisfaction and treatment outcome in that satisfaction is not related to the extent of symptom reduction (see Revicki[133] in this supplement *(page iv40–iv44)*). Thus there are dimensions other than symptom reduction that relate to patient satisfaction in GORD.

(4.5) *As patient satisfaction is multidimensional, questions related to these dimensions and disease focused questions are useful to evaluate different aspects of treatment* (nature of evidence: D).

Strength of recommendation: agree strongly, 14%; **agree, reservation, 86%***; disagree, reservation, 0%; disagree strongly, 0%.*

Indirect evidence for this comes from primary care studies which suggest that there are a number of dimensions to patient satisfaction.[134–137] Studies that address patient satisfaction therefore require evaluation of all of these dimensions, and also need disease specific outcome measures to determine the outcome of treatment.

Editorial comment. This vote emphasises the importance of measuring the many dimensions of patient satisfaction and also highlights the need for disease specific measures of clinical outcome.

Are patients satisfied with current therapy for reflux disease?

(4.6) *Absence of symptoms is a major determinant of patient satisfaction with therapy* (nature of evidence: D).

Strength of recommendation: agree strongly, 11%; agree, reservation, 18%; **disagree, reservation, 64%***; disagree strongly, 7%.*

There is a lack of evidence to support this proposition. Absence of symptoms is paralleled by improved quality of life, and patient willingness to pay is higher for absence rather than reduction of symptoms.[138] However, there is a weak relationship between reduction of symptoms and patient satisfaction (Revicki[133] in this supplement *(see page iv40–iv44)*). The proposition was therefore rejected by the majority of participants.

Editorial comment. Although there is a correlation between reduction of symptoms and satisfaction, the correlation is poor because satisfaction measures other dimensions of care that are unrelated to symptom reduction. Patient satisfaction alone is therefore not a substitute for determining the absence of symptoms and vice versa.

(4.7) *In primary care, more than one third of patients are somewhat dissatisfied with current prescription medical therapy* (nature of evidence: C).

Strength of recommendation: agree strongly, 9%; **agree, reservation, 86%***; disagree, reservation, 5%; disagree strongly, 0%.*

Patient satisfaction with outcome of therapy is a function of symptom reduction rather than healing of oesophagitis. Approximately two thirds of reflux oesophagitis patients treated with PPIs have abolition of symptoms while rates of symptom reduction are generally lower in studies of endoscope negative reflux disease. As there is no good correlation between symptom response and patient satisfaction, this may not be directly relevant to the proposition. A single study by Crawley and Schmitt, based on a cross sectional survey of approximately 20 000 chronic heartburn sufferers,[132] lends some support to the proposition. Of the 11 600 respondents, less than 60% were "totally satisfied with treatment" but not all the respondents were in primary care.

Editorial comment. The causes of dissatisfaction have not been studied and may include the cost of therapy and the

difficulty experienced in obtaining effective therapy in managed care. There are no validated tools that can be used in this setting and acquiescence bias is a potential confounding factor. The Crawley-Schmitt study[132] is interesting because it shows a difference between H_2 receptor antagonists and PPIs. Confounding factors should be similar in the two groups but access to PPI therapy is more difficult than to H_2 receptor antagonists in USA managed care settings.

(4.8) Patients with reflux disease taking PPIs are more satisfied than patients taking H_2 receptor antagonists (nature of evidence: C).

Strength of recommendation: agree strongly, 15%; **agree, reservation, 65%**; *disagree, reservation, 19%; disagree strongly, 0%.*

The superiority of symptom control with PPIs compared with H_2 receptor antagonists is well established although the poor correlation between symptom response and patient satisfaction may mean that this may not be directly relevant to the proposition. The Crawley and Schmitt cross sectional survey[132] again lends some support for the proposition, as 58% of patients taking PPI therapy were "totally satisfied with treatment" compared with 45% taking H_2 receptor antagonists.

Editorial comment. The small number of workshop participants who disagreed with this proposition were uncertain if dissatisfaction rates in the H_2 receptor antagonist group were related to poor symptom control, managed care mandates on the choice of therapy, or other factors. There were reservations about the Crawley and Schmitt study[132] because of the limited amount of information on satisfaction that was sought and the lack of data on the underlying disease (subjects were recruited from a pharmacy database).

How important are patient expectations in determining patient satisfaction?

(4.9) Patient expectations need to be evaluated and discussed before embarking on therapy (nature of evidence: E).

Strength of recommendation: **agree strongly, 76%**; *agree, reservation, 16%; disagree, reservation, 8%; disagree strongly, 0%.*

(4.10) Patient expectations may need to be modified before embarking on therapy (nature of evidence: D).

Strength of recommendation: **agree strongly, 75%**; *agree, reservation, 21%; disagree, reservation, 4%; disagree strongly, 0%.*

Patient satisfaction relates to how well their expectations are met. Indirect evidence of the impact on satisfaction of modifying expectation comes from a study of parents attending acute paediatric care who had a pre-visit desire for antibiotics to be prescribed.[139] When this expectation was discussed with parents, and information provided that the expectation might not be appropriate at that point in time, parent satisfaction was greater after the physician encounter in which antibiotics were not prescribed. These data suggest that patients may have medically incorrect expectations based on poor information. If these are not anticipated, discussed, and modified, patient satisfaction may be poor although medical treatment may be entirely appropriate. The recommendations in these two propositions are good for clinical practice. Care is needed in the conduct of randomised clinical trials of different treatment modalities to ensure that patient satisfaction results do not reflect improper expectations.

Editorial comment. Recent studies on surgery[131] suggest that patients with GORD undergoing surgery have incorrect expectations of the potential outcome. This vote reflects the importance of evaluating and addressing patient expectations.

(4.11) Measurements of patient satisfaction have not been given adequate emphasis in the evaluation of drug therapy of reflux disease (nature of evidence: B).

Strength of recommendation: **agree strongly, 52%**; *agree, reservation, 44%; disagree, reservation, 4%; disagree strongly, 0%.*

Of 152 articles identified in the systematic review by Sharma *et al* of outcome measures in reflux disease, only three randomised clinical trials (2%) measured patient satisfaction (Sharma and colleagues[1] in this supplement **(see page iv58–iv65)**). Where assessments have been made, they have not been adequate or validated.[112 140] This again underlines the need for validated instruments to measure patient satisfaction in GORD.

Editorial comment. This vote reflects the desire of most physicians to have a satisfied patient. Unfortunately, there are no validated instruments to perform such measurements at the present time.

Approximately half of the participants disagreed with the proposition that, in clinical trials, a validated measure of patient satisfaction with heartburn control is an important outcome measure for evaluation of long term treatment (proposition 2.14), largely because no validated instrument exists. However, this clouds the issue that multidimensional validated satisfaction scales are needed and that, when available, they will be of value. This was confirmed by broad acceptance of a subsequent proposition (4.12).

(4.12) A valid and responsive treatment satisfaction scale should be used to assess secondary outcomes in clinical trials in GORD.

Strength of recommendation: **agree strongly, 59%**; *agree, reservation, 33%; disagree, reservation, 7%; disagree strongly, 0%.*

(4.13) There is a virtual absence of rigorous evaluation of patients' satisfaction with antireflux surgery and other physical antireflux therapies (nature of evidence: B).

Strength of recommendation: **agree strongly, 81%**; *agree, reservation, 19%; disagree, reservation, 0%; disagree strongly, 0%.*

The Visick and modified Visick classifications have dominated the literature although they were designed for gastric ulcer surgery. They have not been validated in GORD and there is no uniform evaluation and no independent observer, as patients' satisfaction may often be assessed by the physician who performed the surgical procedure, which introduces a strong acquiescence bias. Global measures of patient satisfaction have not included a rigorous assessment of clinical outcomes.

Editorial comment. This vote reflects the need for patient satisfaction outcomes to be considered along with functional and clinical outcomes in GORD. Patient satisfaction is a limited and secondary measure of the outcome of surgery in GORD, and must be considered with other outcome measures, such as symptom reduction, side effects, etc.

(4.14) Patient satisfaction is a useful measure for the evaluation of treatment algorithms developed by funders of health care (nature of evidence: E).

Strength of recommendation: agree strongly, 11%; **agree, reservation, 52%**; *disagree, reservation, 30%; disagree strongly, 7%.*

Funders of health care may have a different agenda to patients. Switching of PPI maintenance therapy in GORD patients in a Veteran's Administration health care system resulted in significant cost savings but the majority of patients preferred the original PPI.[141] While patient satisfaction may be one useful measure for the evaluation of treatment algorithms, provided it can be measured adequately, it is not the only aspect to be taken into account, and its hierarchy in relation to other aspects is currently unclear. However, there may be treatment algorithms put forward by

health care providers where patient satisfaction is the only way of addressing them.

Editorial comment. The split vote reflects the fact that patient satisfaction may be too general a measure of a treatment algorithm, and that it may overemphasise the process of care. The lack of a validated instrument and absence of data correlating this measure with clinical outcome were causes for concern. Patient satisfaction data could be misused to drive less expensive and less effective therapy.

(4.15) *From the patient's perspective, the ideal outcome of therapy is the abolition of all symptoms, without the introduction of new ones from the therapy itself* (nature of evidence: E).

*Strength of recommendation: agree strongly, 21%; **agree, reservation, 71%**; disagree, reservation, 7%; disagree strongly, 0%.*

Absence of symptoms is paralleled by improved quality of life, and patient willingness to pay is higher for absence rather than reduction of symptoms.[138] However, these studies do not directly address the proposition, which was accepted based on expert opinion.

Editorial comment. While intuitive, this aspect of measurement has been neglected in many studies of surgery or endoscopic therapy. The impact of side effects that develop after the procedure can have a profound impact on quality of life[131] that needs to be considered in the overall evaluation of the treatment modality.

Future directions for research

Several areas of future research were identified. There is a pressing need for validated instruments that measure patient satisfaction. There is a need to measure patient expectations of GORD therapy, and to determine if patient expectations can be modified to adjust to the realities of the treatment available (for example, patients may desire a cure but may have to settle for maintenance therapy). Studies on surgical and endoscopic intervention have used the most rudimentary measures of patient satisfaction, and the development of a multidimensional disease specific instrument is a critical need to allow evaluation of these therapies.

. . . .

Authors' affiliations
J Dent, Department of Gastroenterology, Hepatology, and General Medicine, Royal Adelaide Hospital, Adelaide, Australia
D Armstrong, Division of Gastroenterology, McMaster University Medical Centre, Hamilton, Canada
B Delaney, P Moayyedi, Department of Primary Care and General Practice, University of Birmingham, Birmingham, UK
N J Talley, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, USA
N Vakil, University of Wisconsin Medical School, Milwaukee, USA

REFERENCES
1 **Sharma N**, Donnellan C, Preston C, et al. A systematic review of symptomatic outcomes used in oesophagitis drug therapy trials. *Gut* 2004; **53**(suppl IV):iv58–65.
2 **Shaw M**. Diagnostic utility of reflux disease symptoms. *Gut* 2004; **53**(suppl IV):iv25–7.
3 **McColl E**. Best practice in symptom assessment: a review. *Gut* 2004; **53**(suppl IV):iv49–54.
4 **Wyrwich KW**, Staebler Tardino VM. A blueprint for symptom scales and responses: measurement and reporting. *Gut* 2004;**53**(suppl IV):iv45–8.
5 **Kay L**, Jorgensen T, Jensen KH. Epidemiology of abdominal symptoms in a random population: prevalence, incidence, and natural history. *Eur J Epidemiol* 1994;**10**:559–66.
6 **Talley NJ**, Meineche-Schmidt V, Pare P, et al. Efficacy of omeprazole in functional dyspepsia: double-blind, randomized, placebo-controlled trials (the Bond and Opera studies). *Aliment Pharmacol Ther* 1998;**12**:1055–65.
7 **Lagergren J**, Bergström R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med* 1999;**340**:825–31.
8 **Dent J**, Brun J, Fendrick AM, et al. An evidence-based appraisal of reflux disease management—the Genval Workshop Report. *Gut* 1999;**44**(suppl 2):S1–16.
9 **Lundell L**, Dent J, Bennett JR, et al. Endoscopic assessment of oesophagitis—clinical and functional correlates and further validation of the Los Angeles classification. *Gut* 1999;**45**:172–80.
10 **Junghard O**, Carlsson R, Lind T. Sufficient control of heartburn in endoscopy negative GORD trials. *Scand J Gastroenterol* 2003 (in press).
11 **Klauser AG**, Schindlbeck NE, Müller-Lissner SA. Symptoms in gastro-oesophageal reflux disease. *Lancet* 1990;**335**:205–8.
12 **Schindlbeck NE**, Klauser AG, Voderholzer WA, et al. Empiric therapy for gastroesophageal reflux disease. *Arch Intern Med* 1995;**155**:1808–12.
13 **Schenk BE**, Kuipers EJ, Klinkenberg-Knol EC, et al. Omeprazole as a diagnostic tool in gastroesophageal reflux disease. *Am J Gastroenterol* 1997;**92**:1997–2000.
14 **Johnsson F**, Weywadt L, Solhaug J-H, et al. One-week omeprazole treatment in the diagnosis of gastro-oesophageal reflux disease. *Scand J Gastroenterol* 1998;**33**:15–20.
15 **Fass R**, Ofman JJ, Gralnek IM, et al. Clinical and economic assessment of the omeprazole test in patients with symptoms suggestive of gastroesophageal reflux disease. *Arch Intern Med* 1999;**159**:2161–8.
16 **Moayyedi P**, Forman D, Braunholtz D, et al. The proportion of upper gastrointestinal symptoms in the community associated with *Helicobacter pylori*, lifestyle factors and non-steroidal anti-inflammatory drugs. *Am J Gastroenterol* 2000;**95**:1448–55.
17 **Castell DO**, Kahrilas PJ, Richter JE, et al. Esomeprazole (40 mg) compared with lansoprazole (30 mg) in the treatment of erosive esophagitis. *Am J Gastroenterol* 2002;**97**:575–83.
18 **Lassen AT**, Pedersen FM, Bytzer P, et al. Helicobacter pylori test-and-eradicate versus prompt endoscopy for management of dyspeptic patients: A randomised trial *Lancet* 2000;**356**:455–60.
19 **Farup PG**, Hovde O, Torp R, et al. Patients with functional dyspepsia responding to omeprazole have a characteristic gastro-oesophageal reflux pattern. *Scand J Gastroenterol* 1999;**34**:575–9.
20 **Earlam RJ**. Production of epigastric pain in duodenal ulcer by lower oesophageal acid perfusion. *BMJ* 1970;**iv**:714–16.
21 **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.
22 **Carlsson R**, Dent J, Watts R, et al. Gastro-oesophageal reflux disease in primary care: An international study of different treatment strategies with omeprazole. International GORD Study Group. *Eur J Gastroenterol Hepatol* 1998;**10**:119–24.
23 **Johnsen R**, Bernersen B, Straume B, et al. Prevalences of endoscopic and histological findings in subjects with and without dyspepsia. *BMJ* 1991;**302**:749–52.
24 **Johnsson F**, Roth Y, Damgaard Pedersen N-E, et al. Cimetidine improves GERD symptoms in patients selected by a validated GERD questionnaire. *Aliment Pharmacol Ther* 1993;**7**:81–6.
25 **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.
26 **Moayyedi P**, Axon AT. The usefulness of the likelihood ratio in the diagnosis of dyspepsia and gastroesophageal reflux disease. *Am J Gastroenterol* 1999;**94**:3122–5.
27 **Locke GR**, Talley NJ, Weaver AL, et al. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;**69**:539–47.
28 **Hansen JM**, Bytzer P, Schaffalitzky De Muckadell OB. Management of dyspeptic patients in primary care. Value of the unaided clinical diagnosis and of dyspepsia subgrouping. *Scand J Gastroenterol* 1998;**33**:799–805.
29 **Johnsson F**, Joelsson B, Gudmundsson K, et al. Symptoms and endoscopic findings in the diagnosis of gastroesophageal reflux disease. *Scand J Gastroenterol* 1987;**22**:714–18.
30 **Chow WH**, Finkle WD, McLaughlin JK, et al. The relation of gastroesophageal reflux disease and its treatment to adenocarcinomas of the esophagus and gastric cardia. *JAMA* 1995;**274**:474–7.
31 **Kahrilas PJ**, Quigley EM. Clinical esophageal pH recording: a technical review for practice guideline development. *Gastroenterology* 1996;**110**:1982–96.
32 **Ho KY**, Ng WL, Kang JY, et al. Gastroesophageal reflux disease is a common cause of noncardiac chest pain in a country with a low prevalence of reflux esophagitis. *Dig Dis Sci* 1998;**43**:1991–7.
33 **Fass R**, Fennerty MB, Ofman JJ, et al. The clinical and economic value of a short course of omeprazole in patients with noncardiac chest pain. *Gastroenterology* 1998;**115**:42–9.
34 **Achem SR**, Kolts BE, MacMath T, et al. Effects of omeprazole versus placebo in treatment of noncardiac chest pain and gastroesophageal reflux. *Dig Dis Sci* 1997;**42**:2138–45.
35 **Chambers J**, Cooke R, Anggiansah A, et al. Effect of omeprazole in patients with chest pain and normal coronary anatomy: initial experience. *Int J Cardiol* 1998;**65**:51–5.
36 **So JB**, Zeitels SM, Rattner DW. Outcomes of atypical symptoms attributed to gastroesophageal reflux treated by laparoscopic fundoplication. *Surgery* 1998;**124**:28–32.
37 **Rothwell JF**, Feehan E, Reid I, et al. Delay in treatment for oesophageal cancer. *Br J Surg* 1997;**84**:690–3.
38 **Ruth M**, Månsson I, Sandberg N. The prevalence of symptoms suggestive of esophageal disorders. *Scand J Gastroenterol* 1991;**26**:73–81.
39 **Lindgren S**, Janzon L. Prevalence of swallowing complaints and clinical findings among 50–79-year-old men and women in an urban population. *Dysphagia* 1991;**6**:187–92.

40 **Isolauri J**, Laippala P. Prevalence of symptoms suggestive of gastro-oesophageal reflux disease in an adult population. *Ann Med* 1995;**27**:67–70.

41 **Vakil NB**, Zuckerman S, Levine JG. Dysphagia in uncomplicated reflux disease. *Gastroenterology* 2002;**122**:A583.

42 **Wallace MB**, Durkalski VL, Vaughan J, *et al.* Age and alarm symptoms do not predict endoscopic findings among patients with dyspepsia: a multicentre database study. *Gut* 2001;**49**:29–34.

43 **Adang RP**, Vismans JF-J, Talmon JL, *et al.* Appropriateness of indications for diagnostic upper gastrointestinal endoscopy: association with relevant endoscopic disease. *Gastrointest Endosc* 1995;**42**:390–7.

44 **Numans ME**, van der Graaf Y, de Wit NJ, *et al.* How useful is selection based on alarm symptoms in requesting gastroscopy? An evaluation of diagnostic determinants for gastro-oesophageal malignancy. *Scand J Gastroenterol* 2001;**36**:437–43.

45 **Madsen LG**, Bytzer P. The value of alarm features in identifying organic causes of dyspepsia. *Canad J Gastroenterol* 2000;**14**:713–20.

46 **Csendes A**, Smok G, Burdiles P, *et al.* Prevalence of Barrett's esophagus by endoscopy and histologic studies: a prospective evaluation of 306 control subjects and 376 patients with symptoms of gastroesophageal reflux. *Dis Esophagus* 2000;**13**:5–11.

47 **Ronkainen J**, Aro P, Storskrubb T, *et al.* Prevalence of esophagitis and endoscopy-negative reflux disease in a population. A report from the Kalixanda study. *Gastroenterology* 2002;**122**:A213.

48 **Robinson M**, Earnest D, Rodriguez-Stanley S, *et al.* Heartburn requiring frequent antacid use may indicate significant illness. *Arch Intern Med* 1998;**158**:2373–6.

49 **Winters C Jr**, Spurling TJ, Chobanian SJ, *et al.* Barrett's esophagus: a prevalent, occult complication of gastroesophageal reflux disease. *Gastroenterology* 1987;**92**:118–24.

50 **Lieberman DA**, Oehlke M, Helfand M. Risk factors for Barrett's esophagus in community-based practice. GORGE consortium. Gastroenterology Outcomes Research Group in Endoscopy. *Am J Gastroenterol* 1997;**92**:1293–7.

51 **Romero Y**, Cameron AJ, Schaid DJ, *et al.* Barrett's esophagus: prevalence in symptomatic relatives. *Am J Gastroenterol* 2002;**97**:1127–32.

52 **Farrow DC**, Vaughan TL, Sweeney C, *et al.* Gastroesophageal reflux disease, use of H2 receptor antagonists, and risk of esophageal and gastric cancer. *Cancer Causes Control* 2000;**11**:231–8.

53 **Cameron AJ**, Lomboy CT. Barrett's esophagus: age, prevalence, and extent of columnar epithelium. *Gastroenterology* 1992;**103**:1241–5.

54 **Bytzer P**, Christensen PB, Damkier P, *et al.* Adenocarcinoma of the esophagus and Barrett's esophagus: a population-based study. *Am J Gastroenterol* 1999;**94**:86–91.

55 **Blot WJ**, Devesa SS, Kneller RW, *et al.* Rising incidence of adenocarinoma of the esophagus and gastric cardia. *JAMA* 1991;**265**:1287–9.

56 **Conio M**, Cameron AJ, Romero Y, *et al.* Secular trends in the epidemiology and outcome of Barrett's oesophagus in Olmsted County, Minnesota. *Gut* 2001;**48**:304–9.

57 **Powell J**, McConkey CC, Gillison EW, *et al.* Continuing rising trend in oesophageal adenocarcinoma. *Int J Cancer* 2002;**102**:422–7.

58 **Powell J**, McConkey CC. The rising trend in oesophageal adenocarcinoma and gastric cardia. *Eur J Cancer Prev* 1992;**1**:265–9.

59 **Vizcaino AP**, Moreno V, Lambert R, *et al.* Time trends incidence of both major histologic types of esophageal carcinomas in selected countries, 1973–1995. *Int J Cancer* 2002;**99**:860–8.

60 **Wijnhoven BP**, Louwman MW, Tilanus HW, *et al.* Increased incidence of adenocarcinomas at the gastro-esophageal junction in Dutch males since the 1990s. *Eur J Gastroenterol* 2002;**14**:115–22.

61 **Blot WJ**, Devesa SS, Fraumeni JF Jr. Continuing climbing rates of esophageal adenocarcinoma: an update. *JAMA* 1993;**270**:1320.

62 **Farrow DC**, Vaughan TL, Hansten PD, *et al.* Use of aspirin and other nonsteroidal anti-inflammatory drugs and risk of esophageal and gastric cancer. *Cancer Epidemiol Biomarkers Prev* 1998;**7**:97–102.

63 **van Blankenstein M**. Barrett's esophagus: so what! *Dis Esophagus* 2002;**15**:1–4.

64 **Bani-Hani K**, Sue-Ling H, Johnston D, *et al.* Barrett's oesophagus: results from a 13-year surveillance programme. *Eur J Gastroenterol Hepatol* 2000;**12**:649–54.

65 **Forman D**. Counting cancers at the junction—a problem of routine statistics. *Eur J Gastroenterol Hepatol* 2002;**14**:99–101.

66 **Shaheen NJ**, Crosby MA, Bozymski EM, *et al.* Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology* 2000;**119**:333–8.

67 **Gerson LB**, Shetler K, Triadafilopoulos G. Prevalence of Barrett's esophagus in asymptomatic individuals. *Gastroenterology* 2002;**123**:461–7.

68 **Armstrong D**, Barkun ANG, Chiba N, *et al.* Initial management strategies for reflux-predominant uninvestigated dyspepsia (UD) in primary care practice—the CADET-HR study. *Am J Gastroenterol* 2001;**96**:S3.

69 **Armstrong D**, Barkun AN, Chiba N, *et al.* and CADET-HR Investigators. On demand therapy for heartburn-dominant uninvestigated dyspepsia in primary care practice: the CADET-HR Study. *Can J Gastroenterol* 2003;**17**(suppl A):105A.

70 **Lind T**, Havelund T, Carlsson R, *et al.* Heartburn without oesophagitis: efficacy of omeprazole therapy and features determining therapeutic response. *Scand J Gastroenterol* 1997;**32**:974–9.

71 **Venables TL**, Newland RD, Patel AC, *et al.* Omeprazole 10 milligrams once daily, omeprazole 20 milligrams once daily, or ranitidine 150 milligrams twice daily, evaluated as initial therapy for the relief of symptoms of gastro-oesophageal reflux disease in general practice. *Scand J Gastroenterol* 1997;**32**:965–73.

72 **Bate CM**, Griffin SM, Keeling PW, *et al.* Reflux symptom relief with omeprazole in patients without unequivocal oesophagitis. *Aliment Pharmacol Ther* 1996;**10**:547–55.

73 **Frank L**, Kleinman L, Ganoczy D, *et al.* Upper gastrointestinal symptoms in North America—Prevalence and relationship to healthcare utilization and quality of life. *Dig Dis Sci* 2000;**45**:809–18.

74 **Kahrilas PJ**, Falk GW, Johnson DA, *et al.* Esomeprazole improves healing and symptom resolution as compared with omeprazole in reflux oesophagitis patients: a randomized controlled trial. The Esomeprazole Study Investigators. *Aliment Pharmacol Ther* 2000;**14**:1249–58.

75 **Richter JE**, Kahrilas PJ, Johanson J, *et al.* Efficacy and safety of esomeprazole compared with omeprazole in GERD patients with erosive esophagitis: a randomized controlled trial. *Am J Gastroenterol* 2001;**96**:656–65.

76 **Johnson DA**, Vakil N, Richter JE, *et al.* Sustained heartburn resolution during the first 4 weeks of PPI therapy predicts subsequent healing in GERD patients with erosive esophagitis. *Gut* 2002;(suppl III):A166, Abs TUE-G-307.

77 **Hyland ME**, Kenyon CA, Allen R, *et al.* Diary keeping in asthma: comparison of written and electronic methods. *BMJ* 1993;**306**:487–9.

78 **Stone AA**, Shiffman S, Schwartz JE, *et al.* Patient non-compliance with paper diaries. *BMJ* 2002;**324**:1193–4.

79 **Guyatt GH**, Townsend M, Berman LB, *et al.* A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis* 1987;**40**:1129–33.

80 **Ferraz M**, Quaresma M, Aquino L, *et al.* Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *J Rheumatol* 1990;**17**:1022–4.

81 **Huskisson E**. Measurement of pain. *Lancet* 1974;**2**:1127–31.

82 **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.

83 **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.

84 **Veldhuyzen van Zanten SJO**, Chiba N, Armstrong D, *et al.* The responsiveness of the Global Overall Symptom (GOS) score—a 7-point Likert scale for the measurement of dyspepsia symptoms in clinical trials. *Can J Gastroenterol* 2003;**17**(suppl A):113A.

85 **Junghard O**, Wiklund IK. A comparison of patient and clinician ratings of heartburn. *Gastroenterology* 2003;**124**(4 suppl 1):A541, Abs T1653.

86 **Armstrong D**, Paré P, Pericak D, *et al.* Symptom relief in gastroesophageal reflux disease: a randomized, controlled comparison of pantoprazole and nizatidine in a mixed patient population with erosive esophagitis or endoscopy-negative reflux disease. The Canadian Pantoprazole GERD Study Group. *Am J Gastroenterol* 2001;**96**:2849–57.

87 **Eisendrath P**, Tack J, Deviere J. Diagnosis of gastroesophageal reflux disease in general practice: a Belgian national survey. *Endoscopy* 2002;**34**:998–1003.

88 **Allgood PC**, Bachmann M. Medical or surgical treatment for chronic gastroesophageal reflux? A systematic review of published evidence of effectiveness. *Eur J Surg* 2000;**166**:713–21.

89 **Sandbu R**, Khamis H, Gustavsson S, *et al.* Long-term results of antireflux surgery indicate the need for a randomized clinical trial. *Br J Surg* 2002;**89**:225–30.

90 **Lind T**, Havelund T, Lundell L, *et al.* On demand therapy with omeprazole for the long-term management of patients with heartburn without oesophagitis—a placebo-controlled randomized trial. *Aliment Pharmacol Ther* 1999;**13**:907–14.

91 **Talley NJ**, Venables TL, Green JR, *et al.* Esomeprazole 40 mg and 20 mg is efficacious in the long-term management of patients with endoscopy-negative gastro-oesophageal reflux disease: a placebo-controlled trial of on-demand therapy for 6 months. *Eur J Gastroenterol Hepatol* 2002;**14**:857–63.

92 **Spechler SJ**, Lee E, Ahnen D, *et al.* Long-term outcome of medical and surgical treatments of gastroesophageal reflux disease. Follow up of a randomized controlled trial. *JAMA* 2001;**285**:2331–8.

93 **Lord RV**, Kaminski A, Oberg S, *et al.* Absence of gastroesophageal reflux disease in a majority of patients taking acid suppression medications after Nissen fundoplication. *J Gastrointest Surg* 2002;**6**:3–9.

94 **Carlsson R**, Galmiche J-P, Dent J, *et al.* Prognostic factors influencing relapse of oesophagitis during maintenance therapy with antisecretory drugs: a meta-analysis of long-term omeprazole trials. *Aliment Pharmacol Ther* 1997;**11**:473–82.

95 **Talley NJ**, Armstrong D, Junghard O, *et al.* Predictors of treatment response in patients with endoscopy-negative reflux disease (ENRD). *Am J Gastroenterol* 2002;**97**(suppl 1):S12.

96 **Fass R**, Ofman JJ, Sampliner RE, *et al.* The omeprazole test is as sensitive as 24-h oesophageal pH monitoring in diagnosing gastro-oesophageal reflux disease in symptomatic patients with erosive oesophagitis. *Aliment Pharmacol Ther* 2000;**14**:389–96.

97 **Fass R**. Empirical trials in treatment of gastroesophageal reflux disease. *Dig Dis* 2000;**18**:20–6.

98 **Johanson JF**, Minor P, Orr W, *et al.* Efficacy of rabeprazole for heartburn and associated symptoms in non-erosive gastroesophageal reflux disease. *Gastroenterology* 2002;**122**:A1294.

99 **Vakil NB**, Zuckerman S, Levine JG, *et al.* Nocturnal heartburn in patients with erosive esophagitis and its resolution with proton pump inhibitors. *Gastroenterology* 2002;**122**:A-200 (S1291).

100 **Fink P**. Surgery and medical treatment in persistent somatizing patients. *J Psychosom Res* 1992;**36**:439–47.

101 **Armstrong D**, Barkun AN, Chiba N, *et al*. Predictors of treatment response and relapse for heartburn-dominant uninvestigated dyspepsia in primary care practice: the CADET-HR study. *Am J Gastroenterol* 2002;**97**(suppl 1):S36.

102 **Harding SM**, Richter JE, Guzzo MR, *et al*. Asthma and gastroesophageal reflux: acid suppressive therapy improves asthma outcome. *Am J Med* 1996;**100**:395–405.

103 **Wong RK**, Hanson DG, Waring PJ, *et al*. ENT manifestations of gastroesophageal reflux. *Am J Gastroenterol* 2000;**95**(suppl):S15–22.

104 **Harding SM**, Sontag SJ. Asthma and gastroesophageal reflux. *Am J Gastroenterol* 2000;**95**(suppl):S23–32.

105 **Irwin RS**, Richter JE. Gastroesophageal reflux and chronic cough. *Am J Gastroenterol* 2000;**95**(suppl):S9–14.

106 **Belafsky PC**, Postma GN, Koufman JA. Laryngopharyngeal reflux symptoms improve before changes in physical findings. *Laryngoscope* 2001;**111**:979–81.

107 **Chiba N**, De Gara C, Wilkinson JM, *et al*. Speed of healing and symptom relief in grade II to IV gastroesophageal reflux disease: a meta-analysis. *Gastroenterology* 1997;**112**:1798–810.

108 **Branski RC**, Bhattacharyya N, Shapiro J. The reliability of the assessment of endoscopic laryngeal findings associated with laryngopharyngeal reflux disease. *Laryngoscope* 2002;**112**:1019–24.

109 **Ours TM**, Kavuru MS, Schilz RJ, *et al*. A prospective evaluation of esophageal testing and a double-blind, randomized study of omeprazole in a diagnostic and therapeutic algorithm for chronic cough. *Am J Gastroenterol* 1999;**94**:3131–8.

110 **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.

111 **Kulig M**, Leodolter AL, Schulte E, *et al*. Quality of life in patients with gastroesophageal reflux disease. *Qual Life Res* 2002;**11**:649.

112 **Mathias SD**, Colwell HH, Miller DP, *et al*. Health-related quality-of-life and quality-days incrementally gained in symptomatic nonerosive GERD patients treated with lansoprazole or ranitidine. *Dig Dis Sci* 2001;**46**:2416–23.

113 **Wiklund I**, Dimenäs E, Wahl M. Factors of importance when evaluating quality of life in clinical trials. *Control Clin Trials* 1990;**11**:169–79.

114 **Farup C**, Kleinman L, Sloan S, *et al*. The impact of nocturnal symptoms associated with gastroesophageal reflux disease on health-related quality of life. *Arch Intern Med* 2001;**161**:45–52.

115 **Laycock WS**, Mauren S, Waring JP, *et al*. Improvement in quality of life measures following laparoscopic antireflux surgery. *Gastroenterology* 1995;**108**(suppl 4):A1228.

116 **Guillemin F**, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;**46**:1417–32.

117 **Szabo S**, Orley J, Saxena S. An approach to response scale development for cross-cultural questionnaires. The WHOQOL Group. *Eur Psychologist* 1997;**2**:270–6.

118 **Stacey JH**, Miocevich ML, Sacks GE. The effect of ranitidine (as effervescent tablets) on the quality of life of GORD patients. *Br J Clin Pract* 1996;**50**:190–4.

119 **Talley NJ**, Junghard O, Wiklund I. Why do patients with gastroesophageal reflux disease have a poor health-related quality of life? *Gastroenterology* 2001;**120**:A423.

120 **Lauritsen K**, Junghard O, Eklund S, *et al*. The effect of esomeprazole 40 mg on healing, reflux symptoms and quality of life in patients with reflux esophagitis. *Gastroenterology* 2002;**122**(suppl):A200.

121 **McDowell I**, Newell, C. *Measuring health: A guide to rating scales and questionnaires*, 2nd edn. New York: Oxford University Press, 1996:23–4.

122 **Koch T**. Life quality vs the 'quality of life': assumptions underlying prospective quality of life instruments in health care planning. *Soc Sci Med* 2000;**51**:419–27.

123 **O'Connor AM**, Rostom A, Fiset V, *et al*. Decision aids for patients facing health treatment or screening decisions: systematic review. *BMJ* 1999;**319**:731–4.

124 **Havelund T**, Lind T, Wiklund I, *et al*. Quality of life in patients with heartburn but without esophagitis: effects of treatment with omeprazole. *Am J Gastroenterol* 1999;**94**:1782–9.

125 **Revicki DA**, Wood M, Maton PN, *et al*. The impact of gastroesophageal reflux disease on health-related quality of life. *Am J Med* 1998;**104**:252–8.

126 **Wilhelmsen I**, Hatlebakk JG, Olafsson S, *et al*. On demand therapy of reflux oesophagitis—a prospective study of symptoms, patient satisfaction and quality of life. *Aliment Pharmacol Ther* 1999;**13**:1035–40.

127 **Wiersma D**, van Busschbach J. Are needs and satisfaction of care associated with quality of life? An epidemiological survey among the severely mentally ill in the Netherlands. *Eur Arch Psychiatry Clin Neurosci* 2001;**251**:239–46.

128 **Redekop WK**, Koopmanschap MA, Stolk RP, *et al*. Health-related quality of life and treatment satisfaction in Dutch patients with type 2 diabetes. *Diabetes Care* 2002;**25**:458–63.

129 **Elf M**, Wikblad K. Satisfaction with information and quality of life in patients undergoing chemotherapy for cancer. The role of individual differences in information preference. *Cancer Nurs* 2001;**24**:351–6.

130 **Muir-Gray JA**. *Evidence-based health care*. New York: Churchill Livingstone, 1997:126–9.

131 **Vakil N**, Kirby R, Shaw M. Clinical effectiveness of laparoscopic fundoplication in a US community. *Am J Med* 2003;**114**:1–5.

132 **Crawley JA**, Schmitt CM. How satisfied are chronic heartburn suffers with their prescription medications? Results of the patient unmet needs survey. *J Clin Outcomes Manag* 2000;**7**:29–34.

133 **Revicki DA**. Patient assessment of treatment satisfaction: methods and practical issues. *Gut* 2004;**53**(suppl IV):iv40–4.

134 **Lydeard S**, Jones R. Factors affecting the decision to consult with dyspepsia: comparison of consulters and non-consulters. *J R Coll Gen Pract* 1989;**39**:495–8.

135 **Hungin APS**, Thomas PR, Bramble M, *et al*. What happens to patients following open access gastroscopy. An outcome study from general practice. *Br J Gen Pract* 1994;**44**:519–21.

136 **Delaney BC**. Why do dyspeptic patients over the age of 50 consult their general practitioner? A qualitative investigation of health beliefs relating to dyspepsia. *Br J Gen Pract* 1998;**48**:1481–5.

137 **Westbrook JI**, McIntosh J, Talley NJ. Factors associated with consulting medical or non-medical practitioners for dyspepsia: an Australian population-based study. *Aliment Pharmacol Ther* 2000;**14**:1581–8.

138 **Kleinman L**, McIntosh E, Ryan M, *et al*. Willingness to pay for complete symptom relief of gastroesophageal reflux disease. *Arch Intern Med* 2002;**162**:1361–6.

139 **Mangione-Smith R**, McGlynn EA, Elliott MN, *et al*. Parent expectations for antibiotics, physician-parent communication, and satisfaction. *Arch Pediatr Adolesc Med* 2001;**155**:800–6.

140 **Bate CM**, Green JR, Axon AT, *et al*. Omeprazole is more effective than cimetidine for the relief of all grades of gastro-oesophageal reflux disease-associated heartburn, irrespective of the presence or absence of endoscopic oesophagitis. *Aliment Pharmacol Ther* 1997;**11**:755–63.

141 **Condra LJ**, Morreale AP, Stolley SN, *et al*. Assessment of patient satisfaction with a formulary switch from omeprazole to lansoprazole in gastroesophageal reflux disease maintenance therapy. *Am J Manag Care* 1999;**5**:631–8.