A systematic review of symptomatic outcomes used in oesophagitis drug therapy trials

N Sharma, C Donnellan, C Preston, B Delaney, G Duckett, P Moayyedi

Symptoms are an important outcome for measurement in clinical trials into gastro-oesophageal reflux disease, but the optimal approach to symptom assessment has not been formally evaluated. The authors conducted a systematic review to assess how reflux symptoms have been evaluated and how well these correlate with oesophagitis healing and relapse.

SUMMARY
The Cochrane Controlled Trials Register, MEDLINE, EMBASE, and CINAHL electronic databases were searched for randomised controlled trials of drug therapies in reflux oesophagitis. Experts in the field and pharmaceutical companies were contacted for information on any unpublished eligible trials. Predefined eligibility and validity criteria determined inclusion of studies in this analysis. Data were extracted on the symptom assessment methodologies used, including scales, methods of data collection, duration of assessments, individual and global symptoms assessed, whether reduction or absence of symptoms was the main outcome measure, types of symptoms assessed, and frequency and severity of symptoms. The proportions of patients with a successful outcome according to these different symptom measures were compared with the proportions of patients in whom oesophagitis had healed (initial therapy studies) or in whom oesophagitis had not relapsed (maintenance studies). The results are primarily evaluated in the form of L’Abbé plots.

Data were extractable from 140 eligible trials. Absent or minimal symptoms correlated well with oesophagitis healing and absence during maintenance therapy, whereas symptom reduction overestimated treatment effects. Trials that measured symptoms over a stated time period showed better correlation of symptom status with oesophagitis healing and relapse than those that did not define the time period. Heartburn was the most useful symptom for prediction of oesophagitis healing but the L’Abbé plots suggested additional information may also be obtained from regurgitation and dysphagia.

Published trials of drug therapy of reflux oesophagitis use a wide range of symptom outcome measures. Comparisons with oesophagitis status suggest that some of the measures used are suboptimal. Methodologies used for acquisition of symptom data are also diverse, and frequently inconsistent with general principles derived from formal research into the processes of symptom evaluation.

INTRODUCTION
Trials that evaluate the efficacy of drug treatments in reflux oesophagitis patients use healing of oesophagitis as the main outcome measure. Although many oesophagitis classifications exist, there has been progress recently with critical evaluation and standardisation of the severity grading of oesophagitis so that results from different studies can be compared. More recently, the major impacts of reflux induced symptoms have been better recognised, and, indeed, these are the primary outcome measures for trials on endoscopy negative reflux disease. Unfortunately, there have been no attempts to optimise, and then to standardise methods of symptom evaluation in the assessment of gastro-oesophageal reflux disease (GORD). It is, therefore, difficult to compare studies. Trials have measured single symptoms, such as heartburn, or have evaluated global upper gastrointestinal symptoms. Treatment success has been variably defined as a subjective reduction or a complete absence of symptoms. There has been also variation in the time periods over which symptoms have been assessed, the scales used, and whether information was obtained from diary cards, self administered questionnaires, or investigator interview. These issues are pertinent to trials that assess the efficacy of therapy for initial treatment of GORD and for subsequent maintenance of patients in remission.

We are currently conducting Cochrane systematic reviews1 2 into the efficacy of medical therapies for the treatment of GORD. The primary aim of these reviews is to determine the most efficacious therapy for reflux disease. In this study, we have used these databases to gain a comprehensive overview of the outcome measures that have been used in these trials and to examine which of these appears to be the most appropriate. For this, we have compared the outcomes of the different methods of symptom assessment with the reference outcome of healing and prevention of relapse of oesophagitis, the most objective end point of successful treatment of GORD that is available.

Abbreviations: GORD, gastro-oesophageal reflux disease; PGWBI, psychological general well being index; SF-36, short form 36
METHODS
This study used data collected from two systematic reviews of short term therapy for reflux oesophagitis4 and of long term maintenance therapy in oesophagitis patients.2 Though this latter review also collates data from endoscopy negative reflux disease patients, data from this patient group have not been used in this analysis.

Randomised controlled trials were identified from the Cochrane Controlled Trials Register, MEDLINE (1966–2000), EMBASE (1988–2000), and CINAHL (1982–2000). The search terms used have been outlined elsewhere.1 Papers identified by the search strategy were reviewed for eligibility (see below) by two researchers using predefined criteria. The researchers’ initial judgements were blinded to the conclusions on eligibility reached by their colleague. Any discrepancies between the two researchers were then resolved by discussion of the paper. Eligible trials underwent data extraction using a predesigned data extraction form, which was checked by a second reviewer.

Short term trials
Initial therapy trials on patients with oesophagitis were potentially eligible if they assessed symptoms 2–12 weeks after start of therapy. The interventions evaluated were proton pump inhibitors, H2 receptor antagonists, prokinetics, or sucralfate, and these were compared either with each other or with placebo, with or without antacids.

Long term or maintenance trials
Long term trials were eligible for inclusion if they were randomised controlled trials of maintenance therapy in patients in whom initial therapy had healed reflux oesophagitis.3 The same drug therapies as for initial therapy were allowed and patients had to have taken at least 12 weeks of continuous therapy. Symptom assessment had to occur within 12 and 52 weeks.

Statistical analysis
Data were extracted on symptom scales used, methods of data collection, durations of assessment, whether individual or global symptoms were used, whether reduction or absence of symptoms was the main outcome measure, the types of symptoms assessed, whether frequency and/or severity were evaluated, and whether quality of life and patient satisfaction measures were used. The proportions of patients with a successful outcome according to these different symptom measures were compared with the proportions of patients with healing of oesophagitis, or, in the case of maintenance therapy, the proportion of patients in whom oesophagitis did not relapse. The results are shown in the form of L’Abbé plots4 (fig 1), a technique primarily used to explore heterogeneity among trials. In this study, this technique has been used to display graphically the association between the symptom measure and absence of oesophagitis. Each box represents a single intervention within each trial. For example, a trial that evaluated the efficacy of placebo compared to a proton pump inhibitor would generate two boxes on the plot. The size of the box is related to the inverse of the standard error of the proportion healed and, therefore, represents the size of the study. The diagonal line represents the line of equivalence. If a box falls on this equivalence line it suggests that both outcomes were found in the same proportion in each intervention group. If a box is below the line it suggests that the outcome measure being tested against the reference outcome (always plotted on the vertical axis—for example, oesophagitis status) overestimates the treatment effect. Conversely, if boxes appear above the equivalence line, that particular outcome is underestimating the treatment effect compared to the reference outcome—for example, oesophagitis status.

RESULTS
Short term trials
Of 167 trials that were potentially relevant, 126 were eligible. The data were extracted from 1086–113 papers for this review. Of the trials not included, data required were not extractable from the original texts in eight. The other 10 trials were excluded because they were not published as an English text at the time of writing. The primary symptomatic outcome was explicitly stated in the methods in only 11/108 (10%) papers.9–21 25 26 49 64 65 81 105 107 The main outcomes that could be inferred from the papers are given in table 1. Only one paper21 used a validated questionnaire as part of the main symptomatic outcome assessment. The most common symptoms to be assessed either individually or as part of a global symptom score were heartburn (n = 93), regurgitation (n = 69), dysphagia (n = 55), and epigastric pain (n = 39). Other symptoms assessed included vomiting (n = 22), odynophagia (n = 20), nausea (n = 19), and belching (n = 15).

Symptom severity and/or frequency were assessed with a modified Likert scale in 71 of the trials. A 4 point modified Likert scale was most commonly used (n = 53) followed by a 5 point scale (n = 12), then a 3 point scale (n = 4), with 6 and 7 point scales used on only one occasion each. Ordinal scales were used in seven trials and visual analogue scales were used in 12 trials (all used a 100 mm length scale). Fifty three trials mentioned use of a patient diary card, but investigator interview was also often implied and it was unclear if the reported final outcome related to the patient or the investigator assessment.

Quality of life was formally assessed in six trials, two of which did not use a validated questionnaire. Two trials measured generic quality of life, one with the psychological general well being index (PGWBI) and the other with the short form 36 (SF-36) questionnaire. Patient satisfaction was reported in two trials.

Reduction versus absence of global symptoms as an outcome measure
Twenty five trials3 reported both absence of global symptoms and oesophagitis healing. Authors did not usually define which upper gastrointestinal symptoms were evaluated in their global assessment. Where these were defined, all included heartburn and at least one other upper gastrointestinal symptom (usually regurgitation). Similar proportions of patients had absence of symptoms and healing of oesophagitis according to the L’Abbé plot (fig 1) although there was a trend for absence of global symptoms to underestimate oesophagitis healing when the effect of treatment was small (fig 1). We also identified 13 trials35 37 41 51 55 63 81 94 96 101 102 105 that
Table 1  Main symptomatic outcomes* inferred from the papers included in this systematic review

<table>
<thead>
<tr>
<th>Main outcome measure</th>
<th>No of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heartburn and regurgitation reported separately</td>
<td>4</td>
</tr>
<tr>
<td>Heartburn, regurgitation-one other symptom reported separately</td>
<td>9</td>
</tr>
<tr>
<td>Heartburn, regurgitation-at least two other symptoms reported separately</td>
<td>18</td>
</tr>
<tr>
<td>Heartburn and night symptoms reported separately</td>
<td>16</td>
</tr>
<tr>
<td>Heartburn frequency and severity reported separately</td>
<td>1</td>
</tr>
<tr>
<td>Heartburn day, night, frequency and severity reported separately</td>
<td>5</td>
</tr>
<tr>
<td>Heartburn as sole symptom</td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td>5</td>
</tr>
<tr>
<td>Reduction</td>
<td>0</td>
</tr>
<tr>
<td>Absence and reduction</td>
<td>12</td>
</tr>
<tr>
<td>Change in score</td>
<td>0</td>
</tr>
<tr>
<td>Global symptoms (symptoms not defined)</td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td>2</td>
</tr>
<tr>
<td>Reduction</td>
<td>0</td>
</tr>
<tr>
<td>Absence and reduction</td>
<td>6</td>
</tr>
<tr>
<td>Change in score</td>
<td>3</td>
</tr>
<tr>
<td>Global symptoms (heartburn, regurgitation-one other)</td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td>3</td>
</tr>
<tr>
<td>Reduction</td>
<td>0</td>
</tr>
<tr>
<td>Absence and reduction</td>
<td>1</td>
</tr>
<tr>
<td>Change in score</td>
<td>0</td>
</tr>
<tr>
<td>Global symptoms (heartburn regurgitation-at least two other symptoms)</td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td>3</td>
</tr>
<tr>
<td>Reduction</td>
<td>2</td>
</tr>
<tr>
<td>Absence and reduction</td>
<td>3</td>
</tr>
<tr>
<td>Change in score</td>
<td>3</td>
</tr>
<tr>
<td>Most marked symptom</td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td>1</td>
</tr>
<tr>
<td>Reduction</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
</tr>
<tr>
<td>Unclear</td>
<td>8</td>
</tr>
</tbody>
</table>

*Subjective assessment by one reviewer (PM) of the main symptom outcome measures used in short term randomised controlled trials of medical therapy for oesophagitis.

reported both global symptom reduction and oesophagitis cure. Global symptom reduction overestimated response to treatment compared with oesophagitis healing, as most trials were situated below and to the right of the “equivalence” line of the l’Abbé plot (fig 2). Global symptom absence is, therefore, a better predictor of oesophagitis healing than symptom reduction according to this analysis.

Individual reflux symptoms as outcome measures

We identified 31 trials31 39 40 46 47 50 61 64 65 68–72 90 93 96 97 99 that reported both absence of heartburn and healing of oesophagitis. The proportions of patients with heartburn and oesophagitis healing were very similar across a wide range of treatment effects (fig 3). Absence of regurgitation also showed good correlation with oesophagitis healing in the 20 trials31 39 40 46 47 50 61 64 65 68–72 90 93 96 97 99 that reported the two outcomes (fig 4). There was, however, a tendency for absence of regurgitation to overestimate oesophagitis healing (fig 4), a finding similar to the 20 trials26 31 39 40 46 47 50 61 64 65 68–72 90 93 96 97 99 in which absence of heartburn could be compared with absence of regurgitation, as the latter tended to slightly overestimate response to treatment (fig 5). There were, however, a few trials in which measures of heartburn overestimated response to treatment compared to regurgitation, so this latter symptom may be useful in some cases (fig 5). Thirteen trials31 39 46 47 60 62 69 70 72 90 97 99 102 evaluated absence of dysphagia and heartburn as outcomes. When absence of dysphagia was used, this overestimated treatment response compared with absence of heartburn (fig 6).

Comparison of frequency and severity of symptoms as outcome measures

There is a paucity of information on whether either symptom frequency or severity is the best measure of treatment in GORD. We identified two trials35 38 that reported symptom frequency, symptom severity, and oesophagitis healing. Symptom frequency alone underestimated response to treatment compared with oesophagitis healing (fig 7). Symptom severity gave much closer correlation with oesophagitis healing (fig 8).

Long term trials

Of the 157 papers that were potentially relevant, 48 were eligible and data could be extracted from 37 of these.26 29 31 39 40 46 47 50 61 64 65 68–72 90 93 96 97 99 The major cause of ineligibility was that studies enrolled only patients with endoscopy-negative reflux disease. Quality of life outcomes were assessed in three trials, two of which used the PGWBL, and the third used the SF-36 questionnaire. Disease specific quality of life was not measured. Patient satisfaction was recorded in only one trial.

The eligible long term trials were also used to assess the optimum definition of symptom relapse. It is unclear whether symptoms should be assessed over a defined time period. There are also few data on whether relapse should be defined as the presence of any symptom or as the occurrence of moderate or even severe symptoms. We identified seven trials101–107 that defined relapse as the presence of any symptom (either heartburn alone or global symptoms); this measure correlated well with the proportion of patients having relapse of oesophagitis (fig 9). All of these trials also defined the period of symptom assessment as between 1–7 days. There were six trials115 122–125 that defined relapse as occurrence of moderate to severe symptoms during an...
undefined time period. These trials underestimated relapse rates compared with oesophagitis relapse (fig 10). Four trials\textsuperscript{126–129} defined relapse as moderate or severe symptoms over 1–7 days; a similar proportion of patients were in remission using this definition compared with absence of oesophagitis (fig 11).

**DISCUSSION**

This is the first review that surveys and assesses the symptom outcome measures and the methodologies used for obtaining symptom data used in randomised controlled trials. This comprehensive, systematically performed overview of the literature has given new insights into the optimal evaluation of symptoms. The methods used in trials to date vary widely. In some instances—most importantly, the number of response options offered to patients—current practices do not reflect what is considered to be best practice. The data suggest that in short term trials, absence of reflux symptoms is a better predictor of healing of oesophagitis than symptom reduction. In long term trials, absence of symptoms also correlates well with remission of oesophagitis, although the time period over which symptoms are evaluated also appears to be important. On the basis of a small number of trials, there also appears to be an additional advantage to measuring severity as well as frequency of reflux symptoms. These data suggest the optimal evaluation would assess symptoms over a set time period and would define treatment success as no more than mild symptoms (for example, reflux symptoms assessed over 1 week with treatment success defined as no more than mild symptoms for 1 day/week).

Clinical trials of GORD drug therapies usually select patients who have predominant heartburn and/or oesophagitis at endoscopy. Patients recruited into trials may, therefore, have a symptom profile that is not exactly comparable with unselected GORD patients seen in primary care. This must be borne in mind when data from such cohorts are evaluated for the full range of reflux symptoms. With this caveat, heartburn appears to be the most important symptom and, happily, this gives the closest correlation with oesophagitis healing. Absence of heartburn predicts the absence of other reflux symptoms, such as regurgitation and dysphagia, although in a few trials these symptoms were present in patients in whom heartburn was no longer present as a result of therapy. Separate evaluation of regurgitation, and possibly dysphagia, may, therefore, provide additional information. This possibility is supported by our data that suggest that global reflux symptoms also correlate with oesophagitis healing and may be a more conservative measure of treatment success suitable for therapies with low efficacy.

Most studies used a diary card to assess symptoms, which is likely to be the optimal method for gathering information on subjective symptoms (see McColl on page iv49–iv54),\textsuperscript{130} but, frequently, no information was given as to which of these data were being drawn on. When both a diary card and an investigator administered symptom questionnaire are used, authors should clarify which assessment is being reported. Modified Likert scales were nearly always employed to gather information on severity of symptoms; this would also seem the most appropriate method. Importantly, though, the structures of the modified Likert scales used were usually suboptimal, as a 4 point scale was used in the great majority of studies, whereas data suggests a 6 to 7 point scale is more appropriate (see Wyrwich & Staebler Tardino on page iv55–iv56).
Figure 8 L’Abbé plot of trials reporting heartburn severity and oesophagitis healing.

Figure 9 L’Abbé plot of maintenance trials reporting absence of symptoms and prevention of oesophagitis relapse.

Figure 10 L’Abbé plot of maintenance trials defining moderate/severe symptoms as relapse (with no time period of assessment stated), and prevention of oesophagitis relapse.

Figure 11 L’Abbé plot of maintenance trials defining moderate/severe symptoms as relapse (over a defined time period) and prevention of oesophagitis relapse.

Some caveats apply to this review of reflux oesophagitis clinical trials. Although we evaluated randomised controlled trials, the randomisation process was not used in the analyses, each treatment arm being considered separately. The information presented is, therefore, observational rather than randomised, controlled data. Outcomes were evaluated in groups of patients rather than individuals, so this ecological analysis could be subject to the ecological fallacy. For example, if oesophagitis healing is seen in 50% of patients within a group and absence of heartburn is also seen in 50%, there appears to be perfect agreement between the two outcomes. However, it is possible that the 50% of cases with oesophagitis healing were all the patients that continued to have heartburn. This is less of a problem when a successful outcome is found in a high proportion of the group. It can then be assumed we are looking at mostly the same individuals. Furthermore, individual patient data suggest that symptom absence is a good predictor of oesophagitis healing, although there is a 10–20% discrepancy between the two outcomes.

Healing of oesophagitis was used as the reference standard for success of therapy against which symptom outcomes were compared. This may not be the best outcome to assess, as some patients can remain symptomatic in the absence of erosive oesophagitis. However, oesophagitis is a very specific marker for GORD, and the most objective outcome measure that is available. This systematic review could not have identified all relevant papers as the searches on which it was based extended only into the year 2000. It is unlikely that inclusion of papers published since then would have any major impact on our conclusions. Despite its limitations, this systematic review provides a comprehensive summary of the symptom outcome measures that have been used in randomised controlled trials of medical therapy. Our findings suggest that there is considerable potential for reduction and point to how this could be achieved.


A systematic review of symptomatic outcomes


