

ORIGINAL CONTRIBUTIONS

Assessing Symptoms in Gastroesophageal Reflux Disease: How Well Do Clinicians' Assessments Agree with Those of Their Patients?

Elaine McColl, M.Sc., Ola Junghard, Ph.D., Ingela Wiklund, Ph.D., and Dennis A. Revicki, Ph.D.
Centre for Health Services Research, School of Population and Health Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom; Outcomes Research, AstraZeneca, Mölndal, Sweden; and Center for Health Outcomes Research, MEDTAP International, Bethesda, Maryland

- OBJECTIVES:** This study aimed to determine the extent of agreement between clinicians and patients regarding assessments of reflux symptom severity in patients with gastroesophageal reflux disease.
- METHODS:** Data were analyzed from four randomized clinical trials involving 2,674 patients treated with esomeprazole, omeprazole, ranitidine, or placebo. The extent of agreement was determined for symptom severity before and after 4–8 wk of treatment, and for the absence of symptoms after treatment. Agreement was further analyzed by determining weighted kappa values, which were interpreted according to the criteria of Landis and Koch.
- RESULTS:** Before treatment, clinician-patient agreement regarding symptom severity in the four studies was slight to moderate (kappa: 0.17–0.53); 48–52% of assessments agreed for heartburn, 24–35% for epigastric pain, 36–43% for regurgitation, and 63% agreed for dysphagia. Poor agreement reflected clinician underestimation of symptom severity relative to patient ratings in three studies and clinician overestimation in one study. Agreement regarding symptom severity improved following treatment, and was fair to substantial (kappa: 0.31–0.73); 58–78% of assessments agreed for heartburn, 42–60% for epigastric pain, 66–76% for regurgitation, and 86% agreed for dysphagia. After treatment, agreement was greatest for patients reporting absence of symptoms and decreased with increasing severity of symptoms.
- CONCLUSIONS:** The agreement between clinicians and patients in their assessments of the severity of reflux symptoms is poor, particularly before treatment and for more severe symptoms. Improvements in clinician-patient communication may help to bridge this gap, and greater reliance on patient assessments may be appropriate.

(Am J Gastroenterol 2005;100:11–18)

INTRODUCTION

Gastroesophageal reflux disease (GERD) affects more than one in ten of the population and is associated with a broad range of gastrointestinal symptoms, including heartburn, regurgitation, dysphagia, and epigastric pain (1). In up to half of all cases, GERD is associated with endoscopic esophagitis (1). However, individuals with GERD suffer significant pain and discomfort whether or not esophagitis is present, and the resulting impairment of quality of life is not dependent on endoscopic findings (2, 3). Thus, international guidelines recommend that GERD should be diagnosed and managed on the basis of symptom frequency and severity (4–6).

Although the symptoms of GERD are experienced by the patient, assessments of symptom severity have, traditionally, been carried out by the clinician. This process involves the clinician asking appropriate questions to elicit relevant information from the patient, interpreting the response, and

then making a judgment as to the severity of the symptoms. However, an increasing body of evidence suggests that there is often poor agreement between patients and clinicians in their assessment of symptom severity. Such discrepancies may have a bearing on the management of GERD, where assessments of symptom severity are a major determinant of therapeutic decisions. They also have important implications for metaanalyses, which pool data from studies including different groups of informants.

The reported agreement between patient and clinician ratings of subjective outcomes across a range of diseases has been moderate, at best (7–14). For example, in a study of symptoms such as weight loss, fatigue, and anxiety in a sample of HIV patients, clinician-patient agreement regarding reports of symptoms was no more than slight to moderate (8). Similar results were seen for gastrointestinal symptoms in this patient group (7). To our knowledge, only one study to date presents data from both clinician and patient assessments

of GERD symptoms (15). In this double-blind study of symptom resolution during proton pump inhibitor (PPI) therapy, considerably more clinician assessments (~75%) than patient assessments (~50%) rated symptoms as absent after 4 wk of omeprazole treatment.

Given this paucity of data regarding clinician-patient agreement in the assessment of GERD symptoms, we performed a secondary analysis of clinician and patient symptom-rating data from four clinical trials of pharmacotherapy in the treatment of GERD. The aim of this study was to examine the level of agreement between clinician and patient ratings of symptom severity and freedom from symptoms, and to determine the relationship between symptom severity and the level of clinician-patient agreement.

METHODS

Studies

This paper reports a secondary analysis of reflux-related symptom data from four multicenter, randomized trials (16–21). The main features of the four studies are summarized in Table 1. All four studies involved similar patient populations with the exception that studies 1–3 enrolled patients without endoscopic evidence of esophagitis and study 4 involved patients who were poorly responsive to ranitidine. (In study 4, all patients had received ranitidine for 6 wk prior to the randomized phase of the study.) All patients had experienced episodes of heartburn for at least 6 months (studies

1 and 2) or at least 12 months (study 3), or had a history of GERD with heartburn as the predominant symptom for at least 6 months (study 4). During the week preceding study entry, patients had to have experienced episodes of heartburn on at least 4 days (studies 1 and 2), on at least 2 days (study 3), or at least moderate episodes on at least 4 days despite treatment with ranitidine (study 4). Exclusion criteria included a current or recent history of other gastrointestinal diseases and conditions (*e.g.*, Zollinger-Ellison syndrome, esophageal stricture, duodenal or gastric ulcers, Barrett's metaplasia) (all studies), and treatment with H₂-receptor antagonists or proton pump inhibitors within the month prior to the study (studies 3 and 4).

Symptom Assessments

In each study, patients and clinicians independently rated the severity of a range of symptoms of reflux disease as summarized in Table 1. Investigator ratings were on a 4-point scale: none, mild, moderate and severe. Mild symptoms were described as “awareness of sign or symptom, but easily tolerated,” moderate symptoms as “discomfort sufficient to cause interference with normal activities,” and severe symptoms as “incapacitating, with inability to perform normal activities.” Any patient whom the investigator rated as having no heartburn at baseline was excluded from studies 1, 2, and 3. Likewise, subjects rated as not having heartburn at all at baseline, based on patient diary data for the 2 wk before baseline, were excluded from study 4.

Table 1. Summary of Included Studies

| | Study 1/2 (n = 1,848*) | Study 3 (n = 509 [†]) | Study 4 (n = 317) |
|---------------------------------|---|---|---|
| Interventions | Esomeprazole vs omeprazole | Omeprazole vs placebo | Omeprazole vs ranitidine |
| Duration of treatment | 4 wk | 4 wk | 8 wk |
| Setting | Canada, UK, and Ireland (study 1); Denmark, Sweden, Norway, and Finland (study 2) | Denmark, Sweden | United States |
| Blinding of outcome assessment | Yes—double-blind study | | |
| Symptom assessment by clinician | Heartburn; regurgitation; epigastric pain Rated as: none; mild; moderate; severe | Heartburn; regurgitation; epigastric pain Rated as: none; mild; moderate; severe | Regurgitation; epigastric pain; dysphagia Rated as: none; mild; moderate; severe |
| Symptom assessment by patient | Gastrointestinal Symptoms Rating Scale (22–24)—15 symptoms (modified response categories in study 3) [‡] | | |
| Timing of assessment | Baseline (pre-randomization); 2-wk follow-up; end of study (4 wk) | Baseline (pre-randomization); end of study (4 wk) | Baseline (pre-randomization); 4-wk follow-up [§] ; end of study (8 wk) |

*Patients who had at least one GSRS assessment.

[†]Total number of patients (intention to treat analysis).

[§]Levels of agreement at these interim assessments are not examined here.

[‡]Same assessment method applies across all studies.

In all four studies, patients reported the severity of their symptoms using the Gastrointestinal Symptoms Rating Scale (GSRS), a self-completion measure with good evidence of validity and reliability (22–24). The GSRS includes 15 items addressing different gastrointestinal symptoms (including those assessed by the clinicians), with a recall period of the preceding week. It uses a 7-point Likert-type response scale with verbal descriptors (none at all, minor, mild, moderate, moderately severe, severe, and very severe), and is designed to measure the amount of discomfort a patient has experienced. In study 3, the 7-point scale was slightly different from that used in the other studies, ranging from “not bothersome at all” to “worst imaginable.” For the purposes of comparison with the 4-point investigator assessment of symptom severity, patient ratings for each symptom were collapsed into four categories: “none,” “minor/mild,” “moderate,” and “moderately severe/severe/very severe” (in studies 1, 2, and 4), and “not bothersome at all,” “minimal,” “moderate,” and “quite severe/severe/very severe/worst imaginable” (in study 3).

Data Analysis

Data from studies 1 and 2 were combined since these two studies employed identical inclusion and exclusion criteria, treatment durations, and symptom assessments. Data were analyzed separately for studies 3 and 4. In each analysis, the blinded nature of the studies allowed symptom absence and severity-assessment data from all treatment groups to be pooled.

For each study, the proportion of patients for whom patient and clinician ratings of symptom severity were the same (percentage of clinician-patient agreement) was determined for each symptom, both before and after treatment (*i.e.*, at baseline and at the end of the study). The proportion of clinician ratings, which were (i) higher than patient ratings and (ii) lower than patient ratings were also determined for both time points. At the end of each study, the percentage of clinician-patient agreement for symptom absence (*i.e.*, a rating of “none”) was also determined.

Observed levels of clinician-patient agreement were further investigated by calculating weighted kappa values (25–27). We employed a quadratic weighting system, which gives greater weight to larger discrepancies between raters. Kappa values provide a chance-corrected index of agreement and are calculated as: observed agreement = 1 – (observed disagreement/disagreement expected by chance).

Weighted kappa values were interpreted according to the criteria of Landis and Koch, as follows: 0.00 = poor; 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.80–1.00 = almost perfect (28).

A kappa value of 0 indicates that the observed disagreement is purely due to chance, while a kappa value of 1 denotes perfect agreement between the two raters. However, in practice, the maximum attainable kappa value for a given data set may be less than 1. To facilitate comparisons of levels of agreement across data sets, it is therefore useful to express the observed kappa value as a percentage of the maximum possible for the data set in question and we followed this convention.

In a further analysis, the clinician-patient agreement according to symptom severity was determined (*i.e.*, severity agreement). This was calculated as the number of observations where the clinician and patient both agreed that the symptoms were at a specific severity level divided by the number of observations where one or other, or both, considered the symptoms to be at that level (29).

RESULTS

Patient Characteristics

Patient demographic and baseline clinical characteristics for the three study populations are reported in Table 2. Just under half of patients in each dataset were male, and the mean age ranged from 45.3 yr (study 4) to 50.4 yr (study 3). The severity of reflux symptoms (mean GSRS score in the reflux domain) was also similar across the three populations, ranging from 3.00 to 3.82.

Symptom Assessment before Study Treatment

The agreement between clinician and patient ratings of symptom severity (*i.e.*, percentage of clinician-patient agreement) before treatment was lowest for epigastric pain (agreement in 24–35% of patients), intermediate for regurgitation (36–43%), and heartburn (48–52%), and highest for dysphagia (63%) (Fig. 1). Clinician-patient agreement was stronger in study 1/2 and weaker in study 4. According to the Landis and Koch criteria for kappa values, agreement across all four studies was slight–fair for epigastric pain (0.17–0.34), fair for heartburn (0.36–0.39), fair–moderate for regurgitation (0.38–0.46), and moderate for dysphagia (0.53) (Table 3).

Table 2. Demographic and Baseline Clinical Characteristics

| | Study 1/2 (n = 1,848*) | Study 3 (n = 509 [†]) | Study 4 (n = 317) |
|--|------------------------|---------------------------------|-------------------|
| Gender (% male) | 46.0 | 40.5 | 41.7 |
| Ethnic origin (% Caucasian) | 98.4 | 99.2 | 89.0 |
| Age (years, mean ± SD) | 48.2 ± 13.9 | 50.4 ± 13.7 | 45.3 ± 12.8 |
| Baseline GSRS score, reflux domain (mean ± SD) | 3.82 ± 1.17 | 3.00 ± 0.99 | 3.27 ± 0.8 |

*Patients who had at least one GSRS assessment.

[†]Total number of patients (intention to treat analysis).

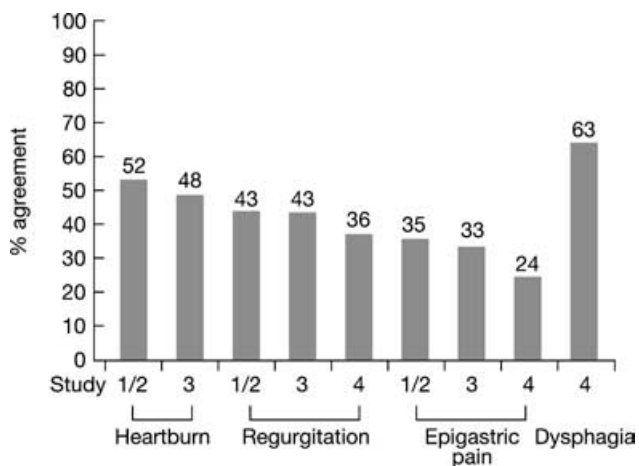


Figure 1. Clinician-patient agreement regarding symptom severity before treatment.

Disagreement between clinician and patient ratings of symptoms in studies 1/2 and 3 reflected clinicians typically underrating symptom severity compared with patient ratings (34–63% of patients), rather than overrating them (4–14%) (Fig. 2). In these studies, clinician underrating was greater for epigastric pain (54–63%) than for heartburn (34–44%) and regurgitation (46–49%). Underestimation was more likely in study 3 than in study 1/2. In study 4, disagreement reflected a tendency of clinicians to overrate symptoms compared with patients (28–64% overrating), rather than underrate them (8–12%) (Fig. 2). In this study, clinician overrating was greater for epigastric pain (64%) and regurgitation (56%) than for dysphagia (28%).

Symptom Assessment after Study Treatment

Clinician-patient agreement regarding symptom severity across all symptoms was greater after treatment than be-

fore treatment, ranging from 42–58% for epigastric pain, through heartburn (63–78%) and regurgitation (66–76%), to 86% for dysphagia (Fig. 3A). Agreement was best in study 1/2. Based on the corresponding weighted kappa values across all studies, agreement was substantial for heartburn (0.71–0.73), fair–moderate for epigastric pain (0.31–0.42), moderate–substantial for regurgitation (0.52–0.67), and moderate for dysphagia (0.58) (Table 3). The extent of clinician underestimation of symptom severity relative to patient ratings was lower than at baseline in study 1/2 (34–54% at baseline; 13–37% after treatment), and in study 3 (43–63% at baseline; 25–56% after treatment). In study 4, clinicians continued to overrate rather than underrate symptom severity relative to patients, but the level of overrating was reduced (from 28–64% at baseline to 7–35% after treatment).

As might be expected, clinician-patient agreement regarding the presence *versus* absence of symptoms following treatment was at least as good as the level of agreement regarding symptom severity. The percentage of agreement ranged from 63–67% for epigastric pain, and 78–82% for regurgitation to 86–88% for heartburn, and 90% for dysphagia (Fig. 3B). Interpreting the kappa values according to the Landis and Koch criteria, agreement was fair for epigastric pain (0.33–0.37), moderate to substantial for regurgitation (0.51–0.64), and substantial for heartburn (0.72–0.75) and dysphagia (0.61) (Table 3).

Clinician-Patient Agreement According to Symptom Severity (Severity Agreement)

In a further analysis, clinician-patient concordance according to symptom severity (*i.e.*, severity agreement) was compared for the four symptom levels (none; minimal/mild; moderate; and moderately severe/severe/very severe). Assessments at the end of the studies showed better agreement for absence of symptoms (symptom level 1, *i.e.*, “none”) compared with

Table 3. Clinician-Patient Agreement Regarding Symptom Severity before and after Treatment, and Agreement Regarding Absence of Symptoms after Treatment

| | Weighted Kappa Values (Weighted Kappa as a Percentage of Maximum Possible Values) | | |
|-------------------------------------|---|--------------|--------------|
| | Study 1/2 | Study 3 | Study 4 |
| Before treatment | | | |
| Heartburn | 0.39 (50.4%) | 0.36 (44.3%) | |
| Epigastric pain | 0.27 (30.7%) | 0.34 (40.8%) | 0.17 (22.8%) |
| Regurgitation | 0.44 (47.6%) | 0.46 (51.6%) | 0.38 (49.4%) |
| Dysphagia | | | 0.53 (60.1%) |
| After treatment | | | |
| Heartburn | 0.73 (74.5%) | 0.71 (78.9%) | |
| Epigastric pain | 0.42 (52.3%) | 0.34 (42.5%) | 0.31 (35.9%) |
| Regurgitation | 0.57 (66.5%) | 0.67 (72.8%) | 0.52 (59.3%) |
| Dysphagia | | | 0.58 (66.6%) |
| Absence of symptoms after treatment | | | |
| Heartburn | 0.72 (72.7%) | 0.75 (82.7%) | |
| Epigastric pain | 0.37 (79.5%) | 0.33 (90.4%) | 0.33 (34.1%) |
| Regurgitation | 0.51 (72.9%) | 0.64 (78.1%) | 0.55 (55.8%) |
| Dysphagia | | | 0.61 (69.2%) |

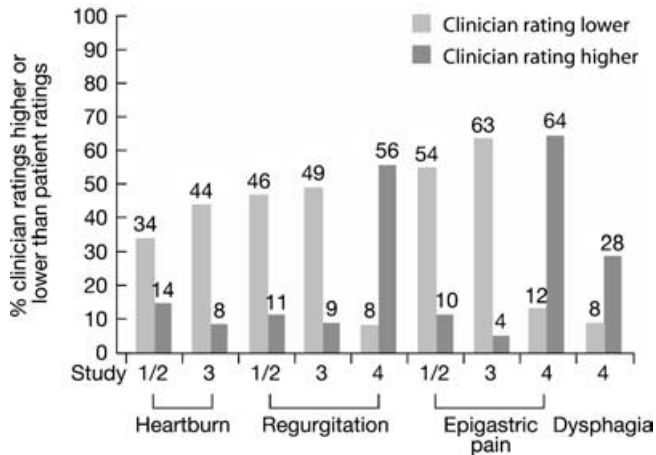


Figure 2. Percentage of clinician ratings that were lower or higher than patient ratings of symptoms before treatment.

higher levels of severity (symptom levels 2–4, *i.e.*, “mild” to “severe”). Observed levels of agreement at the end of the study were 74–79% for absence of heartburn (compared with <54% for heartburn graded “mild” to “severe”), 41–57% for absence of epigastric pain (compared with <30% for “mild” to “severe” epigastric pain), 67–77% for absence of regurgitation (compared with <42% for “mild” to “severe” regurgitation), and 89% for absence of dysphagia (compared with <31% for “mild” to “severe” dysphagia) (Table 4). Across all four symptoms, there was no overlap in the 95% confidence intervals for agreement for symptom level 1 (“none”) compared with higher levels of symptoms. There was also a trend towards better agreement for mild symptoms compared with more severe symptoms (see Table 4) but the 95% confidence intervals overlapped in many cases. In contrast, before treatment, there was no consistent pattern of severity agreement for all three study populations and all four symptoms.

DISCUSSION

Our analysis of data from a total of 2,674 subjects enrolled in four clinical trials showed substantial levels of disagreement between clinicians and patients regarding symptom severity before treatment, with patients and their clinicians agreeing in fewer than half of severity assessments; agreement was poorest for epigastric pain and best (but still only moderate, according to the weighted kappa value) for dysphagia. The particularly low agreement for epigastric pain may reflect differences between patients and clinicians in their understanding of terms such as “pain” and “discomfort” (30–33), and difficulties experienced by patients in locating their stomachs (33). The low agreement regarding heartburn may also reflect different understandings of this term (34–36).

In three of the four studies, the discrepancy in assessment largely resulted from clinicians underestimating symptom severity relative to their patients, reflecting findings from studies of clinician-patient agreement in other diseases

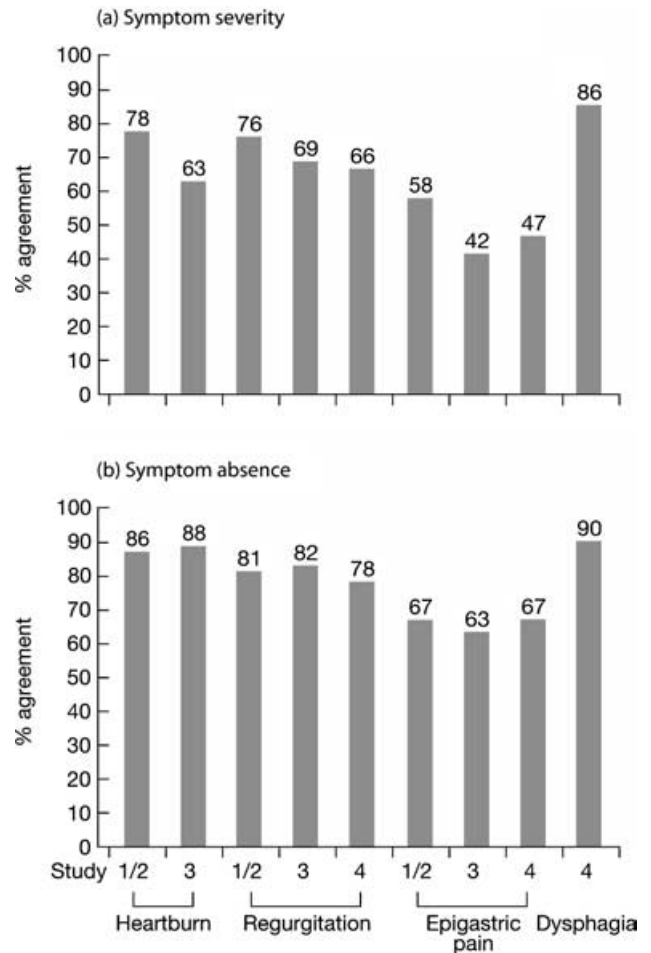


Figure 3. Clinician-patient agreement after treatment regarding (A) symptom severity and (B) symptom absence.

(7, 9–11, 13, 14, 37). Clinicians’ relative overestimation of symptom severity in the fourth study may reflect a difference in study population (this study focused on patients who were known to have failed to respond to prior rantidine therapy, which may have resulted in clinicians rating their symptoms as more severe) or may be due to cultural differences between Europe and North America.

Patient-clinician agreement increased markedly following treatment. At the end of the studies, patient and clinician assessments of reflux symptom severity agreed in over half of patients. This convergence of assessments may, in part, reflect a learning effect on the part of clinicians and/or patients, with agreement improving as a result of improved communication and understanding over time. However, as most patients experienced a marked improvement in symptoms following treatment, it may also reflect the influence of reduced symptom severity on clinician-patient agreement; clinicians and patients tended to have closer agreement for mild disease symptoms than for moderate or severe symptoms. An analysis of agreement relative to degree of reflux symptom severity after treatment showed that patients and physicians were more likely to agree when patients rated their symptoms mild or

Table 4. Clinician-Patient Agreement (%) According to Symptom Severity after Treatment (Severity Agreement)

| Symptom | Symptom Severity | | | |
|-----------------|------------------|------------------|----------|--|
| | None | Minimal/ Mild | Moderate | Moderately Severe/ Severe/Very Severe |
| Heartburn | | | | |
| Study 1/2 | 79 | 53 | 29 | 26 |
| Study 3 | 74 | 41 | 23 | 26 |
| Epigastric pain | | | | |
| Study 1/2 | 57 | 26 | 15 | 16 |
| Study 3 | 41 | 12 | 15 | 14 |
| Study 4 | 46 | 29 | 0 | 8 |
| Regurgitation | | | | |
| Study 1/2 | 77 | 33 | 15 | 15 |
| Study 3 | 72 | 37 | 28 | 17 |
| Study 4 | 67 | 41 | 6 | 0 |
| Dysphagia | | | | |
| Study 4 | 89 | 30 | 0 | 0 |

absent. This is consistent with previous findings in patients with GERD (15) and other conditions (8–10, 13, 38–43). Indeed, the best levels of clinician-patient agreement were achieved for absence of reflux symptoms following treatment, suggesting that patients and physicians are better able to reach a common understanding of the presence or absence of a symptom than the grading of symptom severity, as also observed by Heading and colleagues (44).

It should be noted that in three out of the four studies, patients and clinicians completed an interim assessment of symptom severity halfway through the trial. It is possible that these interim assessments may have influenced the severity ratings after treatment. For example, the assessor may have reasoned along the following lines: “Last time, I rated the symptom as moderate so, if the symptom has diminished, I should rate it as mild on this occasion.” It is notable that the mismatch between clinician and patient assessments after treatment was greatest in the only study in which there was no such interim assessment.

A number of potential limitations should be considered when interpreting the findings reported here. Firstly, physicians and patients completed different instruments when assessing symptom severity and symptom absence, and there were some minor differences in assessment method between the studies; these differences may have introduced variability between the comparisons. Secondly, the 7-point patient ratings were arbitrarily collapsed to a 4-point rating scale for comparison with physician ratings, an approach previously used in other assessments of symptom resolution (45). Thirdly, our data were gathered in the context of clinical trials and predominantly used validated symptom rating scales; therefore, our findings may not be fully representative of the situation in routine clinical practice. Fourthly, there was no formal assessment of interinvestigator agreement in symptom rating, although all investigators were trained in the application of the rating scale.

A number of factors may contribute to the discrepancy between clinician and patient ratings. Accurate assessment of symptom presence and severity by the physician depends on the patient communicating all the pertinent information and the physician comprehending it. But patients may underreport their symptoms if they wish to please the clinician or researcher (11) or exaggerate severity if they perceive that a possible benefit may result, such as receiving further treatment or being certified as unable to work.

Moreover, patients and clinicians may use different frames of reference and criteria when assessing symptoms (11). Patients are likely to use an internal frame of reference—either in relation to their own past experiences of symptoms, or to an ideal state of health—while physicians may implicitly compare a given patient to other patients possibly with more severe disorders (46). Holmes and colleagues (47) have suggested that clinicians may focus primarily on frequency and intensity of symptoms in their history-taking, while patients also consider symptom-related disability and impact on health-related quality of life when judging how severe those symptoms are.

The mismatch between clinicians and patients in the assessment of symptoms has far-reaching implications. In metaanalyses, extreme caution should be exercised in pooling data from clinicians and patients regarding symptom-severity grading. Agreement is better in respect of the binary outcome of symptom resolution, but even those assessments should not be combined unthinkingly.

In clinical studies, underestimation of reflux symptom severity prior to study entry may lead to only those patients with the most severe symptoms being enrolled; hence, findings may not be applicable to the entire spectrum of GERD. Furthermore, underestimation of symptoms prior to therapy, coupled with better agreement following treatment, will lead to understatement of the benefits of treatment and make the differentiation between treatments more difficult. This has led to the recommendation that patient ratings are more appropriate than clinician ratings in this setting (48).

In day-to-day clinical practice, both the decision to offer the patient treatment and the type of treatment offered are often determined by the clinician’s initial assessment of symptoms. Poor clinician-patient agreement prior to therapy may be an obstacle to the appropriate management of GERD in the clinic, and this may contribute to the widespread treatment dissatisfaction experienced by patients with GERD (49). Conversely, the benefits of good doctor-patient communication are likely to extend beyond more accurate symptom assessment, to improved patient outcomes, satisfaction, well-being, and trust (50–54).

Recognition of the poor agreement between patients and physicians regarding reflux symptom severity should serve as a starting point to stimulate research into the potential communication gap between these two groups. More appropriate instruments for symptom assessment may improve agreement between patient and clinician reporting, and the development of a formal patient-completed rating scale for use in clinical

practice may be helpful. Ultimately, recognition that, when supported by appropriate instruments, patients may be the most faithful reporters of their own symptoms could allow a move towards giving greater weighting to patients' own reports. Appropriate support will be needed, so that patients are able to recognize, distinguish, and report their own symptoms in ways that inform clinician decision-making.

ACKNOWLEDGMENTS

We would like to thank Dr. Chris Winchester and Dr. Rowena Hughes for editorial assistance with this paper; Professor Graham Dunn and Dr. Richard Lowry for their invaluable advice on the calculation and interpretation of measures of agreement; Professors Senga Bond and Greg Rubin, and two anonymous referees, for their helpful comments on earlier drafts of this paper. Elaine McColl and Dennis Revicki received travel grants from AstraZeneca to attend a workshop entitled "Symptom evaluation in reflux disease" from which this work originated. Elaine McColl is a Primary Care Career Scientist funded by the United Kingdom's Department of Health Research Capacity Development Programme.

Reprint requests and correspondence: Elaine McColl, M.Sc., Centre for Health Services Research, University of Newcastle upon Tyne, 21 Claremont Place, Newcastle upon Tyne, NE2 4AA, United Kingdom.

Received August 2, 2004; accepted August 12, 2004.

REFERENCES

- Sonnenberg A, El-Serag HB. Clinical epidemiology and natural history of gastroesophageal reflux disease. *Yale J Biol Med* 1999;72:81–92.
- Wiklund I, Talley NJ. Update on health-related quality of life in patients with gastroesophageal reflux disease. *Expert Rev Pharmacoeconomics Outcomes Res* 2003;3:341–50.
- Kulig M, Leodolter A, Vieth M, et al. Quality of life in relation to symptoms in patients with gastro-oesophageal reflux disease—an analysis based on the ProGERD initiative. *Aliment Pharmacol Ther* 2003;18:767–76.
- Szarka LA, DeVault KR, Murray JA. Diagnosing gastroesophageal reflux disease. *Mayo Clin Proc* 2001;76:97–101.
- Dent J. Definitions of reflux disease and its separation from dyspepsia. *Gut* 2002;50 (Suppl 4):iv17–20 (Discussion iv1–2).
- Dent J, Armstrong D, Delaney DC, et al. Symptom evaluation in reflux disease: Proceedings of a workshop held in Marrakech, Morocco. *Gut* 2004;53:iv1–65.
- Corley DA, Cello JP, Koch J. Accuracy of endoscopic databases for assessing patient symptoms: Comparison with self-reported questionnaires in patients infected with the human immunodeficiency virus. *Gastrointest Endosc* 2000;51:129–33.
- Fontaine A, Larue F, Lassauniere JM. Physicians' recognition of the symptoms experienced by HIV patients: How reliable? *J Pain Symptom Manage* 1999;18:263–70.
- Justice AC, Rabeneck L, Hays RD, et al. Sensitivity, specificity, reliability, and clinical validity of provider-reported symptoms: A comparison with self-reported symptoms. Outcomes Committee of the AIDS Clinical Trials Group. *J Acquir Immune Defic Syndr* 1999;21:126–33.
- Justice AC, Chang CH, Rabeneck L, et al. Clinical importance of provider-reported HIV symptoms compared with patient-report. *Med Care* 2001;39:397–408.
- Stephens RJ, Hopwood P, Girling DJ, et al. Randomized trials with quality of life endpoints: Are doctors' ratings of patients' physical symptoms interchangeable with patients' self-ratings? *Qual Life Res* 1997;6:225–36.
- Suarez-Almazor ME, Conner-Spady B, Kendall CJ, et al. Lack of congruence in the ratings of patients' health status by patients and their physicians. *Med Decis Making* 2001;21:113–21.
- Wolfenden LL, Diette GB, Krishnan JA, et al. Lower physician estimate of underlying asthma severity leads to undertreatment. *Arch Intern Med* 2003;163:231–6.
- Nekolaichuk CL, Bruera E, Spachynski K, et al. A comparison of patient and proxy symptom assessments in advanced cancer patients. *Palliat Med* 1999;13:311–23.
- Sandmark S, Carlsson R, Fausa O, et al. Omeprazole or ranitidine in the treatment of reflux esophagitis. Results of a double-blind, randomized, Scandinavian multicenter study. *Scand J Gastroenterol* 1988;23:625–32.
- 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.
- Talley NJ, Junghard O, Wiklund I. Improvement in health-related quality of life following esomeprazole treatment in patients with endoscopy-negative gastro-oesophageal reflux disease (GORD)(abstract). *Gut* 2001;49:2365.
- Talley NJ, Armstrong D, Junghard O, et al. Predictors of treatment response in patients with non-erosive reflux disease (NERD). *Gut* 2002;51:A230.
- 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.
- 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.
- Revicki DA, Sorensen S, Maton PN, et al. Health-related quality of life outcomes of omeprazole versus ranitidine in poorly responsive symptomatic gastroesophageal reflux disease. *Dig Dis* 1998;16:284–91.
- Svedlund J, Sjodin I, Dotevall G. GSRS—a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci* 1988;33:129–34.
- Dimenas E, Glise H, Hallerback B, et al. Quality of life in patients with upper gastrointestinal symptoms. An improved evaluation of treatment regimens? *Scand J Gastroenterol* 1993;28:681–7.
- Revicki DA, Wood M, Wiklund I, et al. Reliability and validity of the gastrointestinal symptom rating scale in patients with gastroesophageal reflux disease. *Qual Life Res* 1998;7:75–83.
- Cohen J. Weighted kappa: Nominal scale agreement with provisions for scales disagreement of partial credit. *Psychol Bull* 1968;70:213–20.
- Dunn G. Design and analysis of reliability studies: The statistical evaluation of measurement errors. New York: Oxford University Press, 1989.
- Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press, 1995.

28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
29. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat Med* 1998;17:857–72.
30. Agreus L, Talley NJ, Svardsudd K, et al. Identifying dyspepsia and irritable bowel syndrome: The value of pain or discomfort, and bowel habit descriptors. *Scand J Gastroenterol* 2000;35:142–51.
31. Agreus L, Borgquist L. The cost of gastro-oesophageal reflux disease, dyspepsia and peptic ulcer disease in Sweden. *Pharmacoeconomics* 2002;20:347–55.
32. Agreus L, Svardsudd K, Nyren O, et al. The epidemiology of abdominal symptoms: Prevalence and demographic characteristics in a Swedish adult population. A report from the Abdominal Symptom Study. *Scand J Gastroenterol* 1994;29:102–9.
33. Stanghellini V, Schmier JK, Jones R, et al. Mode of administration of the patient assessment of upper gastrointestinal disorders-symptom severity (PAGI-SYM) and patient assessment of upper gastrointestinal disorders-quality of life (PAGI-QOL). Annual meeting of the American College of Gastroenterology; 2001.
34. 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.
35. Locke GR, Talley NJ, Weaver AL, et al. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;69:539–47.
36. Spechler SJ, Jain SK, Tendler DA, et al. Racial differences in the frequency of symptoms and complications of gastro-oesophageal reflux disease. *Aliment Pharmacol Ther* 2002;16:1795–800.
37. Wyshak G, Barsky AJ. Relationship between patient self-ratings and physician ratings of general health, depression, and anxiety. *Arch Fam Med* 1994;3:419–24.
38. Jachuck SJ, Brierley H, Jachuck S, et al. The effect of hypotensive drugs on the quality of life. *J R Coll Gen Pract* 1982;32:103–5.
39. Larue F, Fontaine A, Colleau SM. Underestimation and undertreatment of pain in HIV disease: Multicentre study. *Br Med J* 1997;314:23–8.
40. Melville JL, Miller EA, Fialkow MF, et al. Relationship between patient report and physician assessment of urinary incontinence severity. *Am J Obstet Gynecol* 2003;189:76–80.
41. Russo F, Di Pasquale B, Romano G, et al. International prostate symptom score: Comparison of doctor and patient. *Arch Ital Urol Androl* 1998;70:15–24.
42. McCammon KA, Kolm P, Main B, et al. Comparative quality-of-life analysis after radical prostatectomy or external beam radiation for localized prostate cancer. *Urology* 1999;54:509–16.
43. Calvo FA, Calvo A, Berrocal A, et al. Self-administered joint counts in rheumatoid arthritis: Comparison with standard joint counts. *J Rheumatol* 1999;26:536–9.
44. Heading RC, Wager E, Tooley PJ. Reliability of symptom assessment in dyspepsia. *Eur J Gastroenterol Hepatol* 1997;9:779–81.
45. Revicki DA, Crawley JA, Zodet MW, et al. Complete resolution of heartburn symptoms and health-related quality of life in patients with gastro-oesophageal reflux disease. *Aliment Pharmacol Ther* 1999;13:1621–30.
46. Steen IN, McColl E, Hutchinson A. Developing symptom-based outcome measures. In: Hutchinson A, McColl E, Christie M, Riccalton C, eds. *Outcome measures in primary and outpatient care*. Reading: Harwood Academic Publishers, 1996:23–44.
47. Holmes WF, MacGregor EA, Sawyer JP, et al. Information about migraine disability influences physicians' perceptions of illness severity and treatment needs. *Headache* 2001;41:343–50.
48. McColl E. Best practice in symptom assessment: A review. *Gut* 2004;53:iv49–54.
49. Crawley JA, Schmitt CM. How satisfied are chronic heartburn sufferers with their prescription medications? Results of the patient unmet needs survey. *J Clin Outcomes Manag* 2000;7:29–34.
50. Fallowfield L, Jenkins V, Farewell V, et al. Efficacy of a cancer research UK communication skills training model for oncologists: A randomised controlled trial. *Lancet* 2002;359:650–6.
51. Rutter DR, Iconomou G, Quine L. Doctor-patient communication and outcome in cancer patients: An intervention. *Psychol Health* 1996;12:57–71.
52. Tarrant C, Stokes T, Baker R. Factors associated with patients' trust in their general practitioner: A cross-sectional survey. *Br J Gen Pract* 2003;53:798–800.
53. Stewart MA. Effective physician-patient communication and health outcomes: A review. *CMAJ* 1995;152:1423–33.
54. Little FB, Koufman JA, Kohut RI, et al. Effect of gastric acid on the pathogenesis of subglottic stenosis. *Ann Otol Rhinol Laryngol* 1985;94:516–9.