

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Comparison of Daily versus Weekly Recording of Gastroesophageal Reflux Disease Symptoms in Patients with a Partial Response to Proton Pump Inhibitor Therapy

Anna Rydén, PhD^{1,*}, Olivia C. Leavy, MSc², Katarina Halling, MSc¹, Arthur A. Stone, PhD³

¹AstraZeneca Gothenburg, Mölndal, Sweden; ²Oxford Brookes University, Oxford, UK; ³Department of Psychology and the USC Dornsife Center for Self-Report Science, University of Southern California, Los Angeles, CA, USA

ABSTRACT

Background: The Reflux Symptom Questionnaire electronic Diary (RESQ-eD) and the Reflux Symptom Questionnaire 7-day recall (RESQ-7) are versions of a patient-reported outcome instrument that was developed and validated for measuring the frequency and intensity of symptoms in patients with gastroesophageal reflux disease (GERD) who have a partial response to proton pump inhibitor (PPI) therapy. **Objective:** The aim of these analyses was to assess the ability of the RESQ-7 to reproduce findings based on RESQ-eD reports of the same symptoms. **Methods:** These analyses are based on data from patients with GERD with a partial response to PPI (ClinicalTrials.gov identifier: NCT00703534). Participants completed the RESQ-eD twice daily for 7 days and the RESQ-7 on day 7. **Results:** Data from 446 patients were available for these analyses. Symptom-level analyses showed that, for intensity, mean domain scores were higher for the RESQ-7 (range 1.49–2.72) than for the RESQ-eD (range 1.45–2.57); for frequency, scores were lower for the RESQ-7 (range 2.58–4.82) than

for the RESQ-eD (range 4.22–6.24). Correspondence analyses of RESQ-7 and RESQ-eD mean domain scores indicated excellent agreement for intensity (correlation-concordance coefficient 0.77–0.83) and fair agreement for frequency (correlation-concordance coefficient 0.40–0.58). Mean RESQ-eD subscale intensity scores for GERD symptoms were higher for symptoms experienced during the daytime than for those occurring at nighttime. Symptom recall was not associated with peak or recency effects. **Conclusions:** Patients with GERD slightly overestimated the intensity of their reflux symptoms and markedly underestimated the frequency on weekly recall compared with twice-daily reporting.

Keywords: ecological validity, frequency, gastroesophageal reflux disease, intensity, patient-reported outcome instruments.

Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Having valid and reliable patient-reported outcome (PRO) instruments for measuring symptoms is paramount in medical research. Disease-specific PRO instruments are developed for particular target populations, and it is thus crucial that they capture the complete symptom pattern of the group of patients of interest [1]. The recall period is also an important aspect of a PRO instrument, and depends on the target population, symptoms, and the general setting in which the PRO instrument is to be implemented. Memory biases (cognitive heuristics) tend to affect retrospective evaluation of symptoms, with the majority of individuals being disproportionately influenced by the most intense and the most recent symptom events when asked to

recall the average intensity of symptoms, an effect termed the “peak-end” rule [2–4].

In its guidance for industry, the US Food and Drug Administration recommends that questionnaires generally use short recall periods or ask patients to describe their current or recent state [5]. Asking patients to think back over a long period can reduce the accuracy of recall for all but the most memorable or stressful events [6], and may also make it more likely that the response is influenced by the patients’ current state [7].

Patients with gastroesophageal reflux disease (GERD) typically have troublesome symptoms of heartburn and regurgitation [8], which usually resolve with proton pump inhibitor (PPI) therapy [9,10]. Results from a systematic review indicated that about 20% to 30% of patients with GERD who participated in primary care

Ethical approval: These post hoc analyses are based on data from the Patient-Reported Outcome Validation Study (ClinicalTrials.gov identifier: NCT00703534). All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written informed consent was obtained from all individual participants included in the study.

Conflict of interest: A. Rydén and K. Halling are employees of AstraZeneca Gothenburg, Mölndal, Sweden. O. C. Leavy is an MSc graduate from Oxford Brookes University, Oxford, UK. Arthur A. Stone is a consultant for ERT Inc. and is a Senior Scientist with the Gallup Organization.

* Address correspondence to: Anna Rydén, AstraZeneca Gothenburg, S-431 83, Mölndal, Sweden.

E-mail: anna.ryden@astrazeneca.com.

1098-3015/\$36.00 – see front matter Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2016.05.007>

studies had persistent heartburn or regurgitation during PPI therapy [11]. Symptoms confirmed by patients with a partial response to PPIs as being relevant to their GERD experience, in addition to heartburn and regurgitation, are hoarseness, cough, difficulty swallowing, and burping [12].

Most patients with GERD report having both daytime and nighttime symptoms [13]. Although the nature of daytime and nighttime GERD symptoms is similar in individuals with a partial response to PPI therapy, one study has reported a greater focus by patients on symptom intensity at night than during the day [14]. This increased attention paid to symptom intensity at night compared with during the day could potentially overestimate nighttime versus daytime symptom intensity on retrospective evaluation.

PRO instruments for evaluating GERD symptoms are used to measure symptom intensity and frequency in clinical trial settings, and are valuable tools when assessing symptoms and deciding on disease management in primary care [6]. When asking patients to recall their GERD symptoms, the duration of the recall period used may be influenced by practical considerations. While daily electronic diaries provide an effective way of assessing conditions such as GERD, in which symptoms fluctuate from day to day [15], this short recall period may not be practical for use in routine clinical practice or in pragmatic clinical trials, for which questionnaires with a 7-day recall period are commonly used. The Reflux Symptom Questionnaire electronic Diary (RESQ-eD) and the Reflux Symptom Questionnaire 7-day recall (RESQ-7) are PRO instruments that were developed and validated for measuring the intensity and frequency of GERD symptoms in patients with a partial response to PPIs [12,16]. Both versions of the instrument are conceptually and structurally identical. However, the RESQ-eD uses a twice-daily electronic recording, which makes it preferable in the clinical trial setting, whereas the RESQ-7 uses a 1-week recall period, making it more practical in routine clinical care [12,16].

The aim of the post hoc analyses presented in this article was to assess the ability of the RESQ-7 to reproduce faithfully findings based on twice-daily RESQ-eD reports of the same symptoms (i.e., its ecological validity [17]), using data from a large clinical study [16]. First, basic questions about the comparability and correlation of mean levels of aggregated real-time and recall measures were considered separately for intensity and frequency. Next, cognitive heuristics (peak and recency effects) that may be at work in the recall measures were assessed. Last, the role of nighttime symptoms in recall was examined.

Methods

Participants

These post hoc analyses are based on data from part 1 (8–12-day screening phase) of the PRO Validation Study, which included patients with GERD who had a partial response to PPI therapy ($N = 580$; mean age 48 years [range 19–70 years]; 58% women; ClinicalTrials.gov identifier: NCT00703534) [16]. To be eligible for inclusion in the PRO Validation Study, patients had to have a history of GERD symptoms for at least 6 months and a minimum of 4 weeks of PPI therapy. Patients must have also reported (using the RESQ-7) a minimum of 3 days on which they experienced an at least mildly intense burning feeling behind the breastbone, and/or unpleasant movement of material upward from the stomach over the previous 7 days.

Measures

Participants completed the RESQ-eD twice daily during the screening phase (in the morning and before bedtime) and the

RESQ-7 at the end of the screening phase. The RESQ-eD and the RESQ-7 were developed and validated for use in clinical trials in patients with a partial response to PPI therapy [12,16]. Both versions have the same 13 symptom items and use the same intensity scoring. However, the RESQ-eD is a twice-daily diary, whereas the RESQ-7 has a 7-day recall period.

For each version of the instrument, the 13 symptom items combine into four separate domains: Heartburn (five items: burning feeling behind breastbone; pain behind breastbone; heartburn; burning feeling in upper stomach; and pain in upper stomach); Regurgitation (four items: acid taste in mouth; bitter taste in mouth; unpleasant movement of material upward from the stomach; and stomach contents [liquid or food] moving upward to throat or mouth); Hoarseness, cough, difficulty swallowing (three items: hoarseness; cough; and difficulty swallowing); and Burping (one item: burping). The intensity of each symptom item is recorded using a six-point scale (0 = did not have; 1 = very mild; 2 = mild; 3 = moderate; 4 = moderately severe; and 5 = severe). The RESQ-7 also captures information on symptom frequency during the past week (have not had; 1 day; 2 days; 3–4 days; 5–6 days; daily [with a day being defined as a 24-hour period]).

Symptom Analyses

For RESQ-eD symptom analyses, data were assessed for the 7 days ending with the last day of the screening phase (i.e., the 7 days covered by the RESQ-7 recall period). Each day was defined as the 24-hour period covered by the symptom scores reported on the RESQ-eD in the evening (which covered symptoms experienced during the daytime) and those reported the next morning (which covered symptoms experienced during the preceding nighttime).

Daily (i.e., 24-hour) RESQ-eD symptom intensity was defined in the PRO Validation Study as the higher of the two daily item intensity scores (i.e., either the evening score or the following morning score, whichever was the higher). Mean weekly RESQ-eD symptom intensity was defined as the mean of the daily (24-hour) intensities during the week. Daily RESQ-eD domain intensity was defined as the mean (morning and evening) intensity values for all the items in that domain. Mean weekly RESQ-eD domain intensity was defined as the mean of the daily domain intensities during the week. RESQ-eD symptom frequency was computed on the basis of RESQ-eD intensity scores, with a daily item intensity score of at least very mild (score ≥ 1) indicating a symptomatic day for that symptom item. For domain frequency, an intensity score of at least very mild on a particular day for any item in that domain was taken as indicating a day with symptoms for that domain.

Statistical Analyses

To be included in the analyses, patients had to have completed both the RESQ-eD and the RESQ-7 during the screening phase. Imputation was performed on the RESQ-eD data set if a single observation was missing in the sequence of morning or evening registrations, by replacing the missing value with the larger of the two surrounding values. If two or more consecutive observations were missing, no imputation was performed and the patient was excluded from these analyses.

Mean \pm SD domain scores were calculated for the RESQ-eD and the RESQ-7. For weekly frequency data from the RESQ-7, midpoints were created for the RESQ-7 response options 3 to 4 days (3.5 days) and 5 to 6 days (5.5 days). For the weekly frequency data from the RESQ-eD, the variable was categorized such that the 3-day response option and the 4-day response option were recoded into the 3- to 4-day interval, and the 5-day

response option and the 6-day response option were recoded into the 5- to 6-day interval, and the midpoints were then used.

For correlation and symptom-level analyses, RESQ-eD intensity and frequency scores were compared using correlation-concordance coefficient (CCC) analyses and the paired *t* test, respectively. Given their similarity, CCC values were interpreted using the same ranges as used for standard intraclass correlation coefficient values (i.e., 0.4–0.75, fair-to-good; >0.75, excellent) [18–20]. To assess cognitive heuristics, day 7 (recency) and maximum (peak) diary day intensity scores for the RESQ-eD were compared with mean RESQ-7 intensity scores.

Mean evening RESQ-eD intensity scores (i.e., for symptoms experienced during the day) and mean morning RESQ-eD intensity scores (i.e., for symptoms experienced during the night) for each of the four symptom domains were assessed for comparability both during days 1 to 3 and during days 5 to 7, using the paired *t* test. In addition, the comparability of mean evening and mean morning intensity scores on days 1 to 3 versus on days 5 to 7 was assessed for each of the four symptom domains, using the paired *t* test.

Results

Data Characteristics

Of the 580 patients who participated in part 1 of the PRO Validation Study, 46 (8%) discontinued prematurely. A further 88 patients (15%) were excluded, either because they had two or more consecutive observations missing on the RESQ-eD (e.g., they were missing observations for the first day or the last

day of the screening phase) or because they did not complete the RESQ-7 at the end of the screening phase. A total of 446 patients (77%) completed both the RESQ-eD and the RESQ-7 as required, and were thus included in these analyses. The mean age of included patients was 49 ± 11 years and 40% were men.

Correlation Analyses

The CCC for mean intensity ranged from 0.77 to 0.83 for the four symptom domains (Table 1), indicating excellent correlation between the RESQ-eD and the RESQ-7 for each of the four domains for symptom intensity. For mean frequency, the CCC ranged from 0.40 to 0.58 for the four symptom domains (Table 2), indicating fair correlation between the RESQ-eD and the RESQ-7 for each of the four domains for symptom frequency.

Differences in Symptom Level between the RESQ-eD and the RESQ-7

Mean domain scores for intensity were lower for the RESQ-eD than for the RESQ-7 (Table 1). Paired *t*-test results for intensity showed a significant difference between mean RESQ-eD and RESQ-7 scores for each of the domains Heartburn, Regurgitation, and Burping (all $P < 0.001$; Table 1). Intensity scores did not differ significantly for the Hoarseness, cough, difficulty swallowing domain ($P = 0.277$).

Mean domain scores for frequency were higher for the RESQ-eD than for the RESQ-7 (Table 2). Paired *t*-test results for frequency showed a significant difference between mean RESQ-eD and RESQ-7 scores for all four domains (all $P < 0.001$).

Table 1 – Level and correspondence analyses of mean, day 7, and peak RESQ-eD vs RESQ-7 domain intensity scores.

Domain/statistics	Mean RESQ-eD vs RESQ-7 (n = 446)	Day 7 RESQ-eD vs RESQ-7 (n = 436)	Peak RESQ-eD vs RESQ-7 (n = 446)
Heartburn			
RESQ-eD score	2.38 ± 1.04	2.27	3.22
RESQ-7 score	2.54 ± 1.04	2.54	2.54
Score difference	−0.16	−0.27	0.62
<i>t</i>	5.47 [*]	7.06 [*]	20.51 [*]
CCC	0.80	0.72	0.63
Regurgitation			
RESQ-eD score	2.21 ± 1.15	2.12	3.06
RESQ-7 score	2.35 ± 1.16	2.35	2.35
Score difference	−0.14	−0.23	0.71
<i>t</i>	4.47 [*]	5.71 [*]	19.15 [*]
CCC	0.83	0.75	0.65
Hoarseness, cough, difficulty swallowing			
RESQ-eD score	1.45 ± 1.18	1.39	2.09
RESQ-7 score	1.49 ± 1.22	1.49	1.49
Score difference	−0.04	−0.10	0.60
<i>t</i>	1.09	2.44 [†]	14.93 [*]
CCC	0.82	0.78	0.71
Burping			
RESQ-eD score	2.57 ± 1.19	2.50	3.35
RESQ-7 score	2.72 ± 1.28	2.72	2.72
Score difference	−0.15	−0.22	0.63
<i>t</i>	3.75 [†]	4.25 [*]	14.05 [*]
CCC	0.77	0.68	0.63

CCC, correlation-concordance coefficient; RESQ-eD, Reflux Symptom Questionnaire electronic Diary; RESQ-7, Reflux Symptom Questionnaire 7-day recall.

* $P < 0.001$.

† $P < 0.05$.

Table 2 – Level and correspondence analyses of RESQ-eD vs RESQ-7 domain frequency scores (n = 446).

Domain/statistics	Frequency (d/wk)
Heartburn	
RESQ-eD score, mean ± SD	5.98 ± 1.43
RESQ-7 score, mean ± SD	4.49 ± 1.77
Score difference	1.49
t	21.89*
CCC	0.42
Regurgitation	
RESQ-eD score, mean ± SD	5.71 ± 1.79
RESQ-7 score, mean ± SD	4.09 ± 2.01
Score difference	1.62
t	22.92*
CCC	0.50
Hoarseness, cough, difficulty swallowing	
RESQ-eD score, mean ± SD	4.22 ± 2.40
RESQ-7 score, mean ± SD	2.58 ± 2.14
Score difference	1.64
t	20.68*
CCC	0.58
Burping	
RESQ-eD score, mean ± SD	6.24 ± 1.53
RESQ-7 score, mean ± SD	4.82 ± 2.18
Score difference	1.42
t	16.07*
CCC	0.40

CCC, correlation-concordance coefficient; RESQ-eD, Reflux Symptom Questionnaire electronic Diary; RESQ-7, Reflux Symptom Questionnaire 7-day recall.
* P < 0.001.

Testing of Cognitive Heuristics

Potential recency heuristic was assessed by comparing intensity scores for the last diary day (RESQ-eD day 7) with the weekly recall (RESQ-7). Mean RESQ-eD day 7 intensity scores differed significantly from mean RESQ-7 intensity scores for each of the four symptom domains (Table 1). The CCC values ranged from 0.68 to 0.78, indicating mostly fair-to-good correlation.

Potential peak heuristic was assessed by comparing maximum diary day intensity scores with the weekly recall (RESQ-7). Maximum RESQ-eD diary day intensity scores differed significantly from mean RESQ-7 intensity scores for each of the four symptom domains (Table 1). The CCC values ranged from 0.63 to 0.71, indicating fair-to-good correlation.

Comparability of Nighttime and Daytime Symptoms

Mean evening RESQ-eD intensity scores (i.e., for symptoms experienced during the day) were significantly higher than mean morning RESQ-eD intensity scores (i.e., for symptoms experienced during the night) for each of the four symptom domains, both during days 1 to 3 (Heartburn: $t = 8.810$; Regurgitation: $t = 4.833$; Hoarseness, cough, difficulty swallowing: $t = 4.603$; Burping: $t = 11.532$; all $P < 0.001$; $df = 445$) and during days 5 to 7 (Heartburn: $t = 9.315$; Regurgitation: $t = 6.955$; Hoarseness, cough, difficulty swallowing: $t = 5.704$; Burping: $t = 10.478$; all $P < 0.001$; $df = 445$).

For each of the four symptom domains, mean morning RESQ-eD intensity scores were not significantly different for days 1 to 3 versus days 5 to 7 (paired t test, Heartburn: $t = 1.839$ [$P = 0.067$]; Regurgitation: $t = 0.110$ [$P = 0.912$]; Hoarseness, cough, difficulty swallowing: $t = -0.485$ [$P = 0.628$]; Burping: $t = 0.085$ [$P = 0.933$];

$df = 445$). For three of the four symptom domains, mean evening RESQ-eD intensity scores were not significantly different for days 1 to 3 versus days 5 to 7 (Regurgitation: $t = -0.707$ [$P = 0.480$]; Hoarseness, cough, difficulty swallowing: $t = -0.675$ [$P = 0.500$]; Burping: $t = 1.136$ [$P = 0.257$]; $df = 445$). However, the difference between mean intensity scores for days 1 to 3 versus days 5 to 7 was borderline significant for Heartburn ($t = 1.988$ [$P = 0.047$]; $df = 445$).

Discussion

In these analyses, patients with GERD slightly overestimated the intensity of their reflux symptoms and markedly underestimated symptom frequency on weekly recall when compared with daily reporting. Correlation between the mean RESQ-eD scores and RESQ-7 scores was excellent for symptom intensity, suggesting that individuals with high (or low) scores on the RESQ-eD also had high (or low) scores on the RESQ-7. The correlation for symptom frequency was only fair (“moderate”).

The slightly higher symptom intensity reported on weekly recall compared with daily reporting is not surprising, and is consistent with other studies assessing recall periods for pain, headache, respiratory symptoms, fatigue, and the impact of symptoms on daily life [2,21–24]. In the current analyses, the difference in mean domain intensity scores between the two versions of the instrument was small, differing by, at most, 0.16 units on a scale ranging from 0 to 5. Thus, although patients overestimated the intensity of their GERD symptoms on weekly recall compared with daily reporting, the discrepancy was slight and is unlikely to be of much relevance in clinical practice, although it could be important in clinical trials.

On average, patients were worse at recalling the frequency than the intensity of their symptoms. Mean domain scores for frequency (which could range from 0 to 7 d/wk) were between 1.42 and 1.64 d/wk lower for the RESQ-7 than for the RESQ-eD. Thus, patients tended to underestimate the number of days per week on which they experienced symptoms by about 1.5 days when recalling their GERD symptoms from the previous week. Few published data are available regarding the effects of recall period on the reporting of symptom frequency [24,25]. Generally, the recalling of frequency information is affected more by omission than by inappropriate inflation, leading to underestimation rather than overestimation when recalling symptom frequency [25]. However, in a study conducted in patients with migraine, a similar mean headache frequency was obtained from daily and 4-weekly reporting (15.1 vs 14.7 days in the 4-week period, respectively) [24], suggesting that effects of recall period on reporting of symptom frequency may depend on the type of symptom that is being assessed. In that study, the accuracy of recall was not affected by whether symptoms occurred more or less frequently [24]. The discrepancy between daily reporting and weekly recall of symptom frequency supports the use of the RESQ-eD rather than the RESQ-7 for assessing GERD symptom frequency in the clinical trial setting. With its longer recall period, the RESQ-7 still has a place in routine clinical practice or when conducting large, pragmatic trials, where weekly recording may be more practical than daily recording; however, patients’ underestimation of GERD symptom frequency on recall needs to be kept in mind when interpreting results.

Our analyses went beyond the basic question of ecological validity and examined cognitive heuristics that may be at work in the RESQ-eD. Retrospective reporting of symptom intensity has been reported to be associated with the level of the most intense (“peak”) event that occurred during the recall period, an effect that is likely to be exacerbated by patients ignoring occasions when the symptom is absent, thus giving peak events undue

influence [2]. Symptom recall can also be affected by the level of intensity of the most recent (“end”) event [25]. In the current analyses, both maximum RESQ-eD intensity scores and intensity scores recorded on day 7 on the RESQ-eD differed significantly from mean RESQ-7 intensity scores for each of the four symptom domains. Thus, neither the maximum symptom intensity nor the intensity on the last day of reporting had a pronounced influence on recall. Agreement of both peak RESQ-eD scores and day 7 RESQ-eD scores with mean RESQ-7 scores was fair-to-excellent, indicating that the peak daily symptom score and the score on the last day of reporting correlated well with the weekly recall.

Mean RESQ-eD domain intensity scores for GERD symptoms were higher for the evening reporting (which covered symptoms experienced during the daytime) than for the morning reporting (which covered symptoms experienced during the preceding nighttime). In contrast, information gathered as part of an interview study in individuals with a partial response to PPI therapy showed that patients experienced a greater intensity of heartburn and regurgitation during the night than during the day [14]. In that study, patients reported that the most troublesome effect of GERD symptoms on their daily lives was the impact on sleep [14].

A strength of our post hoc analyses is that they included a well-defined patient population. Limitations are that they were not originally planned as part of the PRO Validation Study, and that they did not compare the use of the RESQ-eD and the RESQ-7 for capturing changes over time. Such data would be of particular relevance when considering the suitability of these tools for assessing response to treatment. The need to recode response options for weekly symptom frequency to allow comparison between the two instruments presents another potential limitation.

In conclusion, patients with GERD slightly overestimated the intensity of their reflux symptoms and markedly underestimated the frequency when recalling symptoms from the previous week. Symptom recall using the RESQ-7 was not associated with peak or recency effects.

Acknowledgment

We thank Dr. Reza Oskrochi for his supervision of Olivia C. Leavy's MSc project.

Source of financial support: This study was funded by AstraZeneca Gothenburg, Sweden. Writing support was provided by Dr. Anja Becher, from Oxford PharmaGenesis, Oxford, UK, and was funded by AstraZeneca Gothenburg, Sweden.

REFERENCES

- Vakil NB, Halling K, Becher A, Rydén A. Systematic review of patient-reported outcome instruments for gastroesophageal reflux disease symptoms. *Eur J Gastroenterol Hepatol* 2013;25:2–14.
- Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain* 2004;107:61–9.
- Stone AA, Broderick JE, Kaell AT, et al. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritis? *J Pain* 2000;1:212–7.
- Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers Soc Psychol Bull* 2005;31:1340–6.
- Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available from: <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf>. [Accessed March 11, 2015].
- Chassany O, Shaheen NJ, Karlsson M, et al. Systematic review: symptom assessment using patient-reported outcomes in gastroesophageal reflux disease and dyspepsia. *Scand J Gastroenterol* 2012;47:1412–21.
- Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Curr Med Res Opin* 2009;25:929–42.
- Vakil N, Veldhuyzen van Zanten S, Kahrilas P, et al. The Montreal definition and classification of gastro-esophageal reflux disease (GERD) – a global evidence-based consensus. *Am J Gastroenterol* 2006;101:1900–20.
- van Pinxteren B, Numans ME, Bonis PA, Lau J. Short-term treatment with proton pump inhibitors, H₂-receptor antagonists and prokinetics for gastro-oesophageal reflux disease-like symptoms and endoscopic negative reflux disease. *Cochrane Database Syst Rev* 2004;4:CD002095.
- Donnellan C, Sharma N, Preston C, Moayyedi P. Medical treatments for the maintenance therapy of reflux oesophagitis and endoscopic negative reflux disease. *Cochrane Database Syst Rev* 2005;2:CD003245.
- El-Serag H, Becher A, Jones R. Systematic review: persistent reflux symptoms on proton pump inhibitor therapy in primary care and community studies. *Aliment Pharmacol Ther* 2010;32:720–37.
- Rydén A, Denison H, Karlsson M, Vakil N. Development and validation of a patient-reported outcome instrument in partial responders to proton pump inhibitors. *Scand J Gastroenterol* 2013;48:1018–26.
- Shaker R, Castell DO, Schoenfeld PS, Spechler SJ. Nighttime heartburn is an under-appreciated clinical problem that impacts sleep and daytime function: the results of a Gallup survey conducted on behalf of the American Gastroenterological Association. *Am J Gastroenterol* 2003;98:1487–93.
- Rydén A, Martin M, Halling K, Niklasson A. Night-time symptoms and their impact on sleep in patients with gastroesophageal reflux disease who have a partial response to proton pump inhibitors: a qualitative patient interview study. *Patient* 2013;6:307–14.
- McCull E. Best practice in symptom assessment: a review. *Gut* 2004;53 (Suppl. 4):iv49–54.
- Vakil N, Björck K, Denison H, et al. Validation of the reflux symptom questionnaire electronic diary in partial responders to proton pump inhibitor therapy. *Clin Transl Gastroenterol* 2012;3:e7.
- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1–32.
- Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trials* 1991;12(Suppl. 4):142S–58.
- Bennett AV, Patrick DL, Bushnell DM, et al. Comparison of 7-day and repeated 24-h recall of type 2 diabetes. *Qual Life Res* 2011;20:769–77.
- Fleiss JL. *The Design and Analysis of Clinical Experiments* (Wiley Classics Library ed.). New York: Wiley, 1999.
- Broderick JE, Schneider S, Schwartz JE, Stone AA. Interference with activities due to pain and fatigue: accuracy of ratings across different reporting periods. *Qual Life Res* 2010;19:1163–70.
- Broderick JE, Schwartz JE, Vikingstad G, et al. The accuracy of pain and fatigue items across different reporting periods. *Pain* 2008;139:146–57.
- Bennett AV, Amtmann D, Diehr P, Patrick DL. Comparison of 7-day recall and daily diary reports of COPD symptoms and impacts. *Value Health* 2012;15:466–74.
- McKenzie JA, Cutrer FM. How well do headache patients remember? A comparison of self-report measures of headache frequency and severity in patients with migraine. *Headache* 2009;49:669–72.
- Gorin AA, Stone AA. Recall biases and cognitive errors in retrospective self-reports: a call for momentary assessments. In: Baum A, Revenson TA, Singer JE, eds., *Handbook of Health Psychology*. Mahwah, NJ: Erlbaum, 2001.