

# Validation of the Reflux Symptom Questionnaire Electronic Diary in Partial Responders to Proton Pump Inhibitor Therapy

Nimish Vakil, MD<sup>1</sup>, Karin Björck, MSc<sup>2</sup>, Hans Denison, MD<sup>2</sup>, Katarina Halling, MSc<sup>3</sup>, Maria Karlsson, MSc<sup>2</sup>, Jean Paty, PhD<sup>4</sup>, Debra G. Silberg, MD, PhD<sup>5</sup> and Anna Rydén, PhD<sup>2</sup>

**OBJECTIVES:** We aimed to develop and validate the Reflux Symptom Questionnaire electronic Diary (RESQ-eD) for use in clinical trials in patients with a partial response to proton pump inhibitor (PPI) therapy, using methods that meet US Food & Drug Administration (FDA) regulatory standards.

**METHODS:** Patient interviews were performed to elicit new items and evaluate existing items from the Reflux Disease Questionnaire. The instrument's measurement properties were evaluated, based on data from two clinical trials of patients with gastroesophageal reflux disease (GERD) with a partial response to PPIs who received lesogaberan or placebo as an add-on to PPI therapy.

**RESULTS:** The content validity phase resulted in 13 RESQ-eD items. Principal component analysis supported a four-domain structure. All domains had a high inter-item correlation (Cronbach's alpha lower 95% confidence limit: 0.87–0.95). Test-retest reliability was good to excellent (intraclass correlation coefficient: 0.65–0.85). Convergent and discriminant validity was confirmed by correlation assessments referencing the Gastrointestinal Symptom Rating Scale. The RESQ-eD demonstrated a good ability to capture change in mean intensity and proportion of symptom-free days. Confirmatory psychometric evaluation verified internal consistency reliability, test-retest reliability, and ability to capture change.

**CONCLUSIONS:** The RESQ-eD demonstrated good content validity and psychometric properties in the clinical trial setting in patients with GERD who have a partial response to PPI therapy. To our knowledge, the RESQ-eD is the first electronic symptom diary for use in partial responders to PPI that has been developed in line with the FDA guidance on patient-reported outcomes.

*Clinical and Translational Gastroenterology* (2012) 3, e7; doi:10.1038/ctg.2012.1; published online 26 January 2012

**Subject Category:** Esophagus

## INTRODUCTION

Most patients with gastroesophageal reflux disease (GERD) experience resolution of their heartburn and regurgitation when on proton pump inhibitor (PPI) therapy.<sup>1,2</sup> However, a recent systematic review showed that, in interventional primary care trials, approximately 20–30% of patients with GERD experience only a partial response of their heartburn or regurgitation symptoms to PPI therapy.<sup>3</sup> Potential pharmacological targets in the treatment of partial response to PPI therapy include esophageal mechanisms of visceral sensitivity and nociception, and inhibition of transient lower esophageal sphincter relaxations (TLESRs).<sup>3–6</sup> Other therapeutic approaches to partial response to existing medications include more potent or longer acting acid inhibitors. Using esophageal impedance-pH monitoring, clinical trials in patients with GERD and in healthy volunteers have shown that the reflux inhibitor lesogaberan (AZD3355) reduces the numbers of TLESRs, increases lower esophageal sphincter pressure and decreases the number of reflux episodes.<sup>7,8</sup> However, the majority of reflux episodes that are detected using impedance-pH monitoring are asymptomatic and the

role of new therapeutic approaches in partial responders to PPIs is uncertain.<sup>9,10</sup>

GERD is a symptom-driven disease that has to be evaluated based on the presence, frequency, and severity of GERD symptoms.<sup>11</sup> These are best captured and measured using patient-reported outcome (PRO) instruments. The US Food and Drug Administration (FDA) recently published its guidance for industry on the use of PRO instruments in medical product development to support labelling claims.<sup>12</sup> The guidance emphasizes the importance of obtaining evidence of the relevancy (i.e., content validity) of the PRO instrument from the target patient population, and of ensuring that the population studied in the PRO instrument development and documentation process is comparable with that in the clinical study setting in which the instrument is to be used (i.e., that the instrument is “fit for purpose”).

The FDA guidance calls for patient input during the development of PRO instruments and establishment of their content validity.<sup>12</sup> To assess content validity of a PRO instrument, the FDA intends to review the derivation of items, transcripts from focus groups and cognitive interviews, and the composition of patient groups involved in content

<sup>1</sup>University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; <sup>2</sup>AstraZeneca R&D, Mölndal, Sweden; <sup>3</sup>PRO Consulting, Askim, Sweden; <sup>4</sup>PRO Consulting, Pittsburgh, PA, USA and <sup>5</sup>AstraZeneca LP, Wilmington, DE, USA

Correspondence: N Vakil, MD, University of Wisconsin School of Medicine and Public Health, Aurora Summit Hospital, 36500 Aurora Drive, Summit, WI 53066, USA. E-mail: nvakil@wisc.edu

Received 8 December 2011; accepted 27 December 2011

development. Additional measurement properties that will need to be established are internal consistency and test-retest reliability, convergent and discriminant validity, known-groups validity, and responsiveness to change. Furthermore, the FDA guidance encourages planning for clinical trial interpretation using an *a priori* responder definition.

Several instruments have undergone psychometric evaluation in patients with GERD for use in clinical practice and in clinical trials. One of the most extensively evaluated instruments in GERD is the Reflux Disease Questionnaire (RDQ).<sup>13–15</sup> However, neither the RDQ nor, as far as we know, any other PRO instrument has been developed for the specific target population of patients with GERD with a partial response to PPI therapy. Consequently, these instruments may not capture the complete symptom pattern of this specific group of patients. This paper describes how methodology outlined in recent regulatory guidelines was applied in the development and psychometric evaluation of a new PRO instrument, the Reflux Symptom Questionnaire electronic Diary (RESQ-eD), intended for use in clinical trials in patients with a partial response to PPI therapy. The RDQ was used as a basis for the development of the RESQ-eD.

## METHODS

**RESQ-eD development.** The RESQ-eD was developed in two stages: (i) patient interviews were performed to elicit new items and to evaluate existing items from the RDQ (content validity); and (ii) evaluation of the instrument's measurement properties (including reliability, validity, and responsiveness) were performed. Both of these stages included an exploratory and a confirmatory phase.

**Exploratory interview study: exploratory content validity phase.** Eighty patients diagnosed with GERD, or having GERD according to their physician's judgement, participated in an independent interview study in the USA, France, and Japan. All patients had remaining symptoms of heartburn and/or regurgitation despite continuous PPI therapy during the previous 6 weeks. Patients were identified through a commercial recruitment agency that enlisted gastroenterologists and primary care physicians to help with recruitment. The age range was 25–85 years, and 50 (63%) patients were women. Patients were excluded if they had lower gastrointestinal symptoms consistent with those of irritable bowel syndrome or if they had a history of, or current, peptic ulcer disease. Symptom concepts relevant to partial responders to PPI therapy were elicited in four languages (US English, US Spanish, French, and Japanese). Semi-structured guides that contained a series of open-ended questions in 48 individual interviews and four focus group interviews (32 patients) were used. Patient statements were coded to identify relevant symptom concepts, and saturation of concepts was measured. Concept selection and confirmation were guided by expert opinion (one US and one Australian gastroenterologist) and a review of published GERD studies, including observational studies, clinical trials, and instrument development and validation studies.

## PRO Validation Study: confirmatory content validity phase and exploratory psychometric validation phase.

The PRO Validation Study (ClinicalTrials.gov identifier: NCT00703534) was conducted between May and December 2008 at 77 centers in the USA. The aim of the study was to evaluate prospectively the domain structure performance of the instrument in the target patient population. The study had a randomized, double-blind, placebo-controlled, parallel-group design consisting of an 8–12-day screening phase (part 1), followed by randomization into a 4-week treatment phase (part 2). As an add-on to PPI therapy (see below), patients received either lesogaberan 65 mg twice daily or matching placebo.

Overall, 580 patients were eligible for part 1 (mean age: 48 years [range: 19–70]; 58% women), and 478 patients from part 1 were further randomized into part 2 (mean age: 49 years [range: 19–70]; 59% women). Patients were eligible for inclusion if they had a history of GERD symptoms for at least 6 months and had received a minimum of 4 weeks of PPI therapy (8 weeks of treatment if diagnosed with reflux esophagitis within the previous 8 weeks), individually optimized according to the physician's judgement, within the approved dose range for any GERD indication. Patients whose symptoms did not improve at all after PPI therapy were excluded from the study. For enrollment into part 1 and randomization into part 2, patients should have experienced at least 3 days of a burning feeling behind the breastbone and/or unpleasant movement of material upwards from the stomach over the previous 7 days; symptoms had to be of at least mild intensity (score  $\geq 2$  on a 6-point scale ranging from 0 [did not have] to 5 [severe]). For inclusion into part 1, symptoms were assessed with an instrument with the same items as the RESQ-eD, but with a 7-day recall period (RESQ-7). For inclusion into part 2, the RESQ-eD was used. Patients were required to complete the RESQ-eD twice daily (morning and before bedtime) during the study period to assess symptom intensity (ranging from 0 [did not have] to 5 [severe]). In addition, patients were requested to complete the Gastrointestinal Symptom Rating Scale (GSRS) at enrollment and at randomization, and the Overall Treatment Evaluation (OTE) 2 weeks and 4 weeks after randomization, including an evaluation of the importance of perceived change.

Confirmatory individual interviews were conducted with 42 of the patients included in the PRO Validation Study, to confirm the relevance of the symptom items that were generated during the exploratory interview study. Patients had been asked at enrolment at participating clinical sites whether they would be interested in taking part in two interviews about their experience with GERD in addition to the other study procedures. The draft instrument at this point contained 13 symptoms to be rated on a 6-point scale ranging from 0 to 5. Cognitive interviews were performed after patients had completed the RESQ-eD twice daily for about 1 week. Spontaneous reports of symptoms were noted, and patients were also asked whether symptoms measured in the study were relevant to their GERD experience. As part of the individual interviews, patients were also asked about their treatment expectations before their treatment commenced, and were asked about the importance and meaningfulness

of any improvement in their GERD symptoms at the end of treatment.

**Lesogaberan Dose-finding Study—confirmatory psychometric validation phase.** Psychometric evaluation of the RESQ-eD was confirmed in the lesogaberan Dose-finding Study (ClinicalTrials.gov identifier: NCT01005251), which was conducted between October 2009 and July 2010 at multiple centers in the USA, Canada, and Europe. The study had a randomized, double-blind, placebo-controlled, parallel-group design consisting of an 8–26-day screening phase, followed by randomization into a 4-week treatment phase. Patient inclusion and exclusion criteria were the same as for the PRO Validation Study (described above), except that symptoms for enrollment and randomization had to be of at least moderate intensity (score  $\geq 3$  on a 6-point scale ranging from 0 [did not have] to 5 [severe]) and PPI treatment had to be within the GERD label for the country of the enrolling site. In addition to the USA and Canadian English language versions, the following language versions of the RESQ-eD were translated and linguistically validated using procedures recommended by the International Society for Pharmacoeconomics and Outcomes Research Task Force for Translation and Cultural Adaption,<sup>16</sup> including forward/backward translation and cognitive debriefing: French (France and Canada), German, Hungarian, Romanian, Spanish (USA), Latvian, and Russian (Latvia). In total, 661 patients were randomized in the lesogaberan Dose-finding Study (mean age: 48 years [range: 18–70 years]; 57% women). Most ( $n=504$ ; 76.2%) participants were from the USA. Patients were required to complete the RESQ-eD twice daily (morning and before bedtime) during the study period, and to complete the OTE 2 weeks and 4 weeks after randomization.

#### Reference measures for validation

**Overall Treatment Evaluation.** The OTE is an instrument that rates the magnitude and importance of changes in symptoms on a 15-point scale.<sup>17,18</sup> For magnitude of change, patients were grouped as follows: “worse” (score:  $-7$  to  $-2$ ); “unchanged” (score:  $-1$ ,  $0$ ,  $+1$ ); “small improvement” (score:  $+2$ ,  $+3$ ); “moderate improvement” (score:  $+4$ ,  $+5$ ); and “large improvement” (score:  $+6$ ,  $+7$ ). For importance of change, patients were grouped as follows: “deterioration of any importance” (score:  $-7$  to  $-2$ ); “no important change” (score:  $-1$ ,  $0$ ,  $+1$ ); “improvement of small importance” (score:  $+2$ ,  $+3$ ); “improvement of moderate importance” (score:  $+4$ ,  $+5$ ); and “improvement of large importance” (score:  $+6$ ,  $+7$ ). The OTE is derived from the Global Ratings of Change Questionnaire.<sup>19,20</sup>

**Gastrointestinal Symptom Rating Scale.** The GSRS is an instrument with a 1-week recall that assesses symptom severity using a 7-grade Likert scale, ranging from 1 (“no discomfort at all”) to 7 (“very severe discomfort”).<sup>21</sup> The instrument consists of 15 items that are clustered into five domains: Reflux, Abdominal pain, Indigestion, Diarrhea, and Constipation. Of relevance to the present Validation Study, the Reflux domain consists of heartburn and regurgitation items, and the Indigestion domain includes an item for

burping. Based on percentiles from previous data in order to get three groups of similar size, GSRS domain scores (derived as the mean of the separate item scores) were grouped as follows: “low” (score:  $\leq 4.0$ ); “medium” (score:  $> 4.0$  and  $\leq 5.0$ ); and “high” (score:  $> 5.0$ ).

**Statistics.** Statistical analyses were performed using SAS<sup>®</sup> version 8.2 (SAS Institute Inc., Cary, NC, USA). Unless otherwise stated, psychometric properties of the RESQ-eD were evaluated in the 7 days before randomization, (i.e., before the start of treatment). Psychometric evaluation of the RESQ-eD was performed on data irrespective of treatment.

**Calculation of variables.** Missing values were imputed if the length of the gap was one registration in the sequence of morning and bedtime registrations respectively, in which case the value was imputed as the maximum of the two surrounding values in the respective sequence. Daily intensity values were then derived as the maximum of a bedtime value and the following morning value, but set to missing if one of the two values were still missing after imputation of morning and bedtime registrations. Mean intensity over a time period was then calculated unless the proportion of daily values missing was greater than 30%. In calculations of proportion of days without symptoms or symptoms above a certain intensity, a day with missing values was regarded as a day with symptoms above the given threshold. RESQ-eD domain intensity scores were derived as the mean value of the item intensity scores within the domain.

**Principal component analysis.** Principal component analysis with varimax and oblimin rotation was performed on variables of intensity and frequency of symptoms (mean intensity and number of days with symptoms, respectively) to identify items that correlate strongly enough to form homogeneous domains.

**Reliability.** Internal consistency (i.e., to what extent items within a domain are interrelated) was evaluated using Cronbach’s alpha coefficients. Cronbach’s alpha is a measure of reliability and an alpha value  $> 0.7$  was considered as supporting internal consistency reliability.<sup>22</sup> Test-retest reliability (i.e., the stability of an instrument over time) was assessed between randomization and 2 weeks after randomization, using intraclass correlation coefficients (ICC) in patients considered to be in a stable condition (“unchanged” according to the OTE classification). ICC values of  $0.4$ – $0.75$  were considered to represent fair to good reliability and values  $> 0.75$  represented excellent reliability.<sup>23</sup>

**Construct validity.** Spearman’s rho was used to evaluate convergent and discriminant validity. Convergent validity refers to how well constructs that theoretically should be related to each other are observed to be related. Discriminant validity can be viewed as the counterpart to convergent validity and is based on the assumption that constructs theoretically not related to each other are not observed to be related to each other and therefore demonstrate low correlations. The GSRS domains and single item “burping”

were used as references for these assessments. Known-groups validity (i.e., an instrument's ability to differentiate between groups of patients known to be clinically distinct) was evaluated graphically using the GSRS reflux domain and burping item scores to separate patients into subgroups by symptom severity.

**Responsiveness to change.** The instrument's responsiveness to change was evaluated by graphical depiction of the effect size using the variables intra-patient change in mean domain intensity from baseline (the 7 days before randomization) to the treatment period (approximately 4 weeks treatment) and the intra-patient change in proportion of symptom-free days from baseline to treatment period. The effect size was defined as the mean individual change over time divided by the standard deviation at baseline, and was calculated on subgroups of patients determined by the OTE classifications at the end of treatment.

**Ethical considerations.** The studies were performed in accordance with the Declaration of Helsinki and the International Conference on Harmonisation Good Clinical Practice guide. The study protocol was approved by a central, free-standing board—the Schulman Associates Institutional Review Board—prior to patient enrollment. All patients included in the analysis gave their written informed consent prior to participation.

## RESULTS

**Exploratory interview study: exploratory content validity phase.** During the exploratory interviews, all six symptom expressions from the RDQ were mentioned by at least 20 patients (Table 1). Six additional symptom expressions emerged from the interviews: heartburn, difficulty swallowing, cough, nausea, hoarseness, and burping. "Bitter taste in mouth" was mentioned by fewer than 20 patients, but was

**Table 1** The 13 items of the Reflux Disease Questionnaire electronic Diary

RESQ-eD item	Item retained from RDQ	New item added to RESQ-eD
Burning feeling behind the breastbone	✓	
Pain, breastbone	✓	
Heartburn		✓
Acid taste in the mouth	✓	
Bitter taste in the mouth		✓
Unpleasant movement of material upwards from the stomach	✓	
Stomach contents (liquid or food) moving upwards towards your throat or mouth		✓
Burning feeling in the center of the upper stomach	✓	
Pain in the center of the upper stomach	✓	
Hoarseness		✓
Cough		✓
Difficulty swallowing		✓
Burping		✓

RDQ, Reflux Disease Questionnaire; RESQ-eD, Reflux Disease Questionnaire electronic Diary.

deemed relevant based on empirical evidence and expert input. "Stomach contents (liquid or food) moving upwards towards your throat or mouth" was added, based on patient and expert input, to specify the location of regurgitation. All expressions, except nausea, which was deemed too non-specific, were included in the RESQ-eD. All RDQ items were retained. Literature searches and expert opinion confirmed the proposed 13 RESQ-eD items as relevant. Furthermore, a vast majority of the interviewed patients had no difficulty understanding the instructions or response options. The exploratory content validity phase therefore resulted in 13 items. Items were scored for intensity on a 6-point scale (0 = did not have; 1 = very mild; 2 = mild; 3 = moderate; 4 = moderately severe; 5 = severe).

### PRO Validation Study: confirmatory content validity phase and exploratory psychometric validation phase

**Confirmatory interviews.** Individual interviews confirmed the relevance of the symptom items of the RESQ-eD. Patients collectively reported all symptoms included in the RESQ-eD, and endorsed all items as being relevant to their experience of GERD. The majority (86%) of patients understood all 13 items of the RESQ-eD and did not make suggestions for modifications. Before treatment, 53% of responses from the patients indicated that any improvement in GERD symptoms would be considered important and meaningful; following treatment, 91% (29/32) of interviewed patients who had any improvement in their GERD symptoms expressed that the change was both important and meaningful.

**Principal component analyses.** In the principal component analyses using varimax and oblimin rotation, an overall evaluation of scree-plots, eigenvalues, and proportion of the total variance accounted for suggested retaining two or three components for rotation (Table 2).

Also, the three-factor solution was conceptually appealing when reviewing the rotated factor pattern. Taken together, the principal component analysis and prior assumptions converged in suggesting that a three-component solution may be appropriate. The highest factor loading for the item burping was found in the component with regurgitation items. However, the loadings of burping were not of the same magnitude as the loadings of the other items in that component.

A four-domain structure was thus deemed most suitable for the 13 items of the RESQ-eD, consisting of Heartburn (5 items: burning feeling, breastbone; pain, breastbone; heartburn; burning feeling, upper stomach; pain, upper stomach), Regurgitation (4 items: acid taste in mouth; bitter taste in mouth; unpleasant movement of material; stomach contents, liquid or food), Hoarseness, cough, difficulty swallowing (3 items), and Burping (1 item). Subsequent analyses of reliability, validity, and responsiveness of the RESQ-eD were conducted on the four symptom domains.

Frequency tabulation analyses of the number of days with RESQ-eD symptoms by domain in the 7 days before randomization showed that 95% of enrolled eligible patients had Heartburn, 90% had Regurgitation, 77% had Hoarseness, cough, or difficulty swallowing, and 90% had Burping for 5–7 days during the week. All 13 items were also combined into an Overall symptoms domain showing that 100% of



**Table 2** Principal component analysis of the mean intensity and frequency of symptoms in the Reflux Disease Questionnaire electronic Diary during the 7 days prior to randomization (all enrolled eligible patients,  $N=580$ )

RESQ-eD item	Intensity			Frequency		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Burning feeling, breastbone	<b>86</b>	27	24	13	<b>89</b>	11
Pain, breastbone	<b>83</b>	25	27	19	<b>77</b>	18
Heartburn	<b>82</b>	35	18	16	<b>84</b>	0
Acid taste in mouth	34	<b>80</b>	31	<b>79</b>	24	26
Bitter taste in mouth	33	<b>79</b>	34	<b>79</b>	21	27
Unpleasant movement of material	35	<b>82</b>	29	<b>88</b>	19	19
Stomach contents, liquid or food	33	<b>82</b>	30	<b>86</b>	17	16
Burning feeling, upper stomach	<b>77</b>	39	24	35	<b>71</b>	18
Pain, upper stomach	<b>75</b>	40	27	43	<b>62</b>	26
Hoarseness	23	30	<b>85</b>	21	18	<b>86</b>
Cough	23	27	<b>83</b>	15	9	<b>84</b>
Difficulty swallowing	32	34	<b>76</b>	34	15	<b>74</b>
Burping	44	<b>60</b>	24	<b>52</b>	25	12

RESQ-eD, Reflux Disease Questionnaire electronic Diary.

**Table 3** Internal consistency and test-retest reliability of the Reflux Disease Questionnaire electronic Diary

Domain	PRO Validation Study		Dose-finding Study	
	Cronbach's alpha (95% CI) ( $n=580$ )	ICC (95% CI) ( $n=126$ )	Cronbach's alpha (95% CI) ( $n=625$ )	ICC (95% CI) ( $n=188$ )
Heartburn	0.94 (0.94–0.95)	0.65 (0.54–0.74)	0.93 (0.92–0.94)	0.80 (0.75–0.85)
Regurgitation	0.95 (0.94–0.96)	0.74 (0.65–0.81)	0.89 (0.88–0.91)	0.81 (0.75–0.85)
Hoarseness, cough, difficulty swallowing	0.89 (0.87–0.91)	0.85 (0.79–0.89)	0.85 (0.83–0.87)	0.85 (0.81–0.89)
Burping	NA	0.73 (0.63–0.8)	NA	0.77 (0.70–0.82)
Overall symptoms	0.96 (0.95–0.96)	0.75 (0.66–0.82)	0.93 (0.92–0.93)	0.81 (0.75–0.85)

CI, confidence interval; ICC, intraclass correlation coefficient; NA, not applicable (single item domain); PRO, patient-reported outcome.

**Table 4** Correlation between Reflux Disease Questionnaire electronic Diary and Gastrointestinal Symptom Rating Scale domains/items (all enrolled eligible patients,  $N=580$ )

RESQ-eD domain <sup>a</sup>	PRO Validation Study GSRS domain/item <sup>b</sup>					
	Reflux	Abdominal pain	Indigestion	Diarrhoea	Constipation	Burping
Heartburn	<b>0.60</b>	0.57	0.52	0.28	0.36	0.37
Regurgitation	<b>0.64</b>	0.50	0.48	0.28	0.32	0.35
Hoarseness, cough, difficulty swallowing	0.44	0.49	0.42	0.29	0.34	0.28
Burping	0.52	0.47	<b>0.68</b>	0.27	0.32	<b>0.71</b>
Overall symptoms	<b>0.66</b>	<b>0.60</b>	0.58	0.33	0.39	0.43

<sup>a</sup>Weekly RESQ-eD intensity score of the 7 days before randomization. <sup>b</sup>Reported at randomization. GSRS, Gastrointestinal Symptom Rating Scale; PRO, patient-reported outcome; RESQ-eD, Reflux Disease Questionnaire electronic Diary.

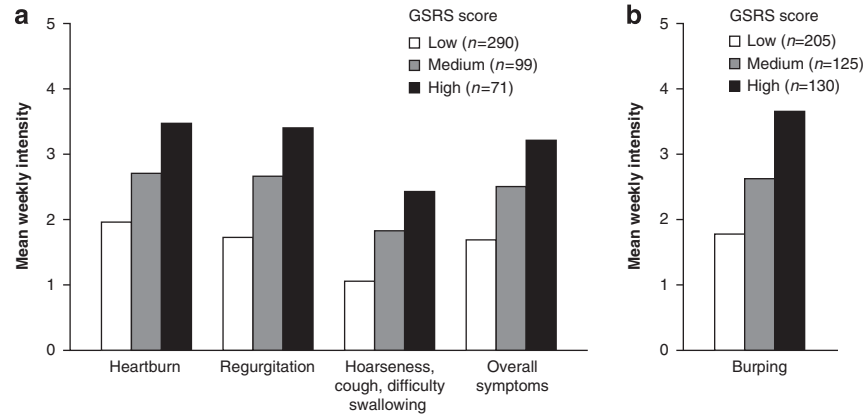
enrolled eligible patients had symptoms for 5–7 days during the week.

**Reliability.** All symptom domains of the RESQ-eD had a high inter-item correlation, with the lower 95% confidence limit of Cronbach's alpha in the range 0.87–0.95, indicating high internal consistency reliability (Table 3). The RESQ-eD showed good to excellent test-retest reliability for all domains (Table 3).

**Construct validity.** The RESQ-eD burping item correlated highly with both the GSRS burping item and Indigestion domain (Table 4). The RESQ-eD Heartburn, Regurgitation, and Overall symptoms domains showed high correlations

with the GSRS Reflux domain. The lowest correlations were observed between all RESQ-eD domains and the GSRS Diarrhoea and Constipation domains, thus supporting the discriminant validity of the RESQ-eD. Increasing GSRS scores were associated with increased symptom intensity in all RESQ-eD domains and in the RESQ-eD burping item, thus supporting the known-groups validity of the RESQ-eD (Figure 1).

**Responsiveness to change.** The RESQ-eD demonstrated good ability to capture change in mean intensity (Figure 2a) and proportion of symptom-free days (Figure 2b). However, no discrimination was observed between small and moderate improvements.



**Figure 1** Known-groups validity of the RESQ-eD in the PRO Validation Study, using the GRSR domains and single item “burping” as reference (all enrolled eligible patients,  $N = 580$ ). GRSR score (0–6) groups: low, if score  $\leq 4.0$ ; medium, if score  $> 4.0$  to  $\leq 5.0$ ; high, if score  $> 5.0$ . GRSR, Gastrointestinal Symptom Rating Scale; PRO, patient-reported outcome; RESQ-eD, Reflux Disease Questionnaire electronic Diary.

**Lesogaberan Dose-finding Study: confirmatory psychometric validation phase.** Frequency tabulation of the number of days with RESQ-eD symptoms by domain in the 7 days before randomization in the lesogaberan Dose-finding Study showed that 100% of randomized patients had Overall symptoms on 5–7 days during the week, 99% had Heartburn, 98% had Regurgitation, 88% had Hoarseness, cough, or difficulty swallowing, and 95% had Burping.

**Reliability.** High inter-item correlations were observed for all symptom domains of the RESQ-eD, thus confirming high internal consistency reliability observed in the PRO Validation Study (Table 3). Intraclass correlation coefficients confirmed the good to excellent test-retest reliability for all domains of the RESQ-eD (Table 3).

**Responsiveness to change.** The RESQ-eD demonstrated good ability to capture change in intensity in all domains in the lesogaberan Dose-finding Study, including an ability to discriminate between small and moderate improvements (Figure 3), thus confirming its responsiveness to change.

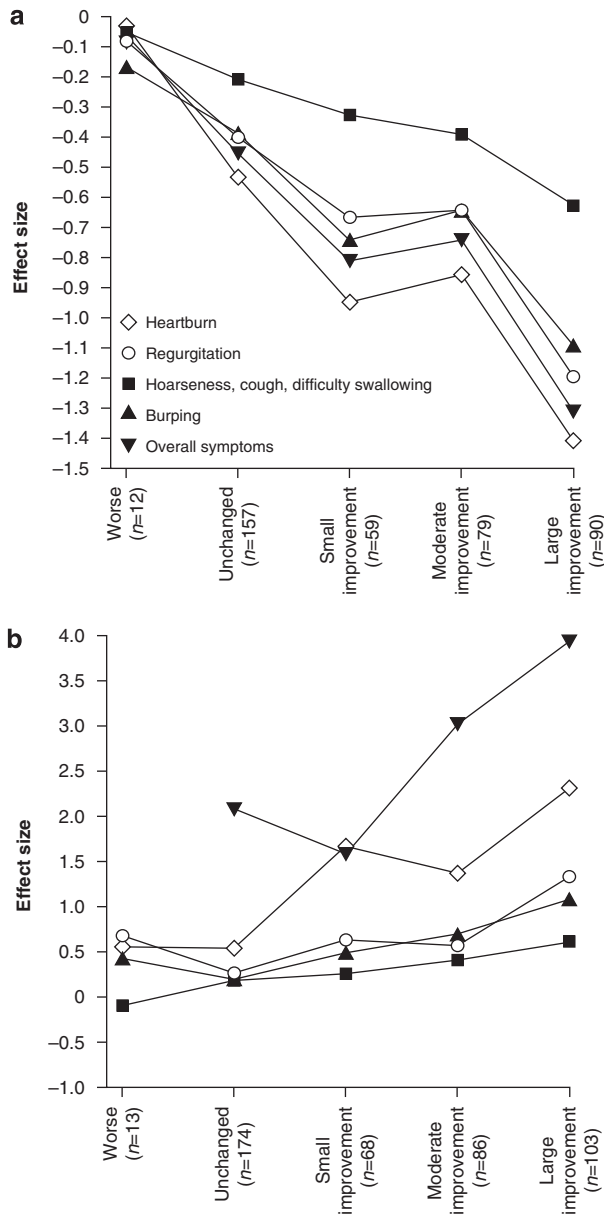
## DISCUSSION

The RESQ-eD was deemed to be valid, reliable, and responsive to change in the clinical trial setting in partial responders to PPI therapy. The 13 items of the RESQ-eD were shown to measure the symptoms most relevant to the target population, and can be meaningfully combined into an Overall symptom domain and four separate domains (Heartburn; Regurgitation; Hoarseness, cough, and difficulty swallowing; and Burping). Although a tool exists for use in clinical practice,<sup>24</sup> the RESQ-eD is, to our knowledge, the first electronic symptom diary designed for use in clinical trials in partial responders to PPI therapy, and has been developed and documented in line with the FDA guidance on the development and use of PRO instruments.<sup>12</sup> The RESQ-eD is available in the public domain and is free for use in the non-commercial setting, which is in keeping with GERD trial design consensus recommendations on better access to

PROs.<sup>25</sup> A version of the instrument, with the same items, but with a 7-day recall period (RESQ-7), is also available. Daily symptom recording is the preferred method of capturing patients’ symptom experience in the clinical trial setting and the RESQ-7 may be more appropriate for use in routine clinical care.

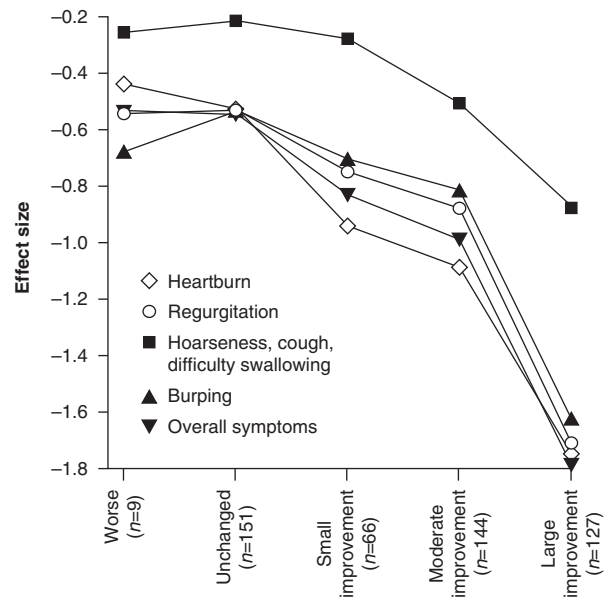
In accordance with the FDA guidance,<sup>12</sup> it was first determined through a systematic literature search that no adequate PRO instrument exists for use in clinical trials of patients on PPI therapy with persistent GERD symptoms. The RDQ was then used as a basis for developing the RESQ-eD, in line with the FDA guidance, which stipulates that if no adequate instrument exists, then a new instrument can be developed or modified from an existing instrument. The RDQ is one of the most extensively used instruments in GERD, and a paper-based version with a 7-day recall period has been psychometrically evaluated in patients with GERD for use in clinical trials and clinical practice.<sup>13–15</sup> However, the content validity of the RDQ has not been established, nor has it been evaluated in patient populations with a partial response to PPI, whose symptom pattern may differ from those with complete response or no response. The format (i.e., electronic diary) and recall period of the RESQ-eD also differ from those of the RDQ. The FDA guidance emphasizes the importance of establishing content validity (that the instrument captures everything relevant to measure what is intended) before other measurement properties are evaluated, as evidence of other types of validity will not overcome problems with content validity.

The content validity of the RESQ-eD was evaluated in the independent interview study and in the patient interviews in the PRO Validation Study. The interviews reinforced the relevance for the target patient population of the six RDQ items. However, the exploratory interviews indicated that relevant symptoms were missing. Therefore, seven more items were extracted from the interviews and from expert input, and included in the RESQ-eD. In the principal component analysis of the RESQ-eD, the two dyspepsia items in the original RDQ (i.e., pain and burning feeling in the center of the upper stomach) clustered with those for



**Figure 2** Responsiveness to change of the RESQ-eD in the PRO Validation Study (full analysis set,  $N = 478$ ). Effect size of the RESQ-eD domains is shown for (a) change in mean intensity and (b) change in proportion of symptom-free days from the 7 days before randomization to treatment period. Patients were grouped by OTE at the end of treatment. Note: the effect size for worse symptoms by OTE for the Overall symptoms domain could not be computed, as all patients had 0% of symptom-free days during the 7 days prior to first dose. OTE, Overall Treatment Evaluation; PRO, patient-reported outcome; RESQ-eD, Reflux Disease Questionnaire electronic Diary.

heartburn, and were thus included in the Heartburn domain. In another study evaluating upper gastrointestinal symptoms, epigastric pain and burning have shown what could be perceived as inconsistent associations, where these symptoms have clustered in the heartburn or the dyspeptic domains depending on whether patients were classified as heartburn predominant or non-heartburn predominant.<sup>15</sup> One explanation for this perceived inconsistency may be that various



**Figure 3** Confirmatory analysis of responsiveness to change of the RESQ-eD in the lesogaberan Dose-finding Study (all randomized patients,  $N = 661$ ). Patients were grouped by OTE at the end of treatment. OTE, Overall Treatment Evaluation; RESQ-eD, Reflux Disease Questionnaire electronic Diary.

underlying mechanisms, such as reflux or dysmotility, generate epigastric symptoms.<sup>26</sup>

When evaluating symptoms as the primary efficacy variable in a disease where symptoms are fluctuating, daily symptom recording is likely to be the most reliable method of capturing patients' symptom experience. The FDA guidance notes that items with short recall periods or items that ask patients to describe their current or recent state are preferable to those with longer recall periods; PRO instruments that call for patients to rely on memory are likely to undermine content validity.<sup>12</sup> Daily recording in an electronic diary may be superior to questionnaires when data are to be recorded contemporaneously, and has been shown previously to be well suited to capturing symptoms that fluctuate over time and to facilitate calculation of symptom-free days.<sup>27</sup> Electronic devices for data entry also help avoid prospective and retrospective diary entry and associated recall bias.

No general European regulatory recommendations currently exist for PRO instrument development and documentation. However, in its section on PRO instruments, the European Medicines Agency (EMA) draft guidelines on drug evaluation recommend that fully validated GERD-specific instruments should be used that focus on symptoms only when quantifying symptoms to assess the efficacy of drug candidates; assessment of health-related quality of life should be kept separate from symptom assessment.<sup>28</sup> The EMA recommendations also advise that symptom evaluation should include frequency and severity of symptoms, and that both heartburn and regurgitation should be included,<sup>28</sup> all of which apply to the RESQ-eD.

Several aspects of this study deserve comment. This is, to our knowledge, the first study that demonstrates the

feasibility of instrument development using new regulatory and research standards for symptom assessments in GERD. Our study provides the basis for the evaluation of new drugs in the treatment of patients with GERD who have an incomplete response to acid suppressive therapy. The instrument has been designed to meet regulatory expectations. However, for future use in clinical trials with the goal of establishing labeling language based on the RESQ-eD, it will be important to ensure that the content of the instrument matches any future claims. In addition, responder definitions for the instrument need to be explored in the context of the patient population and the intended claims.

As for a general GERD population, heartburn and regurgitation are the cardinal symptoms for partial responders. However, this validation study also demonstrates that there are other symptoms that need consideration in this patient population, including cough, hoarseness, and burping. These symptoms are associated with GERD syndromes but are not measured with most instruments used in untreated GERD. Studies have indicated that all or some of these additional symptoms may also be relevant for a general GERD population.<sup>29–32</sup> Therefore, it could be of interest for future studies to evaluate the validity (content and psychometric) of the RESQ-eD also in a treatment naïve GERD population including PPI responders.

In conclusion, the RESQ-eD demonstrated good content validity and psychometric properties in the clinical trial setting in patients with GERD who have a partial response to optimized PPI therapy, as characterized by persistent GERD symptoms. The RESQ-eD was validated in studies with lesogaberan; however, it could be applicable for use in the study of other developmental compounds that may be used as treatment for the same target patient population. Linguistic validation of the RESQ-eD in 10 languages and psychometric properties on pooled data from these languages supports the applicability in multinational and multilingual settings. The RESQ-eD is available in the public domain and is free for use in the non-commercial setting.

**Acknowledgements.** Dr Anja Becher, from Oxford PharmaGenesis™ Ltd, provided writing support funded by AstraZeneca R&D, Mölndal, Sweden.

## CONFLICT OF INTEREST

**Guarantor of the article:** Nimish Vakil, MD.

**Specific author contributions:** Nimish Vakil—study concept and design, acquisition of data, analysis and interpretation of data, and critical revision of the manuscript for important intellectual content. Karin Björck—study concept and design, analysis and interpretation of data, critical revision of the manuscript for important intellectual content, and statistical analysis. Hans Denison—study concept and design, analysis and interpretation of data, critical revision of the manuscript for important intellectual content, and study supervision. Katarina Halling—study concept and design, analysis and interpretation of data, and critical revision of the manuscript for important intellectual content. Maria Karlsson—study concept and design, analysis and interpretation of data, and critical

revision of the manuscript for important intellectual content. Jean Paty—study concept and design, analysis and interpretation of data, and critical revision of the manuscript for important intellectual content. Debra G. Silberg—study concept and design, analysis and interpretation of data, critical revision of the manuscript for important intellectual content, and study supervision. Anna Rydén—analysis and interpretation of data, drafting of the manuscript, and critical revision of the manuscript for important intellectual content. All authors approved the final draft submitted.

**Financial support:** This study was funded by AstraZeneca R&D, Mölndal, Sweden. Dr Anja Becher, from Oxford PharmaGenesis™ Ltd, provided writing support funded by AstraZeneca R&D, Mölndal, Sweden.

**Potential competing interests:** Nimish Vakil has received consultancy fees from Novartis Pharmaceuticals, AstraZeneca, Takeda Pharmaceutical, Meridian Bioscience, XenoPort, Orexo, Axcan Pharma and Ironwood Pharmaceuticals; has received grant/research support from AstraZeneca and XenoPort; and has ownership interest (e.g., stocks, stock options) in Orexo and Meridian Bioscience. Katarina Halling and Jean Paty have received consultancy fees from AstraZeneca R&D, Mölndal, Sweden. Karin Björck, Hans Denison, Maria Karlsson and Anna Rydén are employed by AstraZeneca R&D, Mölndal, Sweden. Debra Silberg was employed by AstraZeneca LP, Wilmington, USA at the time that the study was conducted.

## Study Highlights

### WHAT IS CURRENT KNOWLEDGE

- ✔ Some patients with GERD experience only a partial response of their heartburn or regurgitation symptoms to PPI therapy.
- ✔ The presence, frequency, and severity of GERD symptoms are best captured and measured using PRO instruments.
- ✔ The FDA recently published its guidance for industry on the use of PRO instruments in medical product development to support labelling claims.
- ✔ No PRO instrument has been developed for the specific target population of patients with GERD with a partial response to PPI therapy.

### WHAT IS NEW HERE

- ✔ The RESQ-eD was developed for use in clinical trials in patients with a partial response to PPI therapy, using methods that meet recent regulatory standards.
- ✔ The RESQ-eD was deemed to be valid, reliable, and responsive to change in the clinical trial setting in partial responders to PPI therapy.
- ✔ The RESQ-eD is, to our knowledge, the first electronic symptom diary designed for use in partial responders to PPI therapy that has been developed and documented in line with the FDA guidance on the development and use of PRO instruments.
- ✔ The RESQ-eD is available in the public domain and is free for use in the non-commercial setting.



1. van Pinxteren B, Numans ME, Bonis PA *et al.* Short-term treatment with proton pump inhibitors, H2-receptor antagonists and prokinetics for gastro-oesophageal reflux disease-like symptoms and endoscopy negative reflux disease. *Cochrane Database Syst Rev* 2004; 4: CD002095.
2. Donnellan C, Sharma N, Preston C *et al.* Medical treatments for the maintenance therapy of reflux oesophagitis and endoscopic negative reflux disease. *Cochrane Database Syst Rev* 2005; 4: CD003245.
3. 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: 720–737.
4. Fass R, Shapiro M, Dekel R *et al.* Systematic review: proton-pump inhibitor failure in gastro-oesophageal reflux disease—where next? *Aliment Pharmacol Ther* 2005; 22: 79–94.
5. Boeckxstaens GE, Smout A. Systematic review: role of acid, weakly acidic and weakly alkaline reflux in gastroesophageal reflux disease. *Aliment Pharmacol Ther* 2010; 32: 334–343.
6. Lehmann A. New pharmacological concepts in the treatment of gastro-oesophageal reflux disease. *Eur Gastroenterol Hepatol Rev* 2010; 6: 100–103.
7. Boeckxstaens GE, Beaumont H, Mertens V *et al.* Effects of lesogaberan on reflux and lower esophageal sphincter function in patients with gastroesophageal reflux disease. *Gastroenterology* 2010; 139: 409–417.
8. Boeckxstaens GE, Rydholm H, Lei A *et al.* Effect of lesogaberan, a novel GABA(B)-receptor agonist, on transient lower esophageal sphincter relaxations in male subjects. *Aliment Pharmacol Ther* 2010; 31: 1208–1217.
9. Weigt J, Monkemuller K, Peitz U *et al.* Multichannel intraluminal impedance and pH-metry for investigation of symptomatic gastroesophageal reflux disease. *Dig Dis* 2007; 25: 179–182.
10. Iwakiri K, Kawami N, Sano H *et al.* Acid and non-acid reflux in Japanese patients with non-erosive reflux disease with persistent reflux symptoms, despite taking a double-dose of proton pump inhibitor: a study using combined pH-impedance monitoring. *J Gastroenterol* 2009; 44: 708–712.
11. Vakil N, Veldhuyzen van Zanten S, Kahrilas P *et al.* The Montreal definition and classification of gastro-oesophageal reflux disease (GERD)—a global evidence-based consensus. *Am J Gastroenterol* 2006; 101: 1900–1920.
12. Food and Drug Administration##Patient-reported outcome measures: use in medical product development to support labeling claims. 2009 8 December 2009 [cited 2009 9 December]; Final guidance document]. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.
13. Shaw M, Dent J, Beebe T *et al.* The Reflux Disease Questionnaire: a measure for assessment of treatment response in clinical trials. *Health Qual Life Outcomes* 2008; 6: 31.
14. Shaw MJ, Talley NJ, Beebe TJ *et al.* Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. *Am J Gastroenterol* 2001; 96: 52–57.
15. Veldhuyzen van Zanten S, Armstrong D, Barkun A *et al.* Symptom overlap in patients with upper gastrointestinal complaints in the Canadian confirmatory acid suppression test (CAST) study: further psychometric validation of the Reflux Disease Questionnaire. *Aliment Pharmacol Ther* 2007; 25: 1087–1097.
16. Wild D, Grove A, Martin M *et al.* Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005; 8: 94–104.
17. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.
18. Juniper EF, Guyatt GH, Willan A *et al.* Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994; 47: 81–87.
19. Guyatt GH, Osoba D, Wu AW *et al.* Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; 77: 371–383.
20. Robinson M, Sahba B, Avner D *et al.* A comparison of lansoprazole and ranitidine in the treatment of erosive oesophagitis. Multicentre Investigational Group. *Aliment Pharmacol Ther* 1995; 9: 25–31.
21. Dimenäs E, Gilse H, Hallerback B *et al.* Well-being and gastrointestinal symptoms among patients referred to endoscopy owing to suspected duodenal ulcer. *Scand J Gastroenterol* 1995; 30: 1046–1052.
22. Fayers PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. Wiley: Chichester, 2000.
23. Fleiss JL. *The Design and Analysis of Clinical Experiments*. Wiley Classics Library ed. Wiley: New York, 1999.
24. Armstrong D, Veldhuyzen SJ, Chung SA *et al.* Validation of a short questionnaire in English and French for use in patients with persistent upper gastrointestinal symptoms despite proton pump inhibitor therapy: the PASS (Proton pump inhibitor Acid Suppression Symptom) test. *Can J Gastroenterol* 2005; 19: 350–358.
25. Dent J, Kahrilas PJ, Vakil N *et al.* Clinical trial design in adult reflux disease—a methodological workshop. *Aliment Pharmacol Ther* 2008; 28: 107–126.
26. Vakil N. Epigastric pain in dyspepsia and reflux disease. *Rev Gastroenterol Disord* 2003; 3 (Suppl 4): S16–S21.
27. McColl E. Best practice in symptom assessment: a review. *Gut* 2004; 53 (Suppl 4): iv49–iv54.
28. Bakowsky VS, Hanly JG. Complications of nonsteroidal antiinflammatory drug gastropathy and use of gastric cytoprotection: experience at a tertiary care health center. *J Rheumatol* 1999; 26: 1557–1563.
29. Smith JA, Decalmer S, Kelsall A *et al.* Acoustic cough-reflux associations in chronic cough: potential triggers and mechanisms. *Gastroenterology* 2010; 139: 754–762.
30. Liu JY, Woloshin S, Laycock WS *et al.* Symptoms and treatment burden of gastroesophageal reflux disease: validating the GERD assessment scales. *Arch Intern Med* 2004; 164: 2058–2064.
31. Thompson SK, Cai W, Jamieson GG *et al.* Recurrent symptoms after fundoplication with a negative pH study—recurrent reflux or functional heartburn? *J Gastrointest Surg* 2009; 13: 54–60.
32. Bredenoord AJ, Weusten BL, Timmer R *et al.* Air swallowing, belching, and reflux in patients with gastroesophageal reflux disease. *Am J Gastroenterol* 2006; 101: 1721–1726.



**Clinical and Translational Gastroenterology is an open-access journal published by Nature Publishing Group.** This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>