CrossMark

ORIGINAL SCIENTIFIC REPORT

Measuring Outcomes of Laparoscopic Anti-reflux Surgery: Quality of Life Versus Symptom Scores?

Jan H. Koetje[1,2] · Vincent B. Nieuwenhuijs[1] · Tanya Irvine[2] · George C. Mayne[2] · David I. Watson[2]

Abstract

Introduction Outcome following fundoplication for gastroesophageal reflux can be measured using objective tests, symptom scores and quality of life (QoL) measures. Which is best and how these assessments correlate is uncertain. To determine the utility of assessment measures we compared a general QoL measure (SF-36) and a disease-specific measure (GERD-hr-QoL) with symptom and satisfaction scores in individuals following fundoplication.

Methods 329 individuals underwent fundoplication between 2000 and 2015 in 2 centres in Australia and the Netherlands. Patients were assessed before and 3, 12 and 24 months after surgery using 10-point Likert scales to assess heartburn and satisfaction, the SF-36 questionnaire and the GERD-hr-QoL questionnaire. SF-36 scores were converted into component scores: Physical Component Scale (PCS) score and Mental Component Scale (MCS) score. Correlations between QoL measures and clinical outcomes were determined.

Results Surgery relieved heartburn (7.0 vs. 0.0 median, $P < 0.001$) and patients were highly satisfied with the outcome (median 9.0). PCS and MCS scores improved after surgery (PCS 40.9 vs. 46.0, $P < 0.001$; MCS 47.6 vs. 50.3, $P = 0.027$). GERD-hr-QoL scores also improved after surgery (15.7 vs. 3.7, $P < 0.001$). Correlations between PCS and MCS scores versus heartburn and satisfaction scores were generally weak or absent. However, correlations between GERD-hr-QoL versus heartburn and satisfaction scores were moderate to strong.

Conclusion Despite improvements in scores, the SF-36 correlated poorly with clinical outcome measures, and its use to measure outcome following fundoplication is questioned. However, the GERD-hr-QoL correlated well with the symptom scores, suggesting this disease-specific QoL measure is a better tool for assessing anti-reflux surgery outcome.

Introduction

Gastroesophageal reflux disease (GERD) is common in Western countries, with up to 40 % of individuals experiencing reflux symptoms at least monthly, and 7 % daily

[1]. A subset with more severe symptoms undergoes a laparoscopic fundoplication. Surgery offers effective reflux control, but can be followed by side effects, and reflux can recur in some individuals [2]. In an attempt to improve outcome, a range of fundoplication variants have been suggested, and many of these have been assessed within randomised trials.

When assessing outcomes, robust comparisons rely on reliable outcome measures. Outcome can be measured using objective investigations (endoscopy or pH monitoring), standardised clinical assessments and symptom scores, or quality of life (QoL) measures [3, 4]. Each

✉ David I. Watson
david.watson@flinders.edu.au

1 Department of Surgery, Isala, Zwolle, Overijssel, The Netherlands

2 Flinders University Department of Surgery, Flinders Medical Centre, Bedford Park, SA 5042, Australia

addresses outcome from a different perspective, and some outcome measures might be better than others. Unfortunately, it is difficult to get high levels of compliance with objective measures [3, 5]. Objective measures such as endoscopy and pH monitoring are invasive and when used for follow-up, compliance is generally poor, and repeated use of these measures is not practical. However, standardised symptom scores and QoL measures are more acceptable to patients and can be used repeatedly to assess outcome at different time points, and high levels of compliance are achievable [6, 7]. For more than 20 years, we have used standardised clinical assessment scores to measure outcome following laparoscopic fundoplication, and have found these outcomes correlate well with the findings from objective studies [2, 3, 5]. We have also used the Short Form-36 (SF-36), a widely used general QoL measure, and the GERD-hr-QoL, a GERD health-related QoL measure, to assess QoL in subsets of patients enrolled in specific studies [4, 8].

QoL is an important clinical research outcome [9, 10]. In 1948, the World Health Organization defined health as a state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity [11]. Hence, QoL differs from symptom assessment and includes physical, psychological and social dimensions. However, as it can be a highly subjective outcome, generic and disease-specific QoL questionnaires have been developed to quantitate QoL so that it can be better assessed in the research setting [12]. As gastroesophageal reflux can impact all three dimensions of QoL, assessing QoL as an outcome for anti-reflux surgery makes sense [13]. QoL can be assessed using general or disease-specific measures. Which method is best applied and how well QoL measures correlate with clinical symptoms is uncertain.

Comparisons of pre- versus post-operative QoL following laparoscopic fundoplication have confirmed improvement after surgery [9, 14, 15]. However, few studies have assessed how well QoL measures correlate with symptoms or global satisfaction measures after surgery, and the validity of the SF-36 in this context is also largely untested. Velanovich et al. previously compared the GERD-hr-QoL with the SF-36 in a small cohort of patients suffering GERD [16]. In their study, patients were treated with medication or surgery, and scores were compared at just one random time point. Correlations between the GERD-hr-QoL and SF-36 with other measures of heartburn and satisfaction pre- and post-operatively, as well as across multiple time points following anti-reflux surgery are not clear. Hence, to determine how well different QoL measures (SF-36 and GERD-hr-QoL) perform as outcome assessment measures following anti-reflux surgery, we compared QoL outcomes with a widely used symptom score and global patient satisfaction measure at a range of pre- and post-operative time points in a much larger cohort of patients.

Methods

Patients

Patients who underwent laparoscopic fundoplication for GERD between 2000 and 2015 in one of two cities in Australia (Adelaide) or the Netherlands (Isala, Zwolle) were identified and included in this study. Data were collected prospectively in both Centres and pooled for analysis.

Patients were included if they

- underwent laparoscopic fundoplication for objectively proven gastroesophageal reflux (pH study showing excessive acid exposure, or ulcerative esophagitis at endoscopy);
- had matched symptom scores and QoL scores (SF-36 and/or GERD-hr-QoL scores) collected at at least one of 3 time points—before surgery (baseline), 3 months after surgery and/or 12 months after surgery.

Patients less than 18 years old at surgery, and those whose first operation at our hospitals was a revision anti-reflux procedure were excluded.

Data from Australia were collected within two previously reported randomised controlled trials; laparoscopic anterior 90° partial versus Nissen fundoplication [17], and laparoscopic anterior 180° partial versus posterior partial versus Nissen fundoplication [18]. The protocol for the second trial was altered by dropping the Nissen fundoplication arm to improve recruitment, but the patients who underwent a Nissen fundoplication were still included in the current study. Data from the Netherlands were collected prospectively as part of a standardised follow-up protocol within the clinical practice of one of the authors (VBN).

Data collected and follow-up methods

Collected data included pre-operative and operative details, the SF-36 questionnaire, the GERD-hr-QoL questionnaire, and symptom scores for heartburn and satisfaction with the surgical outcome. The symptom scores comprised previously described 10-point Likert scales; heartburn—0 = no heartburn, 10 = severe heartburn; and satisfaction—0 = highly unsatisfied with the outcome of surgery, 10 = highly satisfied [6]. Questionnaires were completed before surgery (SF-36, GERD-hr-QoL and heartburn score), and at 3, 12 and 24 months (SF-36, GERD-hr-QoL, heartburn score and satisfaction score). Patients received and completed the questionnaires using a range of

methods; mail, face to face contact at hospital visits, telephone interview, or via e-mail using a secure online database managing system 'Research Manager'.

The surgical procedures were standardised within trial protocols and the procedures performed in the Netherlands were undertaken by a surgeon (VBN) who was trained in anti-reflux surgery in Adelaide, South Australia. A range of fundoplication variants were used, including Nissen, anterior 90° partial, anterior 180° partial and posterior partial fundoplication (Toupet). The technical details for each fundoplication type have been described elsewhere [3, 5, 17, 18].

Two QoL measures were used, SF-36 and the GERD-hr-QoL questionnaire. Version 1.0 of the SF-36 was used. This validated, widely used general QoL questionnaire consists of 36 items. Thirty five of these contribute to eight subscales (Table 1), and the other question stands alone and assesses the Reported Health Transition (RHT). The scores of the eight subscales and the RHT-question are converted into 0–100 scores. A higher score on each subscale indicates a better QoL for that subscale, and a score of 50 is equivalent to the population median. The subscales can also be converted into two summarising component scales: a Physical Component Scale (PCS) and a Mental Component Scale (MCS). The PCS and the MCS have been validated by Ware et al. [19]. The component scales provide a more concise overview of the SF-36 outcomes, and both the subscales and the component scales were calculated and used in this study. For these calculations, we used the population norm scores of the USA 1991, since the included patients were both Australian and Dutch. This approach is advised in the SF-36 manual.

GERD health-related QoL was assessed by using the GERD-hr-QoL questionnaire. This is a disease-specific, validated questionnaire that consists of nine questions. Each question can be scored from 0 to 5. This questionnaire has been used in previously published studies, and often in conjunction with the SF-36 [20, 21].

Statistical analysis

Analysis compared both SF-36-derived outcomes and GERD-hr-QoL outcomes versus clinical symptom outcomes. Statistical analysis was performed using IBM's Statistical Package for Social Sciences (SPSS), version 22 for Apple Macintosh OS (IBM corp., Armonk, New York, USA). Parametrically distributed continuous data are summarised as mean $\pm$ standard deviation (SD). Non-parametric continuous data are summarised as median with interquartile ranges (IQR). Categorical data are summarised as frequencies with percentages. The paired samples t tests were used to compare parametrically distributed paired continuous data. Non-parametrically distributed paired continuous datasets were compared using the Wilcoxon Signed-Rank test. The independent samples t test was used to compare unpaired parametrically distributed continuous data, and unpaired non-parametrically distributed continuous data were assessed using the Mann–Whitney U tests. Post hoc analyses were performed using the Holm–Bonferroni correction. The Chi-square test or Fisher's exact test were used to compare proportions of categorical data.

Correlations between non-parametrically distributed continuous datasets were determined using Spearman's Rank correlation test, and presented as Spearman rho correlation coefficient (r_s). A r_s less than 0.40 was considered to be weak, r_s of 0.40–0.59 to be a moderate correlation, $r_s > 0.60$ to be strong and $r_s > 0.80$ to be very strong. Correlations with $r_s > 0.35$ are considered to be of clinical importance, as for these correlations the association exceeds 10 %. A P value of <0.05 was considered statistically significant. However, to reduce the chance of a type 1 error due to a large number of tests, we only considered a P-value of <0.01 to be statistically significant for the correlation analyses. All calculated P-values were two-sided.

Table 1 Subscales included in the Short Form-36 (SF-36) questionnaire

Concept	Abbreviation	Summary
Physical functioning	PF	Limitations in physical activities because of health problems
Social functioning	SF	Limitations in social activities because of physical or emotional problems
Role functioning—physical	RP	Limitations in usual role activities (work or other daily activities) because of physical health problems
Bodily pain	BP	The effect on pain on work and the severity of pain
General mental health	MH	Psychological distress and well-being
Role functioning—emotional	RE	Limitations in usual role activities because of emotional problems
Vitality	V	Energy and fatigue
General health	GH	Perception of general health
Reported health transition	RHT	Current health compared to the situation 1 year before

Ethical approval

Data were collected with approval of the human research ethics committees of the participating hospitals. All patients provided written informed consent and agreed to the use of their data for research.

Results

329 patients met the study inclusion criteria. 276 had matched pre-operative data, 219 had matched data at 3 months, 141 at 12 months and 115 at 24 months after surgery, for paired analyses.

Baseline characteristics

Table 2 summarises the baseline characteristics of the patients included in the study. Age, height, weight and BMI were at the day of surgery. Mean age in the Australian cohort was older than in the Dutch cohort. Gender distribution was similar for both cohorts. The Dutch patients were significant taller than the Australian patients, and as body weight was similar, the BMI was lower in the Dutch cohort. The fundoplication types differed for the 2 cohorts, with the Dutch cohort mainly undergoing an anterior 180° partial fundoplication (80.2 %), and no Dutch patients undergoing a posterior partial or an anterior 90° partial fundoplication. Re-operation rates were similar for both

Table 2 Baseline characteristics

	Overall cohort (n = 329)	Australian cohort (n = 167)	Dutch cohort (n = 162)	P value
Age[a]	50.15 (13.31)	50.61 (12.54)	49.69 (14.07)	0.536[d]
Gender[b]				
Male	162 (49.2 %)	84 (50.3 %)	78 (48.1 %)	0.696[e]
Female	167 (50.8 %)	83 (49.7 %)	84 (51.9 %)	
Height[a]	1.71 (0.10)	1.68 (0.10)	1.74 (0.10)	<0.001[d]
Weight[a]	82.27 (15.55)	82.63 (13.46)	81.95 (17.19)	0.705[d]
BMI[a]	28.02 (4.84)	29.28 (4.31)	26.99 (5.02)	<0.001[d]
Type of fundoplication[b]				
Anterior 90° partial	58 (17.6 %)	58 (34.7 %)	0 (0 %)	<0.001[d]
Anterior 180° partial	154 (46.8 %)	24 (14.4 %)	130 (80.2 %)	
Posterior partial	39 (11.9 %)	23 (13.8 %)	16 (9.9 %)	
Nissen	78 (23.7 %)	62 (37.1 %)	16 (9.9 %)	
Re-operation[b]	24 (7.3 %)	13 (7.8 %)	11 (6.8 %)	0.880[e]
Heartburn[c]				
Pre-op	7.0 (4.0)	5.0 (5.0)	8.0 (2.0)	<0.001[f]
3 months	0.0 (2.0)*	0.0 (0.0)*	1.0 (2.0)*	<0.001[f]
12 months	0.0 (2.0)*	0.0 (2.0)*	1.0 (3.0)*	<0.001[f]
24 months	0.0 (3.0)*	0.0 (2.0)*	2.0 (2.0)*	<0.001[f]
Satisfaction[c]				
3 months	9.0 (3.0)	9.25 (2.0)	8.0 (3.0)	0.003[f]
12 months	9.0 (3.0)	9.25 (2.0)	8.0 (3.75)	0.039[f]
24 months	9.0 (3.0)	9.0 (3.0)	8.0 (3.75)	0.160[f]

BMI body mass index

* Significant (P < 0.001) difference for comparison of pre-operative versus post-operative scores using the Wilcoxon Signed-Rank test

Data presented as either [a]mean (standard deviation), [b]number (percentage), or [c]median (interquartile range)

P-values are for comparisons between the Australian and Dutch cohorts

[d] P-value using independent samples t test

[e] P-value using Chi-Square test

[f] P-value using Mann–Whitney-U test

cohorts. The Dutch patients reported a higher baseline heartburn score than the Australian patients (median 8.0 vs. 5.0), and also a higher score post-operative. Heartburn scores in the overall group, as well as individually in both cohorts, improved significantly following surgery ($P < 0.001$). Patients from both Centres were highly satisfied with the outcome of their surgery (median satisfaction score 9.0 at 24 months).

SF-36 physical component scale and mental component scale

Table 3 summarises the scores for the PCS and the MCS at the different time points. Compared to pre-operative scores there was a significant increase in PCS score at 3, 12 and 24 months ($P < 0.001$). The increase in the MCS score was also significant at 3 months ($P = 0.002$) and 24 months ($P = 0.027$) post-operative, but not at 12 months ($P = 0.575$). The MCS score was below the population norm (score of 50) before surgery, but just above the population norm at both at 3 and 24 months post-operative. However, the PCS score remained below the population norm score both before and after surgery.

GERD-hr-QoL

Table 4 summarises the scores for the GERD-hr-QoL at the different time points. Compared to pre-operative scores there was a significant improvement (decrease) in the GERD-hr-QoL score at 3, 12 and 24 months ($P < 0.001$).

Correlations between SF-36 and symptom scores

Correlations between the heartburn score and the satisfaction score versus the PCS and MCS scores are summarised

in Table 5. Before surgery, the majority of patients reported high heartburn scores, but a wide variation in PCS and MCS scores. Following surgery, the majority of patients reported low heartburn scores, and again a wide variation across the PCS and MCS scores. Using the satisfaction score, the majority of patients were highly satisfied with results of the surgery, but also with a wide variation seen in PCS and MCS scores. More dissatisfied patients (low satisfaction grade <3) also reported highly variable PCS and MCS scores. The Spearman rho correlation coefficient for the tested correlations did not exceed 0.314, indicating that any statistically significant correlation between the variables was either clinically weak or absent. For the PCS score statistically significant (but clinically weak) correlations were seen with both the heartburn and satisfaction

Table 5 Correlations between SF-36 component scores and clinical outcome scores

	PCS		MCS	
	r_s	P-value	r_s	P-value
0–10 heartburn score				
Pre-operative ($n = 262$)	−0.124	0.046	−0.106	0.088
3 months ($n = 200$)	−0.231	0.001	−0.067	0.347
12 months ($n = 131$)	0.032	0.716	0.087	0.324
24 months ($n = 110$)	−0.087	0.364	−0.054	0.573
0–10 satisfaction score				
3 months ($n = 211$)	0.314	<0.001	0.237	0.001
12 months ($n = 137$)	0.066	0.446	0.083	0.337
24 months ($n = 110$)	0.231	0.015	−0.025	0.795

P-values indicate statistical significance of Spearman Rank's correlation test outcomes for each correlation

r_s Spearman rho correlation coefficient

PCS Physical Component Scale, MCS Mental Component Scale

Table 3 Mean physical and mental component scores for the overall population

	Pre-operative ($n = 276$)	3 months ($n = 219$)	12 months ($n = 141$)	24 months ($n = 115$)
PCS	40.9 (9.8)	44.5 (10.9)**	44.7 (11.4)**	46.0 (11.5)**
MCS	47.6 (11.9)	50.3 (9.8)*	48.7 (11.2)	50.3 (10.3)*

Data presented as mean (SD). PCS = Physical Component Scale; MCS = Mental Component Scale

* $P < 0.05$ for pre-operative versus post-operative QoL score (paired samples t test)

** $P < 0.001$ for pre-operative versus post-operative QoL score (paired samples t test)

Table 4 Mean GERD-hr-QoL scores for the overall population

	Pre-operative ($n = 132$)	3 months ($n = 103$)	12 months ($n = 56$)	24 months ($n = 26$)
GERD-hr-QoL	15.7 (8.3)	3.2 (4.7)*	2.6 (4.9)*	3.7 (3.8)*

Data presented as mean (SD)

* $P < 0.001$ for pre-operative versus post-operative score (paired samples t test)

scores at 3 months post-operative. For the MCS score a statistically significant correlation was seen with the satisfaction score at 3 months, but this was also clinically weak.

Table 6 summarises correlations between the symptom scores and SF-36 subscales. Statistically significant correlations were seen between the satisfaction score and all SF-36 subscales except "Role Emotional" at 3 months follow-up. Statistically significant correlations were seen between the satisfaction score and the "General Health" and "Reported Health Transition" scores at 12 and none at 24 months follow-up. No correlations were identified with any other subscales. The correlations identified with the satisfaction score were all clinically weak ($r_s < 0.3$), except for the "Social Functioning", "General Health" and "Reported Health Transition" scores for which moderate correlations were identified. Statistically significant correlations were seen between the heartburn score and "Social Functioning", "Bodily Pain", "General Health" and "Reported Health Transition" at 3 months follow-up. All of these correlations were clinically weak ($r_s < 0.3$). No correlations with heartburn were identified for any subscale measured before surgery, at 12 months and at 24 months after surgery.

Correlations between GERD-hr-QoL and symptom scores

Table 7 summarises the correlations between the heartburn score and the satisfaction score versus the GERD-hr-QoL scores. Significant correlations were identified for all scores at all time points. Post-operative scores for heartburn correlated moderately or strongly with the GERD-hr-QoL and were likely to be clinically significant. There was a moderate correlation between Satisfaction at 3 and 24 months post-operative versus GERD-hr-QoL, and this

also exceeded the threshold for clinical significance. Pre-operative heartburn scores versus pre-operative GERD-hr-QoL scores, and satisfaction scores versus GERD-hr-QoL score at 12 months post-operative correlated weakly.

Discussion

As GERD is a chronic disease that impairs QoL, formal QoL measurement seems to be an appropriate method for assessment of surgical outcomes [10]. The aim of anti-reflux surgery is to fix each individual's symptoms and at the same time not add any new problem or side effect. An important outcome for surgery is control of reflux symptoms, and this can be quantitated using disease-specific symptom questionnaires [3]. However, focussing on reflux symptoms ignores post-fundoplication side effects which

Table 7 Correlations between GERD-hr-QoL scores and clinical outcome scores

	GERD-hr-QoL	
	r_s	P-value
0–10 heartburn score		
Pre-operative ($n = 121$)	0.360	<0.001
3 months ($n = 86$)	0.581	<0.001
12 months ($n = 49$)	0.464	0.001
24 months ($n = 25$)	0.745	<0.001
0–10 satisfaction score		
3 months ($n = 99$)	−0.539	<0.001
12 months ($n = 56$)	−0.359	0.007
24 months ($n = 26$)	−0.503	0.009

r_s Spearman rho correlation coefficient

P-values indicate statistical significance of Spearman Rank's correlation test outcomes for each correlation

Table 6 Correlations between SF-36 subscales and Clinical outcome scores

	PF	SF	RE	BP	MH	RP	V	GH	RHT
0–10 heartburn score									
Pre-operative ($n = 262$)	−0.109	−0.146	−0.124	−0.010	−0.068	−0.138	−0.119	−0.151	−0.107
3 months ($n = 200$)	−0.084	−0.234*	−0.036	−0.195*	−0.071	−0.159	−0.078	−0.226*	−0.246**
12 months ($n = 131$)	0.032	0.121	0.068	0.066	0.058	0.038	0.149	−0.129	−0.159
24 months ($n = 110$)	−0.111	−0.071	−0.008	−0.057	−0.110	0.012	−0.053	−0.176	−0.266
0–10 satisfaction score									
3 months ($n = 211$)	0.186*	0.323**	0.140	0.236*	0.209*	0.274**	0.280**	0.328**	0.562**
12 months ($n = 137$)	0.009	0.032	0.036	0.097	0.069	−0.043	0.038	0.264*	0.418**
24 months ($n = 110$)	0.178	0.108	−0.126	0.187	0.049	0.083	0.131	0.212	0.331

All data are Spearman rho correlation coefficients; * $P < 0.01$, ** $P < 0.001$

PF physical functioning, SF social functioning, RE role functioning—emotional, BP bodily pain, MH general mental health, RP role functioning—physical, V vitality, GH general health, RHT reported health transition

can also impact on the outcome of surgery. To address this, more general measures of satisfaction such as the Likert satisfaction scale or a general QoL assessment are needed. Previous studies have shown that patients suffering gastroesophageal reflux report a poorer QoL than a matched healthy population, with decreased scores for both the physical and mental components of the SF-36 questionnaire [10, 13]. This suggests that the SF-36 should be a valid method for outcome assessment after anti-reflux surgery, and other studies have shown improvements in SF-36 scores following laparoscopic anti-reflux surgery [6, 14, 15]. Other studies, however, have used disease-specific QoL assessment, with the GERD-hr-QoL frequently used to evaluate the outcome of laparoscopic anti-reflux surgery.

In our study, QoL, quantified by the PCS and MCS scores of the SF-36, improved significantly following laparoscopic fundoplication, as did QoL measured by the GERD-hr-QoL. Heartburn scores also fell significantly following surgery, and anti-reflux surgery was associated with a high rate of satisfaction at 3, 12 and 24 months postoperative. These outcomes are similar to those reported in previous randomised trials and outcome studies [8, 15, 22]. In our current study, the MCS scores 3 and 24 months after surgery were equivalent to the 'healthy population' norms, but the PCS scores remained below this norm score, mainly due to lower scores on the Vitality, General Health and Bodily Pain subscales. Whilst it is expected that these subscales will be impaired during early follow-up, it is to be expected that these scores should reach norm scores by 12–24 months after surgery. This outcome is somewhat at odds with the high level of heartburn control and high level of satisfaction with the outcome of surgery.

In general, clinically significant correlations between the SF-36 outcomes and the heartburn score were not identified. This suggests a poor relationship between general QoL measured by the SF-36, and reflux symptoms. More correlations between the SF-36 component scores and the subscales were seen with the satisfaction score. However, most of the correlations were still only clinically weak, and largely seen at 3 months, but not at 12 or 24 months follow-up.

Somewhat stronger correlations were seen at 3 months for Satisfaction versus "Reported Health Transition" ($r_s = 0.562$), "General Health" ($r_s = 0.328$), "Social Functioning" ($r_s = 0.323$) and the "Physical Composite Scale" ($r_s = 0.314$), and at 12 and 24 months for Satisfaction versus "Reported Health Transition". This might indicate some validity for the SF-36 scores as a global outcome measure, rather than as a disease-specific measure, but the magnitude of the correlations should be recognised to be modest at best.

In comparison, the GERD-hr-QoL correlated more strongly with the Heartburn scores, and also with the

Satisfaction scores at all time points. Although the pre-operative correlation between Heartburn score and GERD-hr-QoL was somewhat weaker, the post-operative correlations were certainly strong, and likely to be indicative of clinical symptom outcomes. This almost certainly reflects the more disease-specific questions used in the GERD-hr-QoL questionnaire, which focus on symptoms like heartburn, rather than more generic QoL issues.

A reasonable conclusion to draw from our study is that generic QoL measures such as the SF-36 are not specific enough to be used as robust outcome assessment tools following surgery for gastroesophagal reflux, whereas disease-specific measures such as the GERD-hr-QoL questionnaire appear better suited to assessment of outcome following anti-reflux surgery, and correlate better with validated symptom scores [12]. In general, the disease-specific measures such as the GERD-hr-QoL appear to grade the severity of outcomes such as Heartburn and other reflux symptoms, in a similar manner to the Likert Heartburn score used in the current study.

A difficulty with our current study is that there is no agreed "gold standard" for assessment of outcome after anti-reflux surgery. We have, however, reported extensively using the Likert scores applied in the current study, and have shown a good correlation between these scores and clinically meaningful outcomes, as well as with the outcomes from objective measures such as pH monitoring [3, 5]. It seems likely therefore, that the conclusion that the SF-36 performs poorly for the assessment of outcomes following surgery for reflux is valid, and a disease-specific QoL measure such as the GERD-hr-QoL would be a better choice for outcome assessment.

As the current study combined cohorts from 2 different countries, there are differences in the populations undergoing surgery, based on case selection and other factors. These differences account for the differences highlighted in Table 2. Obesity is a known risk factor for reflux, and the higher BMI in the Australian cohort might have adversely influenced the pre-operative heartburn scores, but the converse was seen. Differences in pre-operative heartburn scores might reflect earlier referral for surgery in Australia where fundoplication is a more common procedure than in the Netherlands. However, even though there were differences between the 2 cohorts included in the study, it is unlikely that these differences impacted on the study outcomes, as data were pooled and outcomes were not compared between countries.

In conclusion, whilst QoL and heartburn improved following anti-reflux surgery, and patients expressed high rates of satisfaction, the SF-36 questionnaire performed poorly as an outcome assessment tool in this context. This is probably because the SF-36 is a questionnaire which assesses generic QoL, and it is not specific enough to be

used for reliable measurement of outcome after laparoscopic anti-reflux surgery. In general, correlations between clinical outcome scores and SF-36 parameters were either poor or absent, suggesting that the changes in SF-36 scores did not consistently reflect disease-specific outcomes following surgery. In contrast, the disease-specific GERD-hrQoL questionnaire performed much better and, along with disease-specific clinical assessments, is recommended for assessment of outcome following surgery for gastroesophageal reflux.

References

1. Camilleri M, Dubois D, Coulie B, Jones M, Kahrilas PJ, Rentz AM et al (2005) Prevalence and socioeconomic impact of upper gastrointestinal disorders in the United States: results of the US Upper Gastrointestinal Study. Clin Gastroenterol Hepatol 3:543–552

2. Engström C, Cai W, Irvine T, Devitt PG, Thompson SK, Game PA, Bessell JR, Jamieson GG, Watson DI (2012) Twenty years of experience with laparoscopic antireflux surgery. Br J Surg 99:1415–1421

3. Watson DI, Pike GK, Baigrie RJ, Mathew G, Devitt PG, Britten-Jones R, Jamieson GG (1997) Prospective double blind randomized trial of laparoscopic Nissen fundoplication with division and without division of short gastric vessels. Ann Surg 226:642–652

4. Koetje JH, Irvine T, Thompson SK, Devitt PG, Woods SD, Aly A, Jamieson GG, Watson DI (2015) Quality of life following repair of large hiatal hernia is improved but not influenced by use of mesh: results from a randomized controlled trial. World J Surg 39:1465–1473. doi:10.1007/s00268-015-2970-3

5. Watson DI, Jamieson GG, Pike GK, Davies N, Richardson M, Devitt PG (1999) Prospective randomized double-blind trial between laparoscopic Nissen fundoplication and anterior partial fundoplication. Br J Surg 86:123–130

6. Yang H, Watson DI, Lally CJ, Devitt PG, Game PA, Jamieson GG (2008) Randomized trial of division versus non-division of the short gastric vessels during laparoscopic Nissen fundoplication—10 year outcomes. Ann Surg 247:38–42

7. Cai W, Watson DI, Lally CJ, Devitt PG, Game PA, Jamieson GG (2008) Ten-year clinical outcome of a prospective randomized clinical trial of laparoscopic Nissen versus anterior 180 partial fundoplication. Br J Surg 95:1501–1505

8. Woodcock SA, Watson DI, Lally C, Archer S, Bessell JR, Booth M et al (2006) Quality of life following laparoscopic anterior 90 degrees versus Nissen fundoplication: results from a multicenter randomized trial. World J Surg 30:1856–1863. doi:10.1007/s00268-005-0623-7

9. Kamolz T, Wykypiel H Jr, Bammer T, Pointner R (1998) Quality of life after laparoscopic antireflux surgery—Nissen fundoplication. Chirurg 69:947–950

10. Kamolz T, Pointner R, Velanovich V (2003) The impact of gastroesophageal reflux disease on quality of life. Surg Endosc 17:1193–1199

11. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19 June–22 July 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948

12. Velanovich V, Vallance SR, Gusz JR, Tapia FV, Harkabus MA (1996) Quality of life scale for gastroesophageal reflux disease. J Am Coll Surg 183:217–224

13. Revicki DA, Wood M, Maton PN, Sorensen S (1998) The impact of gastroesophageal reflux disease on health-related quality of life. Am J Med 104:252–258

14. Glise H, Hallerback B, Johansson B (1995) Quality-of-life assessments in evaluation of laparoscopic Rossetti fundoplication. Surg Endosc 9:183–188

15. Granderath FA, Kamolz T, Schweiger UM, Pointner R (2002) Quality of life, surgical outcome, and patient satisfaction three years after laparoscopic Nissen fundoplication. World J Surg 26:1234–1238. doi:10.1007/s00268-002-6416-3

16. Velanovich V (1998) Comparison of generic (SF-36) vs. disease-specific (GERD-HRQL) quality-of-life scales for gastroesophageal reflux disease. J Gastrointest Surg. 2:141–145

17. Watson DI, Jamieson GG, Lally C, Archer S, Bessell JR, Booth M, Cade R, Cullingford G, Devitt PG, Fletcher DR, Hurley J, Kiroff G, Martin C, Martin IJG, Nathanson LK, Windsor J (2004) Multicentre prospective double blind randomized trial of laparoscopic Nissen versus anterior 90 degree partial fundoplication. Arch Surg 139:1160–1167

18. Daud WMBW, Thompson SK, Jamieson GG, Devitt PG, Martin IJG, Watson DI (2014) Randomized controlled trial of laparoscopic anterior 180 partial vs. posterior 270 partial fundoplication. ANZ J Surg. doi:10.1111/ans.12476

19. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A (1995) Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care 33(4 Suppl):AS264–AS279

20. Tosato F, Marano S, Mattacchine S, Luongo B, Mingarelli V, Campagna G (2012) Quality of life after Nissen-Rossetti fundoplication. Surg Laparosc Endosc Percutan Tech 22:205–209

21. Rossetti G, Limongelli P, Cimmino M, Napoletano D, Bondanese MC, Romano G, Pratilas M, Guerriero L, Orlando F, Conzo G, Amato B, Docimo G, Tolone S, Brusciano L, Docimo L, Fei L (2014) Outcome of medical and surgical therapy of GERD: predictive role of quality of life scores and instrumental evaluation. Int J Surg 12:112–116

22. Korolija D, Sauerland S, Wood-Dauphinee S, Abbou CC, Eypasch E, Caballero MG et al (2004) Evaluation of quality of life after laparoscopic surgery: evidence-based guidelines of the European Association for Endoscopic Surgery. Surg Endosc 18:879–897