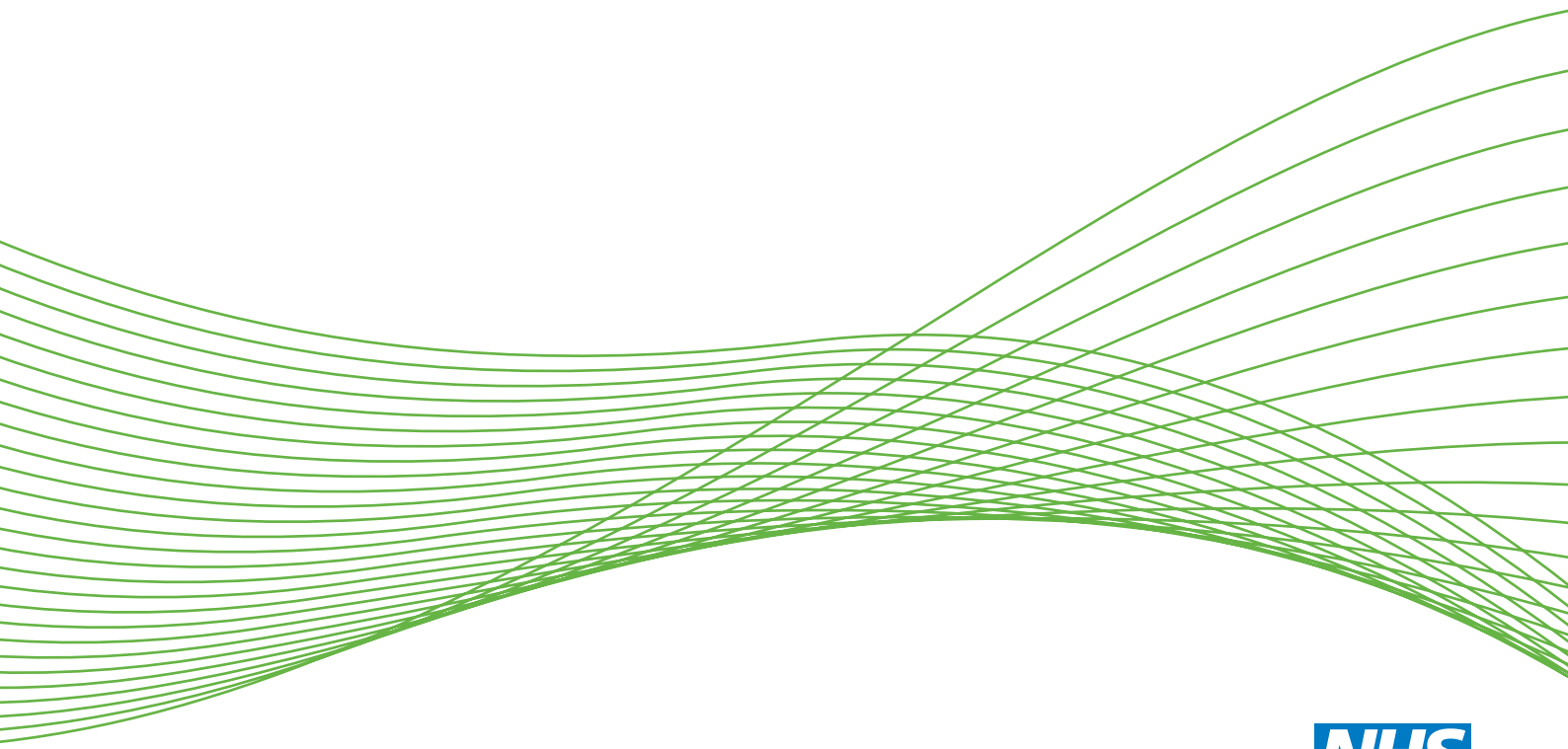
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 17 ISSUE 22 JUNE 2013
ISSN 1366-5278

Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial)

AM Grant, C Boachie, SC Cotton, R Faria, L Bojke, DM Epstein, CR Ramsay, B Corbacho, M Sculpher, ZH Krukowski, RC Heading and MK Campbell on behalf of the REFLUX trial group



National Institute for Health Research

DOI 10.3310/hta17220

Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial)

AM Grant,[1]* C Boachie,[1] SC Cotton,[1] R Faria,[2] L Bojke,[2] DM Epstein,[2] CR Ramsay,[1] B Corbacho,[2] M Sculpher,[2] ZH Krukowski,[3] RC Heading[4] and MK Campbell[1] on behalf of the REFLUX trial group

[1]Health Services Research Unit, Health Sciences Building, University of Aberdeen, Aberdeen, UK
[2]Centre for Health Economics, University of York, York, UK
[3]Aberdeen Royal Infirmary, Aberdeen, UK
[4]School of Medicine and Health, Wolfson Research Institute, Durham University, Stockton-on-Tees, UK

*Corresponding author

Declared competing interests of authors: Adrian Grant has received salary support from the NIHR as director of the NIHR Programme Grants for Applied Research programme.

Published June 2013
DOI: 10.3310/hta17220

This report should be referenced as follows:

Grant AM, Boachie C, Cotton SC, Faria R, Bojke L, Epstein DM, *et al*. Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial). *Health Technol Assess* 2013;**17**(22).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (*SciSearch*®) and *Current Contents*®/Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.596

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 97/10/99. The contractual start date was in May 2007. The draft report began editorial review in October 2011 and was accepted for publication in July 2012. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial)

AM Grant,[1]* C Boachie,[1] SC Cotton,[1] R Faria,[2] L Bojke,[2] DM Epstein,[2] CR Ramsay,[1] B Corbacho,[2] M Sculpher,[2] ZH Krukowski,[3] RC Heading[4] and MK Campbell[1] on behalf of the REFLUX trial group

[1]Health Services Research Unit, Health Sciences Building, University of Aberdeen, Aberdeen, UK
[2]Centre for Health Economics, University of York, York, UK
[3]Aberdeen Royal Infirmary, Aberdeen, UK
[4]School of Medicine and Health, Wolfson Research Institute, Durham University, Stockton-on-Tees, UK

*Corresponding author a.grant@abdn.ac.uk

Background: Despite promising evidence that laparoscopic fundoplication provides better short-term relief of gastro-oesophageal reflux disease (GORD) than continued medical management, uncertainty remains about whether benefits are sustained and outweigh risks.

Objective: To evaluate the long-term clinical effectiveness, cost-effectiveness and safety of laparoscopic surgery among people with GORD requiring long-term medication and suitable for both surgical and medical management.

Design: Five-year follow-up of a randomised trial (with parallel non-randomised preference groups) comparing a laparoscopic surgery-based policy with a continued medical management policy. Cost-effectiveness was assessed alongside the trial using a NHS perspective for costs and expressing health outcomes in terms of quality-adjusted life-years (QALYs).

Setting: Follow-up was by annual postal questionnaire and selective hospital case notes review; initial recruitment in 21 UK hospitals.

Participants: Questionnaire responders among the 810 original participants. At entry, all had documented evidence of GORD and symptoms for >12 months. Questionnaire response rates (years 1–5) were from 89.5% to 68.9%.

Interventions: Three hundred and fifty-seven participants were recruited to the randomised comparison (178 randomised to surgical management and 179 randomised to continued medical management) and 453 to the preference groups (261 surgical management and 192 medical management). The surgeon chose the type of fundoplication.

Main outcome measures: Primary: disease-specific outcome measure (the REFLUX questionnaire); secondary: Short Form questionnaire-36 items (SF-36), European Quality of Life-5 Dimensions (EQ-5D), NHS resource use, reflux medication, complications.

Results: The randomised groups were well balanced. By 5 years, 63% in the randomised surgical group and 13% in the randomised medical management group had received a total or partial wrap fundoplication (85% and 3% in the preference groups), with few perioperative complications and no associated deaths. At 1 year (and 5 years) after surgery, 36% (41%) in the randomised surgical group – 15% (26%) of those who had surgery – were taking proton pump inhibitor medication compared with 87% (82%) in the randomised medical group. At each year, differences in the REFLUX score significantly favoured the randomised surgical group (a third of a SD; $p<0.01$ at 5 years). SF-36 and EQ-5D scores also favoured surgery, but differences attenuated over time and were generally not statistically significant at 5 years. The worse the symptoms at trial entry, the larger the benefit observed after surgery. Those randomised to medical management who subsequently had surgery had low baseline scores that markedly improved after surgery. Following fundoplication, 3% had surgical treatment for a complication and 4% had subsequent reflux-related operations – most often revision of the wrap. Dysphagia, flatulence and inability to vomit were similar in the two randomised groups. The economic analysis indicated that surgery was the more cost-effective option for this patient group. The incremental cost-effectiveness ratio for surgery in the base case was £7028 per additional QALY; these findings were robust to changes in approaches and assumptions. The probability of surgery being cost-effective at a threshold of £20,000 per additional QALY was >0.80 for all analyses.

Conclusions: After 5 years, laparoscopic fundoplication continues to provide better relief of GORD symptoms with associated improved health-related quality of life. Complications of surgery were uncommon. Despite being initially more costly, a surgical policy is highly likely to be cost-effective.

Trial registration: Current Controlled Trials ISRCTN15517081.

Funding: This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 17, No. 22. See the HTA programme website for further project information.

Contents

List of abbreviations

BMI	body mass index
CDSR	Cochrane Database of Systematic Reviews
CI	confidence interval
CONSORT	Consolidated Standards of Reporting Trials
DARE	Database of Abstracts of Reviews of Effects
DMC	Data Monitoring Committee
EQ-5D	European Quality of Life-5 Dimensions
EVPI	expected value of perfect information
GERSS	Gastro-Esophageal Reflux Symptom Score
GORD	gastro-oesophageal reflux disease
GP	general practitioner
GSRS	Gastrointestinal Symptoms Rating Scale
H_2RA	histamine receptor antagonist
HRQoL	health-related quality of life
HTA	Health Technology Assessment
HUI3	Health Utilities Index Mark 3
ICER	incremental cost-effectiveness ratio
ITT	intention to treat
LOTUS	LOng-Term Usage of esomeprazole versus Surgery for treatment of chronic GERD
MAR	missing at random
MCAR	missing completely at random
MCS	mental component score
MICE	multiple imputation using chained equations
MNAR	missing not at random
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMB	net monetary benefit
OLS	ordinary least squares
PCS	physical component score
PGWI	Psychological General Well-Being Index
PP	per protocol
PPI	proton pump inhibitor
QALY	quality-adjusted life-year
QoL	quality of life
QOLRAD	Quality of Life in Reflux and Dyspepsia
RCT	randomised controlled trial
SD	standard deviation
SF-36	Short Form questionnaire-36 items
SF-6D	Short Form questionnaire-6 dimensions

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

In the Health Technology Assessment (HTA)-commissioned REFLUX trial, laparoscopic fundoplication for people with chronic symptoms of gastro-oesophageal reflux disease (GORD) was shown to significantly improve reflux-specific and general health-related quality of life (HRQoL) at least up to 12 months after surgery. However, cost-effectiveness was uncertain without more reliable information about longer-term costs and benefits. Here, we report the findings from longer-term follow-up of the REFLUX trial.

Objective

To evaluate, at 5 years after surgery, the clinical effectiveness, cost-effectiveness and safety of a policy of relatively early laparoscopic surgery compared with continued medical management among people with GORD symptoms that are reasonably controlled by medication and who are judged suitable for both surgical and medical management.

Methods

Design

1. Long-term follow-up of a pragmatic randomised controlled trial (with parallel non-randomised preference groups) comparing a laparoscopic surgery-based policy with a continued medical management policy to assess relative clinical effectiveness.
2. An economic evaluation of laparoscopic surgery for GORD to compare the cost-effectiveness of the two management policies, based on a within-trial (5-year) economic analysis and exploration of the need for a longer-term model.

Setting

Participants had originally been recruited in 21 UK hospitals through local partnership between surgeon(s) and gastroenterologist(s) who shared the secondary care of patients with GORD. After operation (surgical groups) and after optimisation of anti-reflux therapy (medical groups), participants were returned to the care of their general practitioners (GPs). Follow-up was by annual postal questionnaire and selective case notes review when questionnaires indicated reflux-related health-care events.

Participants

Participants in this study were questionnaire responders among the 810 original participants. At trial entry, all had both documented evidence of GORD and symptoms for >12 months. Annual questionnaire response rates (years 1–5) were 89.5%, 77.5%, 76.7%, 69.8% and 68.9%.

Intervention

Of the 810 participants, 357 were recruited to the randomised comparison (178 randomised to surgical management and 179 randomised to continued medical management) and 453 to the parallel non-randomised preference arm (261 surgical management and 192 medical management). The type of fundoplication was left to the discretion of the surgeon.

Main outcome measures

The principal outcome measure was a disease-specific instrument (the REFLUX questionnaire developed specifically for this study). Secondary measures were the Short Form questionnaire-36 items (SF-36), the

European Quality of Life-5 Dimensions (EQ-5D), surgical events including complications, reflux medication use, GP visits, hospital outpatient consultations, day and overnight hospital admissions, and their costs.

Results

At entry to the original trial, participants had been taking GORD medication for a median of 32 months and had a mean age of 46 years, and 66% were men; the randomised groups had been well balanced. Responders at 5 years were older, had been on medication for a shorter time prior to trial entry and had higher baseline quality-of-life scores than non-responders; however, the randomised groups of responders were similar in baseline characteristics. Primary analyses were based on the 'intention-to-treat' (ITT) principle, with secondary per-protocol analyses based on those who, at 1 year, had received their allocated treatment.

By 5 years, 63% ($n = 112$) of the 178 randomised surgery participants and 13% ($n = 24$) of the 179 randomised medical management participants had actually received fundoplication (equivalent figures in the preference groups were 85% and 3%). There had been a mixture of clinical and personal reasons for those allocated surgery not receiving it, sometimes related to long waiting times. A total or partial wrap procedure had been performed depending on surgeon preference; perioperative complications had been uncommon with no deaths associated with surgery.

By the equivalent to 12 months after surgery, 36% in the randomised surgical group (15% among those who had surgery) were taking proton pump inhibitor medication compared with 87% in the randomised medical group. At 5 years, the equivalent figures were 41% (26%) in those randomised to surgery and 82% in those randomised to medical management.

At each year, there were significant differences in the REFLUX score (a third of a SD; $p < 0.01$ at 5 years) favouring the randomised surgical group, reflecting differences in general discomfort (particularly), wind and frequency, nausea and vomiting, and activity limitation subscores. SF-36 and EQ-5D scores also favoured the randomised surgical group, especially SF-36 norm-based general health, but differences attenuated over time and were generally not statistically significant at 5 years [EQ-5D difference (ITT) 0.047, 95% confidence interval (CI) –0.013 to 0.108; $p = 0.13$]. The lower the REFLUX score and hence the worse the symptoms at trial entry, the larger the benefit observed after surgery. Post hoc exploratory analyses showed that those randomly allocated to medical management who subsequently had surgery had worse symptoms (lower baseline scores) than those who continued on medical management as allocated; following surgery, the scores of these patients markedly improved and this explains, at least in part, why differences in outcome between the randomised groups became less marked over time.

The preference surgical group also had low REFLUX scores at baseline. These scores improved substantially after surgery and at 5 years they were slightly better than those in the preference medical group.

Overall, 4% ($n = 16$) of the total 364 in the study who had fundoplication had a subsequent reflux-related operation, of whom two had a further (i.e. third) operation. Reoperation was most often conversion to a different type of wrap or a reconstruction of the same wrap. There were only two cases of reversal of the fundoplication and neither was in the randomised comparison. In total, 3% ($n = 12$) of those who had fundoplication required surgical treatment for a complication directly related to the original surgery, including oesophageal dilatation ($n = 4$) and repair of incisional hernia ($n = 3$). Patterns of 'difficulty swallowing', flatulence and 'wanting to vomit but being physically unable to do so' – all problems that have previously been associated with anti-reflux surgery – were similar in the two randomised groups.

Economic evaluation

Differences in mean costs and mean quality-adjusted life-years (QALYs) at 5 years were used to derive an estimate of the cost-effectiveness of laparoscopic fundoplication and continued medical management from the perspective of the NHS. Conventional decision rules were used to estimate incremental cost-effectiveness ratios (ICERs). Sensitivity analysis (including probabilistic sensitivity analysis) was used to explore and quantify uncertainty in the cost-effectiveness results.

Health-care resource-use data were collected prospectively as part of the clinical report forms and patient questionnaires at each follow-up point. The cost for each individual patient in the trial was calculated by multiplying their use of NHS resources by the associated unit costs (from published sources) and discounting at an annual rate of 3.5%. For the base-case analysis, total costs constituted the costs of surgery, complications due to surgery, reoperations, reflux-related prescribed medication, reflux-related visits to and from the GP and reflux-related hospital inpatient, outpatient and day visits. For the sensitivity analysis, all GP visits and all hospital admissions were included in the calculation of total costs. Health outcomes were expressed in terms of QALYs. HRQoL was assessed at each follow-up point using the EQ-5D. Incremental mean QALYs between randomised treatment groups were estimated with and without adjustment for baseline utility, using ordinary least squares regression.

The extent of missing data throughout the trial follow-up was significant; for this reason, the base case drew on the multiple imputed data set ITT analysis. A separate scenario – the complete-case analysis, in which only participants who returned all questionnaires and completed all EQ-5D profiles are included – was employed for both ITT and per-protocol analyses. Multiple imputation provides unbiased estimates of treatment effect if data are missing at random. Sensitivity analysis was used to test the impact on the cost-effectiveness results if data were missing not at random, that is, if patients with worse outcomes or greater costs were more likely to have missing data.

The results show that, for the base-case analysis (multiple imputed data set), the participants randomised to fundoplication accrued greater costs (incremental mean cost £1518; 95% CI £1006 to £2029) but also reported greater overall HRQoL (incremental mean QALYs 0.2160; 95% CI 0.0205 to 0.4115) than participants randomised to continued medical management. Laparoscopic fundoplication is a cost-effective strategy for GORD patients eligible for the REFLUX trial on the basis of the range of cost-effectiveness thresholds used by the National Institute for Health and Care Excellence (NICE) (£20,000–30,000 per additional QALY). The results for the complete-case analysis concurred with the multiple imputed data set: across analyses adjusted and unadjusted for baseline EQ-5D, ICERs ranged between £5468 and £8410, well below the NICE cost-effectiveness thresholds. For both data sets (multiple imputation and complete case), the probability of surgery being the more cost-effective intervention was >0.82 for incremental analyses unadjusted for baseline EQ-5D and >0.93 once incremental QALYs were adjusted for baseline EQ-5D.

A sensitivity analysis was carried out comparing the groups according to their 'per-protocol' status at 1 year. A per-protocol analysis compares the efficacy of the treatments received, whereas an ITT analysis compares the effectiveness of the strategies as offered to patients. The per-protocol analysis (in complete cases) suggested that surgery was more cost-effective than medical management. Other sensitivity analyses were carried out using a wider set of resource-use data. The results of the first alternative scenario, using the costs of primary care visits for any reason rather than only reflux-related reasons, increased the ICER slightly in relation to the base case. Nevertheless, the ICER remains well below conventional thresholds, and the probability of surgery being cost-effective was >0.85 for both adjusted and unadjusted analyses. In the second alternative scenario, replacing reflux-related hospital costs by all hospital costs, medical management was 'dominated' by the surgical policy; the probability of surgery being cost-effective was >0.90.

The base-case analysis imputes missing data. This assumes that missing data are missing at random, that is, their values can be predicted (with uncertainty) from observed data. This assumption is impossible to confirm or repute but its effect on the results can be tested in sensitivity analysis. The base-case analysis may be biased if the values of a missing variable are different from the observed values (for given values of other covariates). Sensitivity analysis using the multiple imputation data set showed that the cost-effectiveness of surgery was relatively insensitive to any increase in costs: cost-effectiveness changed little when costs were increased for patients with missing data in both treatment groups and when costs were increased just for patients randomly allocated surgery with missing data. A similar result was observed after reducing the total QALYs for all patients with missing data. In contrast, the cost-effectiveness of surgery was highly sensitive to the assumption that patients randomly allocated surgery with missing data experience lower HRQoL than patients with complete data. A 10% decrease in QALYs for patients randomised to surgery with missing data results in the cost-effectiveness increasing above £20,000 per QALY gained. This scenario shows that missing data can have an impact on the results. Nevertheless, although it is impossible to empirically confirm or refute this scenario from the data in the trial, it would seem improbable in practice that surgical patients with poor quality of life are less likely to respond to follow-up questionnaires than similar participants undergoing medical management.

Comparison with similar randomised trials

The findings of the REFLUX trial were considered in the context of the three other randomised trials that have compared laparoscopic surgery with medical management. In respect of benefits, the trials consistently show better relief of GORD symptoms following surgery, with parallel, though less marked, improvements in generic HRQoL. The four trials are also consistent in respect of complications of surgery, with small numbers having associated visceral injuries, postoperative problems and dilatation of the fundoplication wrap. The REFLUX trial suggests that 4.5% have a reoperation and the other trials are broadly consistent with this. Difficulty swallowing (dysphagia), flatulence and bloating have been linked with fundoplication in the other trials. In contrast, although a small number of REFLUX participants had a dilatation of the fundoplication wrap, responses to the questionnaires did not show a difference between those randomised to surgery and those randomised to medical management in these respects.

Conclusions

After 5 years' follow-up, a policy of relatively early laparoscopic fundoplication among patients for whom reasonable control of GORD symptoms requires long-term medication and for whom both surgery and medical management are suitable continues to provide better relief of GORD symptoms with associated better quality of life. Complications of surgery were rare. Despite being initially more costly, a surgical policy is likely to be more cost-effective for such patients suffering from GORD who were eligible for the REFLUX trial.

Implications for health care

Extending the use of laparoscopic fundoplication to people whose GORD symptoms require long-term medication for reasonable control and who would be suitable for surgery would provide health gains that extend over a number of years. The longer-term data reported here indicate that this would also be a cost-effective use of resources. The more troublesome the symptoms, the greater the potential benefit from surgery.

Recommendations for research

Most patients taking anti-reflux medication are managed in general practice. It is uncertain how many of these people might be suitable for surgery and hence what the most efficient provision of future care might be. Further research to explore the feasibility and resource impact of alternative policies for fundoplication within the NHS is therefore recommended.

Trial registration

This study is registered as ISRCTN15517081.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Introduction

This report describes the long-term follow-up of the REFLUX trial assessing the clinical effectiveness and cost-effectiveness of laparoscopic surgery compared with continued medical management for people with gastro-oesophageal reflux disease (GORD). This comparison was identified as a priority for research by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme, which funded the trial in two stages. The first stage, encompassing preliminary economic modelling, outcome development, trial recruitment, initial clinical management, follow-up to a time equivalent to 1 year after surgery and modelling of cost-effectiveness based on results available at that time, was reported in 2008.[1–5] The second stage, reported here, describes analyses based on further follow-up to 5 years after surgery.

Gastro-oesophageal reflux disease

The lower oesophagus, at its junction with the stomach, normally acts as a sphincter to prevent the contents of the stomach flowing back up the oesophagus. When the sphincter does not work adequately, the acid stomach contents leak, or 'reflux', into the oesophagus. The commonest symptom that this causes is heartburn, a burning sensation in the chest or throat. GORD has been defined through an international consensus process as 'a condition which develops when the reflux of stomach contents causes troublesome symptoms and/or complications'; in this consensus, symptoms were considered 'troublesome' 'if they adversely affected a patient's well-being'.[6]

Symptoms caused by gastro-oesophageal reflux are common: between 20% and 30% of a 'Western' adult population experience heartburn and/or reflux intermittently.[7–9]

Treatment of GORD includes both medical and surgical management, the options depending on the severity of symptoms. The majority of people with reflux have only mild symptoms and require little, if any, medication. The simplest is self-administered antacids with advice to alter lifestyle factors such as dietary modification, smoking cessation and weight reduction. A minority have severe symptoms and develop overt complications, despite full medical therapy, and require surgical intervention. Among the remainder, control of symptoms requires regular or continuous acid-suppression therapy using either histamine receptor antagonists (H_2RAs) or proton pump inhibitors (PPIs); initial high-dose therapy may be followed by maintenance treatment using these drugs either intermittently or continuously at a reduced dose sufficient to suppress symptoms. It is from this intermediate group of patients with significant disease requiring maintenance medical treatment that most of the treatment costs for the health service arise.

Laparoscopic fundoplication

Interest in surgery as an alternative to long-term medical therapy for GORD has been considerable since the introduction of the minimal access laparoscopic approach in the early 1990s.[10] Randomised trials conducted comparing laparoscopic with open surgery showed similar improvement in symptoms but with clear benefits of the laparoscopic approach in terms of recovery and fewer postsurgical complications.[11] As a consequence, surgery was suggested as an alternative to long-term maintenance medical treatment with anti-reflux drugs.

The operative method, whether using an open or a laparoscopic approach, involves performing a fundoplication by wrapping the fundus of the stomach around the lower oesophagus to create a high-pressure zone, thus reducing gastro-oesophageal reflux. The wrap created can be either complete (360°)

or partial. Many operative variants have been described. The commonest operation is a 1-cm complete wrap fashioned over a large bougie, the so-called 'short-floppy Nissen'.[12,13] There has been debate about the use of a partial rather than a total fundoplication. The partial approach has a number of potential advantages (such as fewer postoperative complications) but several controlled studies have shown broad equivalence between the two approaches;[14] for the purpose of this study they were therefore regarded as equivalent. Although fundoplication is reported to produce resolution of reflux symptoms in upwards of 90% of patients, like all surgery it carries risks and can have side effects. There is also uncertainty about the durability of benefit and frequency and severity of side effects following surgical therapy. Long-term follow-up to 12 years after open reflux surgery suggested attenuated but continuing better control of reflux symptoms; however, other symptoms such as difficulties swallowing (dysphagia), rectal flatulence and inability to belch or vomit were more common in surgical patients.[15] An important objective of this study was to determine if the long-term pattern of symptoms following laparoscopic surgery was similar.

Medical management

Proton pump inhibitors, sometimes supplemented with prokinetics or alginates, are the most effective medical treatment for moderate to severe GORD. Once started on PPIs, the majority of patients with significant GORD remain on long-term treatment.[16] It is estimated that around 1% or more of the UK adult population are prescribed PPI maintenance therapy.[17–19] The cost to the NHS of medical management of GORD is considerable. In England alone, the cost of PPIs is estimated to be £220M per year.[20] Of this budget, most of this prescribing occurs within the primary care setting.[21,22]

Although PPIs are generally considered safe, there is increasing acknowledgement of their possible adverse effects.[23,24] Gastric acid suppression predisposes to enteric infections and the sustained hypergastrinaemia resulting from PPI use causes rebound acid hypersecretion and the development of acid-related symptoms if the drug is stopped. Acute severe hypomagnesaemia has been recognised relatively recently as a rare adverse reaction to PPIs; the mechanism underlying it is not known. The clinical significance of impaired vitamin B_{12} and iron absorption due to PPIs is uncertain; there is also controversy about the risk of fractures and pneumonia and about the occurrence and significance of gastric mucosal atrophy and intestinal metaplasia, which have been seen in *Helicobacter pylori*-positive patients taking PPIs. Drug–drug interactions have also been a cause for concern,[25] although unequivocal evidence of their occurrence does not in itself establish clinical significance.

For the purpose of this study, medical therapy was taken to mean long-term therapy with PPIs (or H_2RAs if intolerant to PPIs).

Rationale for the study design

The original study design was based on the belief that decisions about the management of GORD should be made using unbiased, statistically precise comparisons of alternative policies. At study entry all patients fulfilled three criteria: they were on long-term acid suppression with PPIs; they had symptoms that were thought to be adequately controlled; and they were suitable in terms of fitness and comorbidity for either surgical or continuing medical treatment for their GORD. At the time that the study was planned, the consensus opinion of clinicians was that these three criteria identified GORD patients for whom surgical and continuing medical treatment could be considered equally acceptable treatment options and that, consequently, the comparison should be undertaken in patients meeting these criteria.

The most likely sources of bias were in the ways in which the groups being compared were selected; how their outcomes were assessed; and how the management was actually delivered. This is the basis for using a pragmatic randomised controlled trial (RCT) design. Random allocation protected against selection bias.

Confining the trial to those with no clear treatment preference limits biased patient-centred assessment of outcome, and pragmatic comparison of alternative policies [with intention-to-treat (ITT) analysis] avoids bias introduced by individual cases of non-compliance. This approach had limitations, however, and for this reason we chose to incorporate two parallel, non-randomised preference groups.

Including those with a clear preference for one policy or the other allows broader extrapolation and generalisability. Study of this group may give insights into the reasons for preference and hence give pointers to patient choices after the study.[26] Furthermore, preference may influence outcome and, if so, this may also help when making treatment decisions.[26,27] A third reason for the parallel, non-randomised preference groups[28] was that the addition of data from the preference groups may reduce imprecision around the estimates from the randomised comparison and this may be particularly useful for rare events, such as complications that can be confidently ascribed to one or other treatment. (The limitation is that the preference groups are not derived by random allocation, and hence the comparisons are exposed to the biases of non-randomised studies.)

Reliable comparisons within and between randomised and preference groups require valid measurement of treatment outcome. Although there were a number of quality-of-life (QoL) tools available, none was sufficiently specific to assess the spectrum of gastrointestinal symptoms associated with the treatment of GORD, particularly those due to surgery. For this reason we developed and validated a new outcome measure (the REFLUX questionnaire). We have continued to use this as the primary outcome measure in the longer-term follow-up reported here. Details of the REFLUX questionnaire and its derivation have been described elsewhere.[1,4]

Gastro-oesophageal reflux disease and its management represent a very significant call on NHS resources. Although clinical effectiveness, acceptability and safety will be important determinants of future policy, the issues of cost and resource use may be over-riding. This is the reason for the economic evaluation component of this study. Policy should be guided by both assessment of the relative cost-effectiveness of alternative policies and assessment of the impact that possible policy changes would have for the NHS and for patients with GORD.

The cost of laparoscopic fundoplication appears to be equivalent to the cost of 2–3 years of maintenance treatment with PPIs, although it is acknowledged that the costs of PPIs are falling.[29] The costs of surgery are related largely to two factors: the incidence of complications/length of hospital stay and the number of patients requiring long-term medical interventions after surgery.

We addressed cost-effectiveness in our report of the first phase of the REFLUX trial.[1] We reported both a within-trial cost-effectiveness analysis based on the results up to 12 months after surgery and an extended cost-effectiveness model that explored a number of scenarios beyond 12 months. The within-trial analysis related the extra mean costs associated with the surgical policy to the estimated increase in mean quality-adjusted life-years (QALYs) associated with surgery up to that time. The incremental cost-effectiveness ratio (ICER) was around £19,000 when the ITT analysis was used. Taking into account the uncertainties around the estimates of both costs and utilities, it was calculated that the chance that the surgical policy would be cost-effective at a threshold of £20,000 per QALY was 46%. This indicated considerable uncertainty at thresholds that are currently commonly applied to costs per QALY. The limitations of the within-trial analysis were discussed in detail in the earlier report, in particular that it ignored costs and benefits that accrued after 1 year.

The economic model was designed to address the limitations of the within-trial analysis. It explored a range of scenarios of varying lifetime benefits and costs, and analyses gave a wide range of incremental costs per QALY of £1000–44,000, again indicative of wide uncertainty. The factors contributing most to this uncertainty were the projected health-related quality-of-life (HRQoL) parameters and the long-term uptake of medication following surgery.

Thus, although data available up to a time equivalent to 1 year after surgery provided promising evidence that surgical management might well be cost-effective, there was too much uncertainty, especially about longer-term costs and benefits, to provide clear guidance for decision-makers. This was the justification for the longer-term follow-up to 5 years reported here.

Chapter 2 Methods

Original study design

The study had two complementary components:

1. a multicentre, pragmatic[30] RCT (with parallel non-randomised preference groups) comparing a laparoscopic surgery-based policy with a continued medical management policy to assess their relative clinical effectiveness
2. an economic evaluation of laparoscopic surgery for GORD to compare the cost-effectiveness of the two management policies, identify the most efficient provision of future care and describe the resource impact that various policies for fundoplication would have on the NHS.

Eligible patients who consented to participate in the RCT were randomly allocated to either laparoscopic surgery or continued medical management. Those patients who had a strong preference for one or other of the two treatment options could be recruited to the preference study. Clinical history was recorded at study entry. Participants completed health status questionnaires at the time of recruitment to the study and then at specified times equivalent to 3 and 12 months and then 2, 3, 4 and 5 years after surgery.

Approval for this study was obtained from the Scottish Multicentre Research Ethics Committee and the appropriate Local Research Ethics Committees.

Clinical centres

Clinical centres were based on local partnerships between surgeons with experience of laparoscopic fundoplication and gastroenterologists, with whom they shared the secondary care of patients with GORD. Centres were eligible if they included:

- a surgeon who had performed at least 50 laparoscopic fundoplication operations
- one or more gastroenterologists who agreed to collaborate with the surgeon(s) in the trial.

Study population

Eligible patients were those for whom care had been provided by a participating clinician who was uncertain which management policy (surgical or medical) was better. In addition, patients had to have documented evidence of GORD (based on endoscopy and/or manometry/24-hour pH monitoring) as well as symptoms for >12 months requiring maintenance PPI therapy for reasonable symptom control. Patients who were intolerant to PPIs and therefore required H_2RA therapy to control their symptoms were also eligible. Patients who were morbidly obese [body mass index (BMI) >40 kg/m^2] or who had Barrett's oesophagus of >3 cm or evidence of dysplasia, a paraoesophageal hernia or an oesophageal stricture were all excluded.

Eligible patients who did not want to take part in the randomised trial because of a strong preference for one type of management or the other were invited to take part in the preference arm of the study. For logistical reasons and to maintain a balance between the sizes of the randomised and the preference groups, we aimed to cap the numbers of participants recruited to the preference arms to 20 per arm in each centre.

All participants gave their informed consent.

Health technology policies being compared

Laparoscopic surgery policy

For those participants allocated to the randomised surgical group or recruited to the preference surgical group of the trial, subsequent deferring or declining of surgery, by either the participant or the surgeon, was always an option (i.e. even after trial entry), particularly among those recruited by a gastroenterologist and referred to a surgeon for consideration of surgery within the trial. Participants who had not had manometry/pH studies underwent these tests before surgery to exclude achalasia.

The surgery was performed either by an experienced surgeon who had undertaken >50 laparoscopic fundoplications or by a less experienced surgeon working under the supervision of an experienced surgeon. It was recommended that crural repair be routine and that non-absorbable synthetic sutures (not silk) be used for the repair. The type of fundoplication used was left to the discretion of the experienced surgeon. For the purposes of the main comparisons, the different surgical techniques for laparoscopic fundoplication were considered as part of a single policy. The study design, however, allowed for indirect comparisons between techniques.

Medical management policy

Those allocated to the medical management policy had their therapy reviewed and adjusted as necessary by the local gastroenterologist to be 'best medical management'. It was recommended that management conformed to the principles of the Genval Workshop Report.[31] These include stepping down antisecretory medication in most patients to the lowest dose that maintained acceptable symptom control. Following the therapy review by the gastroenterologist, trial participants had their medication managed by their general practitioner (GP). Although, in general, trial participants allocated to medical management were managed in this way, the protocol did include the option of surgery if a clear indication for it subsequently developed.

Study registration (and treatment allocation when randomised)

The treatment allocation for participants in the randomised component of the trial was computer generated; it was stratified by centre, with balance in respect of other key prognostic variables – age (18–49 years or 50+ years), sex (male or female) and BMI (≤28 or >29 kg/m^2) – by a process of minimisation. Randomisation was organised centrally at the Health Services Research Unit, University of Aberdeen, and was independent of all clinical collaborators.

Clinical management

Participants who were allocated to surgical management were invited to a consultation with the collaborating surgeon. During this consultation, the surgeon confirmed that there was no contraindication to surgery and discussed the operation in more detail, before arranging an operation date. The surgeon recorded intraoperative details on specially designed study forms. All other in-hospital data collection was the responsibility of the local study nurse. In all respects, other than the trial interventions, clinical management was left to the discretion of the clinician responsible for care. This continued to be the case in the extended follow-up phase, which is the focus of this report, with GPs monitoring subsequent care needs throughout the follow-up period.

Data collection

Follow-up by postal questionnaire was first performed at 3 months after surgery, or at an equivalent time among those who did not have surgery, and then annually. The questionnaire used for the follow-up at 2–5 years was similar to the questionnaire that had been used in the earlier phase of the trial up to 12 months after surgery. Non-responders received up to two reminder telephone calls or letters to encourage return of their postal questionnaires. On occasion, and at the participants' convenience, a shortened version of the questionnaire was completed over the telephone.

From around half-way through the 5-year follow-up, participants were sent a £5 gift voucher with their final postal reminder to compensate for their time in completing the questionnaire. This decision was taken based on the findings of a systematic review of the effects of incentives on postal questionnaire return[32] and specific randomised trials that evaluated the use of vouchers.[33–35]

All data were sent to the trial office in Aberdeen for processing. A random 10% sample of all data was double-entered to check accuracy and no significant errors were identified. Extensive range and consistency checks further enhanced the quality of the data.

The principal study outcome measure

The primary outcomes for measuring the differences in effects between medical and surgical management were:

- a 'disease-specific' measure incorporating assessment of reflux and other gastrointestinal symptoms and the side effects and complications of both therapies (the REFLUX questionnaire was developed specifically for this study[4])
- NHS costs including treatments, investigations, consultations and other contacts with the health service.

The secondary outcome measures were:

- HRQoL – measured using the European Quality of Life-5 Dimensions (EQ-5D)[36] and Short Form questionnaire-36 items (SF-36)[37]
- patient costs, including loss of earnings, reduction in activities and the costs of prescriptions and travel to health care
- other serious morbidity, such as operative complications
- (further) anti-reflux surgery
- mortality.

An example of the annual questionnaire used for collecting this information is provided in *Appendix 1*.

Sample size

The original aim was to recruit 600 participants to the randomised trial to give 80% power to identify a difference between the two groups of 0.25 of a standard deviation (SD) in respect of the disease-specific instrument and other continuous variables such as EQ-5D and SF-36, using a significance level of 5%. Based on the same arguments, it was planned that 300 people would be recruited to each arm of the preference study. The cost savings of a surgical policy largely depend on the number of patients managed surgically who no longer require PPI treatment, and a trial with 300 surgically managed patients would have estimated this proportion to within about 5% with 95% statistical confidence.

However, prompted by a lower rate of recruitment than expected, this target was revised in January 2003 in consultation with the Data Monitoring Committee (DMC) and representatives of the HTA programme. It was agreed that a larger benefit (0.3 of a SD) was clinically plausible based on improvements seen after surgery in the accruing literature among more severely affected people (who were not eligible for the trial). This was calculated to require 196 in each group to give 80% power ($2p = 0.05$).

Statistical considerations

This report describes analyses of annual questionnaire data up to 5 years after surgery (or an equivalent time if managed medically). As a general rule, in the tables and analyses presented in this report, the participants in the randomised groups are separate from those in the preference groups. A sizeable group of patients allocated to surgery did not receive surgery. Therefore, to investigate the potential influence of this non-compliance with allocation, summary statistics in the results tables are given for four main analysis populations (comprising eight groups of participants):

1. Randomised ITT population (groups that were randomised to either surgery or medical management).
2. Per-protocol (PP) population (groups that were either randomised to surgery and received surgery in the first year or randomised to medical management and did not receive surgery in the first year).
3. Preference ITT population (groups that preferred either surgery or medical management at recruitment).
4. Preference PP population (groups that either preferred surgery at recruitment and received surgery in the first year or preferred medical management and did not receive surgery in the first year).

The primary outcome measure (REFLUX QoL score) and secondary outcome measures (SF-36, EQ-5D, REFLUX symptom scores, anti-reflux surgery and use of reflux-related drugs) were analysed using general linear models. The analyses adjusted for the minimisation covariates (age, BMI and sex) and where appropriate (defined by significant at the 5% significance level) also adjusted for baseline measures and baseline measures by treatment interaction. A secondary, pre-stated subgroup analysis explored the differential effects of surgeon's preferred operative procedure on the primary outcome measure. All analyses were reported with 95% confidence intervals (CIs).

The primary analysis of the randomised groups was by ITT. The ITT approach sustains the integrity of the randomisation and gives the least biased estimate of effectiveness of the two forms of management. Given that a sizeable minority of the randomised surgical participants did not receive surgery, we were also interested in estimating the efficacy of the initial treatment received as a secondary comparison (i.e. commonly known as a PP analysis). In an open trial design a PP analysis can have substantial selection bias. To minimise the effects of selection bias we used the method of 'adjusted treatment received' as described by Nagelkerke *et al.*[38] and others.[39,40] The method used a two-stage least-squares approach whereby treatment randomised was regressed onto treatment received and the residuals from that model were used as an independent variable in a second model, together with the treatment received, to estimate the effects on the various primary and secondary outcome measures.

For the preference study, only the primary outcome was analysed statistically. The analysis compared the preference surgical group with the preference medical group and adjusted for the minimisation factors. As described above, for logistical reasons and to maintain balance between the randomised and preference groups, we capped the number of preference participants at 20 per group per centre. The study design was not therefore a true comprehensive cohort. We did consider modelling differences between the randomised and preference groups; however, it is not universally accepted that formal modelling is appropriate in this context. In this case we knew from the randomised arms that there was a strong interaction between treatment effects and baseline REFLUX QoL, and in addition we knew that there was a large difference in QoL between preference arms at baseline (and patient demographics such as age and

sex). We therefore decided that formal modelling of the arms would add little to the comparison given the large confounding between preference groups.

Sensitivity analyses

The sensitivity of the primary outcome analysis result was investigated using two approaches – the effect of excluding a large centre and the effects of missing data. In the first approach the largest recruiting centre, Aberdeen, was excluded and the analysis as described above was rerun. Second, previous work demonstrated that the primary outcome was likely missing at random (MAR) or missing completely at random (MCAR) and that a repeated measures analysis (using all available data) was an appropriate statistical method for analysing data up to 12 months.[41] We therefore used a repeated measures analysis on the primary outcome across all of the follow-up data (12 months to 5 years) to investigate the effect of incorporating a profile of measures for each participant. No further imputation for missing values was necessary.

Data monitoring

During recruitment, an independent DMC met on three occasions and each time saw no reason to recommend any fundamental changes to the protocol. The committee did not meet after recruitment was completed.

Chapter 3 Trial results and clinical effectiveness

Recruitment to the trial

Participants were recruited in 21 clinical centres, all within the UK (their locations are listed on the left-hand side of *Table 1*). Recruitment to the trial was open from March 2001 until the end of June 2004, although not all centres enrolled over the total period because of the staggered introduction of centres and early closure for logistical reasons in a few places.[1]

A total of 357 participants were recruited to the randomised component: 178 allocated to surgery and 179 allocated medical management. 453 participants agreed to join the preference component: 261 choosing surgery and 192 choosing medical management. *Table 1* shows recruitment by centre. Around one-fifth of the randomised participants were enrolled in Aberdeen; no centre contributed >10% of participants in the preference component.

Analysis populations

Throughout the analyses presented later in this chapter, the participants in the randomised component are kept separate from those in the preference component (other than for rare surgical events). The numbers of participants in each of the four main analysis populations are shown in *Table 2*. All 357 who joined the randomised component are in the randomised ITT population; only the 280 within this group who actually received their allocated management over the first year are in the randomised PP population. All 453 participants who joined the preference component are in the preference ITT population; the 407 of these who, by the end of the first year, were managed as originally chosen were in the preference PP population.

Trial conduct

The derivation of the main study groups and their progress through the stages of follow-up in the trial are shown in *Figure 1*. This is in the form of a CONSORT (Consolidated Standards of Reporting Trials) flow diagram. In total, 1078 patients were considered for trial entry and 200 of these were found not to meet one or more of the eligibility criteria. Of the 68 patients eligible for the study but not recruited, 51 declined to participate, six were subsequently deemed inappropriate for the study by the surgeon responsible for care and the remaining 11 were missed.

Details of the clinical management actually received are described later in this chapter.

The mean (SD) time intervals in months between the receipt by the trial office of each subsequent annual postal questionnaire are shown in *Table 3*; all were near 12 months, as would be expected. There was, however, a difference between the randomised groups in the time interval between the 1-year and the 2-year questionnaires (mean 12.2 months surgical group vs 13.9 months medical group). In part, this was due to more late returns in the medical management group – the median intervals were closer: 12.00 and 13.00 months respectively. As described previously,[1] early follow-up was adjusted to be at a time equivalent to 3 and 12 months after surgery. The adjustments in the medical group to match this could be only approximate and this is the explanation for the difference that remained between the randomised groups. An advantage of long-term follow-up to 5 years is that any difference in the timing of follow-up becomes proportionately smaller over time.

TABLE 1 Number of participants by centre

	Randomised participants, *n* (%)		Preference participants, *n* (%)	
	Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)
Aberdeen: Aberdeen Royal Infirmary	38 (21.3)	40 (22.3)	20 (7.7)	21 (10.9)
Belfast: Royal Victoria Hospital	15 (8.4)	14 (7.8)	4 (1.5)	20 (10.4)
Bournemouth: Royal Bournemouth Hospital	4 (2.2)	3 (1.7)	20 (7.7)	3 (1.6)
Bristol: Bristol Royal Infirmary	12 (6.7)	11 (6.1)	18 (6.9)	20 (10.4)
Bromley: Princess Royal Infirmary	3 (1.7)	3 (1.7)	20 (7.7)	17 (8.9)
Edinburgh: Royal Infirmary of Edinburgh	11 (6.2)	11 (6.1)	1 (0.4)	15 (7.8)
Guildford: Royal Surrey County Hospital	10 (5.6)	10 (5.6)	17 (6.5)	10 (5.2)
Hull: Hull Royal Infirmary	7 (3.9)	7 (3.9)	1 (0.4)	2 (1.0)
Inverness: Raigmore Hospital	7 (3.9)	8 (4.5)	2 (0.8)	8 (4.2)
Leeds: Leeds General Infirmary	1 (0.6)	2 (1.1)	10 (3.8)	3 (1.6)
Leicester: Leicester Royal Infirmary	0 (0.0)	0 (0.0)	3 (1.1)	1 (0.5)
London: St Mary's Hospital	8 (4.5)	7 (3.9)	4 (1.5)	10 (5.2)
London: Whipps Cross Hospital	4 (2.2)	3 (1.7)	16 (6.1)	5 (2.6)
Poole: Poole Hospital	10 (5.6)	10 (5.6)	25 (9.6)	13 (6.8)
Portsmouth: Queen Alexandra Hospital	10 (5.6)	10 (5.6)	15 (5.7)	1 (0.5)
Salford: Hope Hospital	0 (0.0)	1 (0.6)	6 (2.3)	3 (1.6)
Stoke-on-Trent: North Staffordshire Hospital	5 (2.8)	6 (3.4)	20 (7.7)	9 (4.7)
Swansea: Morriston Hospital	8 (4.5)	8 (4.5)	14 (5.4)	9 (4.7)
Telford: Princess Royal Hospital	11 (6.2)	12 (6.7)	24 (9.2)	8 (4.2)
Yeovil: Yeovil District Hospital	9 (5.1)	8 (4.5)	18 (6.9)	8 (4.2)
York: York District Hospital	5 (2.8)	5 (2.8)	3 (1.1)	6 (3.1)
Total	178 (100)	179 (100)	261 (100)	192 (100)

TABLE 2 Number of participants in each analysis population

	Surgical, *n* (%)	Medical, *n* (%)	Total, *n*
Randomised ITT	178 (49.9)	179 (50.1)	357
Randomised PP	111 (39.6)	169 (60.4)	280
Preference ITT	261 (57.6)	192 (42.4)	453
Preference PP	218 (53.6)	189 (46.4)	407

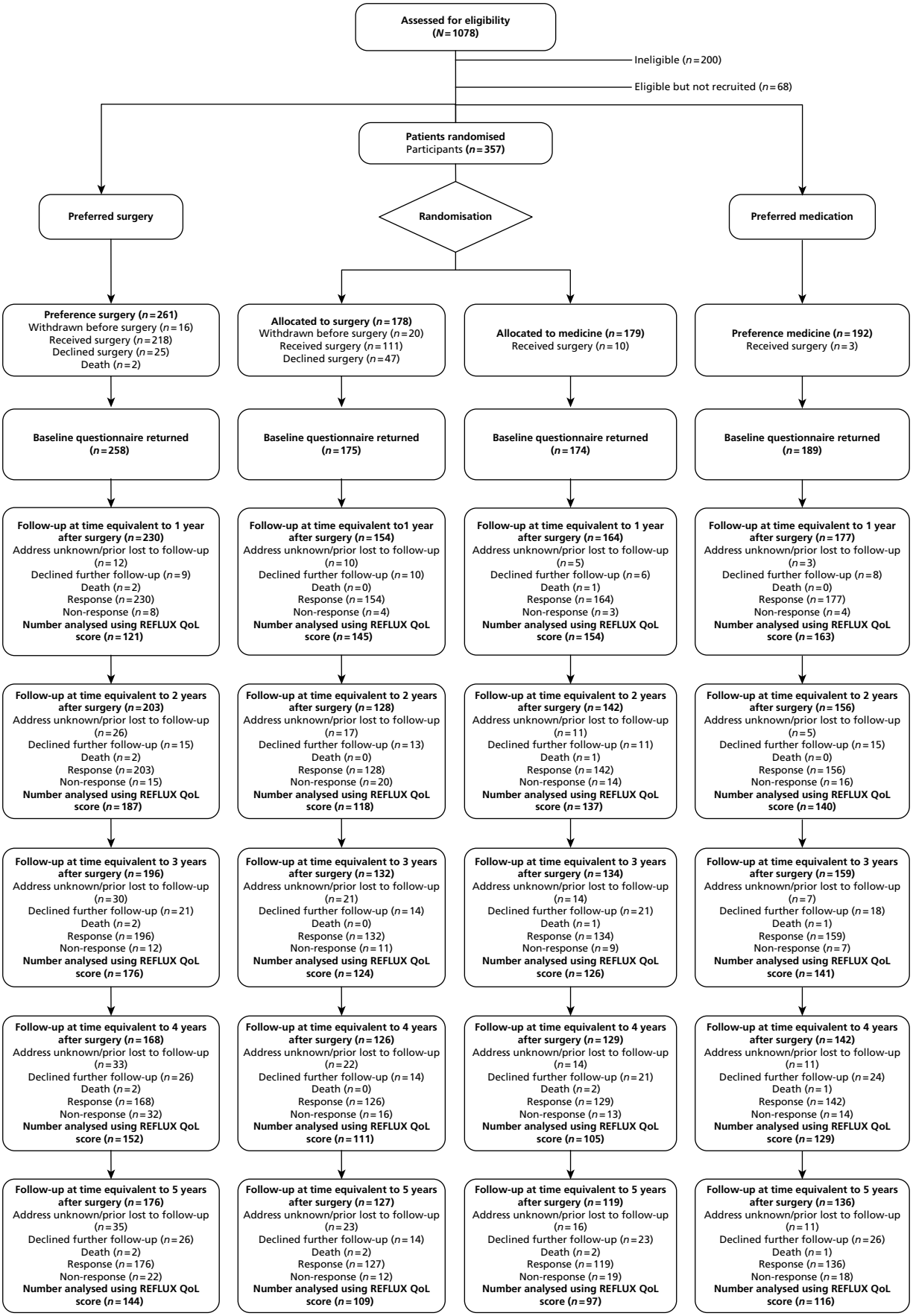


FIGURE 1 The CONSORT diagram.

TABLE 3 Interval between randomisation and follow-up (months), mean (SD)

	Randomised participants		Preference participants	
	Surgical	Medical	Surgical	Medical
	ITT (*n* = 178)	ITT (*n* = 179)	ITT (*n* = 261)	ITT (*n* = 192)
1 year to 2 years	12.2 (1.9)	13.9 (3.1)	12.4 (1.8)	12.9 (4.6)
2 years to 3 years	11.8 (1.2)	11.6 (1.2)	11.6 (1.5)	11.8 (1.2)
3 years to 4 years	12.0 (1.5)	12.0 (1.4)	12.1 (1.2)	12.0 (1.1)
4 years to 5 years	11.8 (1.3)	12.0 (1.3)	12.1 (1.5)	12.0 (1.3)

More details of the response rates to the annual questionnaires are provided in *Table 4*. The overall rates of return of annual follow-up questionnaires (years 1–5) were 89.5%, 77.7%, 76.7%, 69.8% and 68.9% of the study participants. Seven participants are known to have died up to the end of the 5-year follow-up; equivalent response rates among those not known to have died are 89.8%, 77.9%, 77.0%, 70.2% and 69.5%. There were no substantive differences in response rates between the groups.

Three participants died before the 1-year follow-up was reached: two in the preference surgery group and one in the randomised medical group. None of these participants actually had surgery. Four died subsequently; there is no evidence linking these deaths to trial participation.

Description of the groups at trial entry

Sociodemographic and clinical factors

Table 5 provides a description of the groups at trial entry. The main division within the table is between participants in the randomised component and those in the preference component. These two halves of the table are further divided according to the allocation of participants and then subdivided according to ITT or PP.

Randomised arms

Within the randomised groups there were no apparent imbalances between the medical and surgical intervention arms. The patients were, on average, 46 years old, 66% were men, around two-thirds were in full employment and participants had been on GORD medication for a median of 32 months. The baseline characteristics in the randomised PP groups were similar.

Preference arms

The sociodemographic characteristics of the preference participants were broadly similar to those of the randomised participants. However, preference medical participants tended to be older (mean age 50 years) and were more likely to be female, fewer were in full-time employment and participants had been on GORD medication for a shorter period (approximately 6 months less than randomised participants).

Prescribed medications

The prescribed medications at the time of trial entry are shown in *Table 6*. There was a similar profile of prescribed medications across the randomised and preference groups. As would be expected, nearly all participants reported taking a reflux-related drug in the previous 2 weeks. Over 90% had taken a PPI, of which lansoprazole was the most common.

TABLE 4 CONSORT table

Year	Category	Randomised participants, *n* (%)		Preference participants, *n* (%)	
		Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)
1	Responded	154 (87)	164 (92)	230 (88)	177 (92)
	Declined further follow-up	10 (6)	6 (3)	9 (3)	8 (4)
	Deceased	0 (0)	1 (1)	2 (1)	0 (0)
	Address unknown/lost to follow-up	10 (6)	5 (3)	12 (5)	3 (2)
	Non-responder	4 (2)	3 (2)	8 (3)	4 (2)
2	Responded	128 (72)	142 (79)	203 (78)	156 (81)
	Declined further follow-up	13 (7)	11 (6)	15 (6)	15 (8)
	Deceased	0 (0)	1 (1)	2 (1)	0 (0)
	Address unknown/lost to follow-up	17 (10)	11 (6)	26 (10)	5 (3)
	Non-responder	20 (11)	14 (8)	15 (6)	16 (8)
3	Responded	132 (74)	134 (75)	196 (75)	159 (83)
	Declined further follow-up	14 (8)	21 (12)	21 (8)	18 (9)
	Deceased	0 (0)	1 (1)	2 (1)	1 (1)
	Address unknown/lost to follow-up	21 (12)	14 (8)	30 (11)	7 (4)
	Non-responder	11 (6)	9 (5)	12 (5)	7 (4)
4	Responded	126 (71)	129 (72)	168 (64)	142 (74)
	Declined further follow-up	14 (8)	21 (12)	26 (10)	24 (13)
	Deceased	0 (0)	2 (1)	2 (1)	1 (1)
	Address unknown/lost to follow-up	22 (12)	14 (8)	33 (13)	11 (6)
	Non-responder	16 (9)	13 (7)	32 (12)	14 (7)
5	Responded	127 (71)	119 (66)	176 (67)	136 (71)
	Declined further follow-up	14 (8)	23 (13)	26 (10)	26 (14)
	Deceased	2 (1)	2 (1)	2 (1)	1 (1)
	Address unknown/lost to follow-up	23 (13)	16 (9)	35 (13)	11 (6)
	Non-responder	12 (7)	19 (11)	22 (8)	18 (9)

Health status

Randomised arms

The HRQoL scores at study entry are displayed in *Table 7*. The scores were broadly similar in the randomised surgical and randomised medical groups, although they were slightly higher (better health) in the randomised medical group. When the DMC first met after the initial 143 participants had been recruited to the randomised component, the committee did ask us to change the enrolment procedure to ensure that baseline questionnaires were completed *before* formal entry and randomisation. We understand that this was because they were concerned about an apparent imbalance between the randomised groups in baseline health status at that time. After satisfying themselves that this was not due to a breakdown in the randomisation procedure, the DMC surmised that this might be due to prior knowledge of the treatment allocation affecting questionnaire responses (with those allocated surgery tending to project worse health status than those allocated medical management). Certainly, the groups

TABLE 5 Description of groups at trial entry

Characteristic	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
Baseline questionnaire returned, *n* (%)	175 (98.3)	111 (100.0)	174 (97.2)	165 (97.6)	256 (98.1)	216 (99.1)	189 (98.4)	186 (98.4)
Age (years), mean (SD)	46.7 (10.3)	46.3 (10.2)	45.9 (11.9)	45.9 (11.9)	44.4 (12.0)	44.5 (12.2)	49.9 (11.8)	50 (11.7)
Male, *n* (%)	116 (65.2)	68 (61.3)	120 (67.0)	115 (68.0)	170 (65.1)	139 (63.8)	111 (57.8)	110 (58.2)
BMI (kg/m^2), mean (SD)	28.5 (4.3)	28.7 (4.1)	28.4 (4.0)	28.3 (4.0)	27.7 (4.0)	27.5 (3.7)	27.4 (4.1)	27.4 (4.1)
Duration of prescribed medication for GORD (months), median (IQR)	33 (15–83)	30 (16–76)	31 (16–71)	30 (15–71)	35 (14–71)	36 (14–65)	27 (13–60)	26.5 (13–60)
Employment status, *n* (%)								
Employed full-time	116 (66.3)	72 (65.5)	110 (61.8)	104 (61.9)	168 (65.1)	138 (64.2)	100 (52.4)	97 (51.6)
Employed part-time	13 (7.4)	12 (10.9)	16 (9.0)	15 (8.9)	35 (13.6)	29 (13.5)	20 (10.5)	20 (10.6)
Student	5 (2.9)	3 (2.7)	3 (1.7)	3 (1.8)	2 (0.8)	2 (0.9)	3 (1.6)	3 (1.6)
Retired	12 (6.9)	9 (8.2)	22 (12.4)	20 (11.9)	18 (7.0)	16 (7.4)	35 (18.3)	35 (18.6)
Housework	11 (6.3)	6 (5.5)	10 (5.6)	10 (6.0)	17 (6.6)	15 (7.0)	15 (7.9)	15 (8.0)
Seeking work	6 (3.4)	1 (0.9)	3 (1.7)	2 (1.2)	5 (1.9)	5 (2.3)	2 (1.0)	2 (1.1)
Other	12 (6.9)	7 (6.4)	14 (7.9)	14 (8.3)	13 (5.0)	10 (4.7)	16 (8.4)	16 (8.5)

Characteristic	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
Age (years) left full-time education, *n* (%)								
≤16	110 (62.5)	68 (62.4)	108 (60.7)	102 (60.7)	151 (58.5)	128 (59.3)	105 (55.3)	104 (55.6)
17–19	38 (21.6)	24 (22.0)	40 (22.5)	40 (23.8)	63 (24.4)	51 (23.6)	45 (23.7)	43 (23.0)
20+	28 (15.9)	17 (15.6)	30 (16.9)	26 (15.5)	44 (17.1)	37 (17.1)	40 (21.1)	40 (21.4)
Current smoker, *n* (%)	46 (25.8)	29 (26.1)	40 (22.3)	36 (21.3)	71 (27.2)	61 (28.0)	39 (20.3)	39 (20.6)
Erosive oesophagitis, *n* (%)	85 (54.8)	48 (50.0)	97 (62.2)	91 (62.3)	104 (46.4)	80 (43.2)	87 (50.9)	86 (51.2)
Comorbidity: *Helicobacter pylori* status, *n* (%)								
Positive (subsequently treated)	12 (9.0)	5 (6.1)	14 (10.4)	13 (10.3)	18 (8.4)	14 (7.9)	15 (10.5)	15 (10.7)
Positive (subsequently untreated)	1 (0.8)	0 (0.0)	3 (2.2)	3 (2.4)	8 (3.7)	8 (4.5)	2 (1.4)	2 (1.4)
Negative	75 (56.4)	48 (58.5)	73 (54.1)	67 (53.2)	118 (54.9)	101 (56.7)	74 (51.7)	72 (51.4)
Uncertain	45 (33.8)	29 (35.4)	45 (33.3)	43 (34.1)	71 (33.0)	55 (30.9)	52 (36.4)	51 (36.4)
Hiatus hernia present, *n* (%)	94 (57.3)	64 (61.0)	102 (60.4)	94 (59.1)	168 (68.9)	146 (71.2)	101 (59.8)	99 (59.6)
Asthma, *n* (%)	21 (11.9)	14 (12.7)	21 (11.8)	19 (11.3)	30 (11.5)	23 (10.6)	36 (18.8)	36 (19.0)

IQR, interquartile range.

TABLE 6 Description of groups at trial entry: prescribed medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
PPIs, *n* (%)								
Any PPI	161 (92.0)	105 (94.6)	162 (93.1)	153 (92.7)	225 (87.9)	191 (88.4)	173 (91.5)	170 (91.4)
Omeprazole	46 (26.3)	32 (28.8)	46 (26.4)	43 (26.1)	49 (19.1)	36 (16.7)	61 (32.3)	61 (32.8)
Lansoprazole	77 (44.0)	47 (42.3)	72 (41.4)	69 (41.8)	100 (39.1)	92 (42.6)	69 (36.5)	66 (35.5)
Pantoprazole	6 (3.4)	6 (5.4)	11 (6.3)	11 (6.7)	21 (8.2)	17 (7.9)	11 (5.8)	11 (5.9)
Rabeprazole	12 (6.9)	6 (5.4)	13 (7.5)	13 (7.9)	21 (8.2)	16 (7.4)	14 (7.4)	14 (7.5)
Esomeprazole	20 (11.4)	14 (12.6)	20 (11.5)	17 (10.3)	37 (14.5)	33 (15.3)	18 (9.5)	18 (9.7)
H_2RAs, *n* (%)								
Any H_2RA	14 (8.0)	6 (5.4)	12 (6.9)	9 (5.5)	22 (8.6)	16 (7.4)	13 (6.9)	13 (7.0)
Ranitidine	13 (7.4)	6 (5.4)	8 (4.6)	6 (3.6)	11 (4.3)	7 (3.2)	11 (5.8)	11 (5.9)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.4)	1 (0.5)	1 (0.5)	1 (0.5)
Cimetidine	1 (0.6)	0 (0.0)	1 (0.6)	0 (0.0)	1 (0.4)	1 (0.5)	0 (0.0)	0 (0.0)
Nizatidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.2)	3 (1.4)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	0 (0.0)	0 (0.0)	4 (2.3)	3 (1.8)	7 (2.7)	4 (1.9)	2 (1.1)	2 (1.1)

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
Prokinetics, *n* (%)								
Any prokinetics	12 (6.9)	7 (6.3)	8 (4.6)	6 (3.6)	11 (4.3)	10 (4.6)	5 (2.6)	4 (2.2)
Domperidone	8 (4.6)	5 (4.5)	4 (2.3)	3 (1.8)	7 (2.7)	6 (2.8)	4 (2.1)	3 (1.6)
Metoclopramide	4 (2.3)	2 (1.8)	4 (2.3)	3 (1.8)	4 (1.6)	4 (1.9)	1 (0.5)	1 (0.5)
Any reflux-related drug, *n* (%)	170 (97.1)	108 (97.3)	169 (97.1)	160 (97.0)	235 (91.8)	198 (91.7)	184 (97.4)	181 (97.3)
Other prescribed drugs, *n* (%)[a]								
Alginates	22	12	21	18	37	33	14	13
Antispasmodics (e.g. dicycloverine)	0	0	2	2	3	3	0	0
Chelates (e.g. sucralfate)	1	1	0	0	0	0	0	0
Other ulcer-healing drugs	0	0	0	0	1	1	0	0
Mucogel® (Chemidex)	0	0	1	1	1	1	1	1
Asilone® (Thornton & Ross)	0	0	1	1	0	0	0	0
Non-gastrointestinal	7	2	4	4	5	4	6	6
Anti-nausea	0	0	1	1	1	1	1	1

a More than one prescription per person possible.

based on the first 143 participants were well balanced in other respects, and there was subsequently good balance in health status as well. The apparent small imbalance between the randomised groups in health status measures is therefore likely to be a reflection of the imbalance in the first 143 participants.

The most prevalent reflux symptoms (those with lowest scores) were general discomfort and wind. The participants had lower SF-36 and EQ-5D scores than a normal UK population with the same average age and sex characteristics (SF-36 population norm approximately 50 for all domains; EQ-5D norm 0.88).

Preference arms

The preference for surgery participants reported worse REFLUX QoL scores and worse health in general than the preference for medicine participants. It can be seen from *Table 7* that the randomised participants reported QoL measures in between these two extremes.

Baseline characteristics of groups compared at 5 years

There were differences in baseline characteristics between those who had completed a questionnaire at 5 years and those who had not (*Table 8*). For example, responders had a higher mean age (47.9 years vs 43.6 years), had been on prescribed medication for a shorter period at recruitment to the REFLUX trial (50.5 months vs 60.2 months) and had higher QoL scores at baseline (measured on the disease-specific REFLUX instrument, EQ-5D and SF-36).

However, the baseline characteristics of those in the randomised surgical and randomised medical groups who completed a questionnaire at 5 years were very similar, with the only notable difference being in BMI (*Table 9*). The mean baseline BMI among responders in the randomised surgical group was higher (29.0 kg/m^2) than that for responders in the randomised medical management group (27.7 kg/m^2). As described in *Chapter 2*, these results confirmed that a repeated measures analysis assuming no differential loss to follow-up could be considered.

Surgical management

Table 10 summarises the use of surgery in the four study groups over the full 5-year follow-up period. At the end of the first year, 111 participants (62.4%) randomised to surgery had actually undergone fundoplication. Over the next 4 years, one more member of this group had fundoplication, bringing the total to 112 (62.9%). In the randomised medical group, 10 participants (5.6%) had fundoplication in the first year, with a further 14 participants having fundoplication in subsequent years, bringing the total at 5 years to 24 (13.4%). In the preference surgical group, 218 participants (83.5%) had fundoplication in the first year, with four more in the period up to 5 years, taking the percentage to 85.1%. Surgical management applied to only three participants (1.6%) in the preference medical group in the first year, with a further three being operated on in the subsequent 4 years (total 3.1%).

Information about the reasons why participants allocated surgery did not receive it in the first year is available for 47. For 25 of these 47, this was a clinical decision, most commonly the surgeon deciding that surgery was not appropriate; most of the other 22 changed their minds about surgery for a variety of work- or home-related reasons. A further 20 withdrew for unknown reasons. There is no doubt, however, that a number of these participants suffered long delays before being formally offered surgery, and this was an important factor in their eventual decision to choose not to have surgery after all. The trial was conducted at a time when there was great pressure on surgical services in the NHS, with long delays for elective surgery for non-life-threatening benign conditions being common. Indeed, the average time between trial entry and surgery in the trial was 8–9 months.[1]

TABLE 7 Description of groups at trial entry: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
HRQoL instrument	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
REFLUX QoL, mean (SD)	63.6 (24.1)	61.9 (24.5)	66.8 (24.5)	68.2 (24.2)	55.8 (23.2)	55.9 (23.2)	77.5 (19.7)	78.0 (19.1)
REFLUX symptom score, mean (SD)								
General discomfort symptom score	58.5 (24.5)	57.1 (25.1)	61.3 (25.8)	62.4 (25.7)	49.1 (24.4)	48.7 (25.2)	73.1 (21.3)	73.6 (20.9)
Wind and frequency symptom score	48.1 (20.9)	46.2 (20.9)	49.3 (21.4)	49.5 (21.7)	47.1 (21.4)	47.5 (21.2)	59.6 (22.7)	59.8 (22.7)
Nausea and vomiting symptom score	81.5 (19.5)	81.6 (18.8)	80.7 (21.9)	81.6 (21.7)	76.9 (19.9)	77.5 (19.5)	89.7 (13.6)	90.1 (12.9)
Activity limitation symptom score	78.5 (16.9)	77.6 (16.3)	78.9 (17.3)	79.5 (17.1)	74.4 (16.1)	73.9 (16.2)	86.8 (13.0)	87.0 (13.0)
Constipation and swallowing symptom score	77.5 (19.9)	77.3 (20.3)	74.8 (21.0)	75.6 (20.4)	75.8 (22.0)	74.8 (22.6)	83.0 (17.7)	83.3 (17.6)
SF-36 score, mean (SD)								
Norm-based physical functioning	46.8 (10.2)	46.1 (10.3)	47.5 (10.5)	47.7 (10.5)	46.3 (9.4)	46.1 (9.3)	47.1 (10.8)	47.0 (10.9)
Norm-based role physical	46.9 (10.7)	46.6 (10.8)	46.8 (10.6)	47.0 (10.4)	44.7 (10.9)	44.6 (10.7)	46.7 (10.9)	46.6 (10.9)
Norm-based bodily pain	44.4 (10.1)	44.1 (9.9)	44.6 (10.4)	44.9 (10.3)	41.8 (9.5)	41.9 (9.6)	47.1 (9.8)	47.2 (9.8)
Norm-based general health	40.9 (9.9)	40.2 (9.6)	41.1 (10.6)	41.4 (10.6)	40.6 (10.2)	40.8 (10.0)	42.4 (10.0)	42.4 (9.9)
Norm-based vitality	43.5 (10.5)	43.9 (10.3)	44.0 (11.7)	44.4 (11.4)	42.8 (11.1)	42.8 (11.3)	45.5 (10.7)	45.6 (10.7)
Norm-based social functioning	44.4 (11.1)	44.1 (10.6)	44.7 (11.7)	45.2 (11.5)	42.2 (11.6)	42.1 (11.5)	46.8 (10.2)	46.7 (10.2)
Norm-based role emotional	46.6 (11.5)	47.2 (11.5)	45.8 (12.9)	46.3 (12.6)	45.9 (12.2)	46.1 (12.1)	46.9 (11.8)	46.8 (11.8)
Norm-based mental health	46.0 (11.6)	46.9 (11.0)	46.7 (11.6)	47.1 (11.3)	44.6 (11.4)	44.6 (11.6)	46.4 (10.7)	46.3 (10.8)
EQ-5D, mean (SD)	0.711 (0.258)	0.718 (0.239)	0.720 (0.255)	0.732 (0.246)	0.682 (0.259)	0.679 (0.259)	0.750 (0.223)	0.752 (0.222)
EQ-5D$_{VAS}$, mean (SD)	68.6 (17.1)	69.2 (15.9)	70.5 (18.1)	71.2 (17.6)	67.2 (18.5)	67.0 (18.5)	71.3 (16.7)	71.5 (16.6)

VAS, visual analogue scale.

TABLE 8 Baseline characteristics of responders and non-responders at 5 years

Characteristic	Responder (max. *n* = 558)	Non-responder (max. *n* = 252)	*p*-value (two- sided)
BMI (kg/m^2), mean (SD), *n*	27.9 (4.0), 557	28.2 (4.3), 252	0.37
Age (years), mean (SD), *n*	47.9 (11.2), 558	43.6 (12.2), 252	<0.01
Sex, *n/N* (%)			
Male	343/558 (61)	174/252 (69)	0.04
Female	215/558 (39)	78/252 (31)	–
Duration of prescribed medication (months), mean (SD), *n*	50.5 (62.9), 544	60.2 (65.2), 250	0.05
Erosive oesophagitis, *n/N* (%)			
Yes	262/493 (53)	111/213 (52)	0.78
No	231/493 (47)	102/213 (48)	
Helicobacter pylori status, *n/N* (%)			
Positive (subsequently treated)	39/440 (9)	20/186 (11)	0.81
Positive (subsequently untreated)	9/440 (2)	5/186 (3)	–
Negative	238/440 (54)	102/186 (55)	–
Uncertain	154/440 (35)	59/186 (32)	–
Hiatus hernia, *n/N* (%)			
Yes	330/524 (63)	135/222 (61)	0.58
No	194/524 (37)	87/222 (39)	–
Age (years) left full-time education, *n/N* (%)			
≤16	304/552 (55)	170/250 (68)	<0.01
17–19	143/552 (26)	43/250 (17)	–
20+	105/552 (19)	37/250 (15)	–
Employment status, *n/N* (%)			
Full-time	348/551 (63)	146/251 (58)	0.01
Part-time	65/551 (12)	19/251 (8)	–
Student	6/551 (1)	7/251 (3)	–
Retired	62/551 (11)	25/251 (10)	–
Housework	32/551 (6)	21/251 (8)	–
Seeking work	10/551 (2)	6/251 (2)	–
Other	28/551 (5)	27/251 (11)	–
REFLUX QoL, mean (SD), *n*	66.6 (24.2), 533	61.3 (24.1), 226	<0.01
REFLUX symptom score, mean (SD), *n*			
General discomfort symptom score	61.1 (25.5), 544	55.4 (25.4), 231	<0.01
Wind and frequency symptom score	51.5 (21.7), 546	48.9 (23.0), 235	0.13
Nausea and vomiting symptom score	83.8 (18.3), 549	77.1 (21.5), 239	<0.01
Activity limitation symptom score	79.9 (16.1), 547	77.5 (17.4), 232	0.06
Constipation and swallowing symptom score	78.8 (20.0), 550	75.2 (21.7), 236	0.03

TABLE 8 Baseline characteristics of responders and non-responders at 5 years (*continued*)

Characteristic	Responder (max. *n* = 558)	Non-responder (max. *n* = 252)	*p*-value (two- sided)
EQ-5D, mean (SD), *n*	0.735 (0.234), 544	0.662 (0.279), 239	<0.01
SF-36 score, mean (SD), *n*			
SF-36 physical	45.2 (9.5), 530	44.0 (9.7), 232	0.10
SF-36 mental	46.3 (11.2), 530	42.7 (12.9), 232	<0.01
Norm-based physical functioning	47.2 (9.9), 545	46.1 (10.7), 239	0.15
Norm-based role physical	46.6 (10.7), 546	45.0 (11.0), 238	0.06
Norm-based bodily pain	45.1 (10.1), 546	42.3 (9.9), 236	<0.01
Norm-based general health	42.0 (9.8), 544	39.3 (10.7), 236	<0.01
Norm-based vitality	44.3 (10.8), 549	42.8 (11.4), 237	0.07
Norm-based social functioning	45.5 (10.8), 542	41.8 (12.0), 237	<0.01
Norm-based role emotional	47.0 (11.5), 543	44.5 (13.2), 239	0.01
Norm-based mental health	47.0 (10.6), 549	42.9 (12.4), 237	<0.01
Any PPI, *n/N* (%)	508/552 (92)	213/242 (88)	0.07
Any reflux drug, *n/N* (%)	530/552 (96)	225/242 (93)	0.07

max., maximum.

TABLE 9 Baseline characteristics of responders at 5 years by randomised allocation

Characteristic	Surgical (max. *n* = 127)	Medical (max. *n* = 119)	*p*-value (two-sided)
BMI (kg/m^2), mean (SD), *n*	29.0 (4.3), 127	27.7 (3.8), 119	0.01
Age (years), mean (SD), *n*	48.5 (9.3), 127	46.4 (11.6), 119	0.12
Sex, *n/N* (%)			
Male	79/127 (62)	76/119 (64)	0.79
Female	48/127 (38)	43/119 (36)	–
Duration of prescribed medication (months), mean (SD), *n*	57.2 (63.4), 124	46.3 (60.1), 117	0.17
Erosive oesophagitis, *n/N* (%)			
Yes	63/111 (57)	68/107 (64)	0.35
No	48/111 (43)	39/107 (36)	–
Helicobacter pylori status, n/*N* (%)			
Positive (subsequently treated)	6/96 (6)	10/91 (11)	0.52
Positive (subsequently untreated)	1/96 (1)	2/91 (2)	–
Negative	55/96 (57)	45/91 (49)	–
Uncertain	34/96 (35)	34/91 (37)	–
Hiatus hernia, *n/N* (%)			
Yes	73/117 (62)	71/114 (62)	0.99
No	44/117 (38)	43/114 (38)	–

continued

TABLE 9 Baseline characteristics of responders at 5 years by randomised allocation (*continued*)

Characteristic	Surgical (max. *n* = 127)	Medical (max. *n* = 119)	*p*-value (two-sided)
Age (years) left full-time education, *n/N* (%)			
≤16	77/125 (62)	70/119 (59)	0.46
17–19	27/125 (22)	31/119 (26)	–
20+	21/125 (17)	18/119 (15)	–
Employment status, *n/N* (%)			
Full-time	86/124 (69)	76/118 (64)	0.77
Part time	13/124 (10)	10/118 (8)	–
Student	2/124 (2)	1/118 (1)	–
Retired	9/124 (7)	13/118 (11)	–
Housework	4/124 (3)	7/118 (6)	–
Seeking work	4/124 (3)	3/118 (3)	–
Other	6/124 (5)	8/118 (7)	–
REFLUX QoL, mean (SD), *n*	65.9 (23.7), 121	68.6 (24.0), 110	0.38
REFLUX symptom score, mean (SD), *n*			
General discomfort symptom score	60.1 (24.1), 123	63.9 (25.2), 115	0.23
Wind and frequency symptom score	48.0 (19.7), 125	48.7 (20.9), 117	0.78
Nausea and vomiting symptom score	82.9 (18.9), 125	84.7 (18.9), 117	0.46
Activity limitation symptom score	79.9 (15.2), 124	79.9 (16.8), 117	0.99
Constipation and swallowing symptom score	78.2 (19.2), 124	75.9 (20.0), 118	0.35
EQ-5D, mean (SD), *n*	0.736 (0.223), 122	0.755 (0.228), 118	0.51
SF-36 score, mean (SD), *n*			
SF-36 physical	44.8 (10.0), 121	46.1 (9.1), 114	0.30
SF-36 mental	46.6 (11.0), 121	46.5 (11.1), 114	0.98
Norm-based physical functioning	46.8 (10.0), 123	48.4 (10.2), 117	0.22
Norm-based role physical	46.9 (10.8), 124	47.0 (10.8), 116	0.96
Norm-based bodily pain	44.6 (10.1), 123	45.7 (10.1), 117	0.39
Norm-based general health	41.4 (9.4), 124	42.4 (10.2), 116	0.41
Norm-based vitality	43.9 (10.4), 125	44.9 (11.2), 117	0.47
Norm-based social functioning	45.4 (10.5), 124	46.4 (10.8), 115	0.45
Norm-based role emotional	47.2 (11.4), 124	46.7 (12.1), 116	0.74
Norm-based mental health	47.3 (10.9), 125	48.0 (10.6), 117	0.60
Any PPI, *n/N* (%)	120/125 (96)	109/118 (92)	0.23
Any reflux drug, *n/N* (%)	124/125 (99)	113/118 (96)	0.08

max., maximum.

TABLE 10 Initial fundoplication operations

Surgery	Randomised participants		Preference participants	
	Surgical ($n = 178$)	Medical ($n = 179$)	Surgical ($n = 261$)	Medical ($n = 192$)
First fundoplication in first year, *n* (%)	111 (62.4)	10 (5.6)	218 (83.5)	3 (1.6)
First fundoplication after first year, *n*	1	14	4	3
In second year	0	1	2	0
In third year	0	7	1	2
In fourth year	1	4	1	1
In fifth year	0	2	0	0
Fundoplication at any time during 5-year follow-up, *n* (%)	112 (62.9)	24 (13.4)	222 (85.1)	6 (3.1)

Details of the surgery received by the 111 participants (62.4%) randomised to surgery and the 218 preference participants (83.5%) who actually received surgery in the first year, the perioperative complications that they experienced and their hospital stay have been reported previously but are summarised in *Appendix 2* for completeness. There were no perioperative deaths.

Table 11 shows the numbers of those who had fundoplication who subsequently had a second reflux-related operation during the 5 years of follow-up. Overall, this applied to 16 participants (4.4%) among the 364 who had a first operation: five (4.5%) in the randomised surgery group; one (4.2%) in the randomised medical group; eight (3.6%) in the preference surgery group; and two (33.3%) in the preference medical group. In total, five of the 16 operations were reconstructions of the same wrap, three were repairs of hiatus hernia only, six were conversions to a different type of wrap and two were reversals of the fundoplication. Two of these 16 participants had a third reflux-related operation; both were in the preference surgery group – one a reconstruction of the same wrap and one a repair of hiatus hernia only.

Late postoperative complications

Table 12 describes late postoperative complications among those participants who had surgery, in each of the study groups and overall. Of the total 364 who had fundoplication, 12 (3.3%) had a late complication: four (1.1%) were oesophageal dilatations/stricture dilatations; three (0.8%) were repairs of incisional hernias; and five (1.4%) were a heterogeneous group of other complications as detailed in the table.

Medication

Figure 2 summarises reported use of any PPI medication in the previous 2 weeks across the follow-up time points of the trial. Full details are provided in the tables in *Appendix 3*. From the time of the first annual follow-up onwards, rates in both medical groups were consistently around 80%. The rates in the randomised surgical ITT group at the first, second and third annual follow-ups were approximately 36–38%, rising to 43% in the fifth year. The extent to which these rates reflected medication taking among those allocated to surgery and who had fundoplication (rather than those who did not have surgery) can be gauged from the randomised surgery PP group: 7.3% (3 months), 12.5% (1 year), 15.1% (2 years), 19.6% (3 years), 23.9% (4 years) and 25.6% (5 years).

Table 13 allows further exploration of the reasons for the rise in medication use in the randomised surgery group. It distinguishes those reporting taking medication at the end of the first year of follow-up from

TABLE 11 Subsequent reflux-related operations among participants who had fundoplication

Surgery	Randomised participants		Preference participants		
	Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)	Total cohort
First fundoplication operation at any time, *n*	112	24	222	6	364
Second reflux-related reoperation, *n* (%)	5 (4.5)	1 (4.2)	8 (3.6)	2 (33.3)	16 (4.4)
Reconstruction of same wrap	2	1	1	1	5
Repair of hiatus hernia only	1	0	2	0	3
Conversion of type of wrap	2	0	4	0	6
Reversal of fundoplication	0	0	1	1	2
Third reflux-related reoperation, *n*					
Reconstruction of same wrap	0	0	1	0	1
Repair of hiatus hernia only	0	0	1	0	1
Conversion of type of wrap	0	0	0	0	0
Reversal of fundoplication	0	0	0	0	0

TABLE 12 Late postoperative complications (>1 month after surgery)

Complication	Randomised participants		Preference participants		
	Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)	Total cohort
First fundoplication operation at any time	112	24	222	6	364
Late postoperative complications (within first year of original operation), *n*					
Oesophageal dilatation/stricture dilatation	0	0	3	0	3
Repair of incisional hernia	0	0	1	0	1
Other (admission for deep-vein thrombosis/ pulmonary embolism)	0	0	1	0	1
Late postoperative complications (within second year following operation), *n*					
Oesophageal dilatation/stricture dilatation	1	0	0	0	1
Repair of incisional hernia	0	0	0	0	0
Other (pain from operation; hole between stomach and liver)	0	0	1	1	2
Late postoperative complications (beyond second year), *n*					
Oesophageal dilatation/stricture dilatation	0	0	0	0	0
Repair of incisional hernia	0	0	2	0	2
Other (pain due to original wrap shifting; bleed in stomach/bowel)	1	0	0	1	2
Total late postoperative complications, *n* (%)	2 (1.8)	0 (0.0)	8 (3.6)	2 (33.3)	12 (3.3)

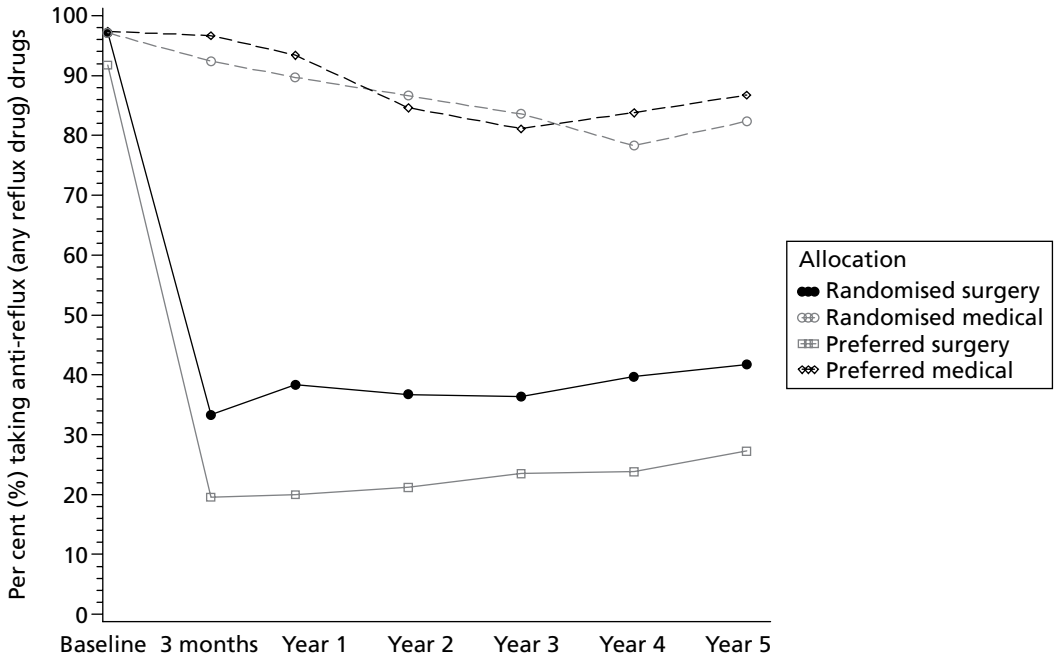


FIGURE 2 Use of PPI medication at baseline and at follow-up points up to 5 years.

those who indicated that they were not taking medication at that time. It shows that around 10–20% of those taking medication at the end of the first year did not report medication use at subsequent annual follow-up. Among those not taking medication at the first annual follow-up in the surgical groups, around 10% rising to around 20% reported medication use at subsequent annual follow-up. This contrasts with the rates in the medical groups, with around 50–60% of those not taking medication at the end of the first year reporting anti-reflux drug use in subsequent annual follow-up. The pattern of type of PPI used changed over the course of the study. Although lansoprazole had been the most commonly used PPI at trial entry, omeprazole use increased over time to become the predominant PPI.

Outcome

Health status

Full details of the health status and QoL measures at each time point of follow-up are in the tables in *Appendix 4*. Details of the statistical testing of the health status and QoL scores can be found in the next section of this chapter.

REFLUX score

Figure 3 summarises changes in the disease-specific REFLUX score over the follow-up period. From this it can be seen that the scores at all time points are highest (indicating fewest symptoms) in the randomised surgical and preference surgical groups. However, the differences between the surgical and medical groups narrow over time. This is due principally to the scores in the randomised medical group improving over the first 3 years and, to a lesser extent, those in the preference medical group improving over the latter end of the follow-up period. The scores for the five components of the measure are summarised graphically in *Figures 4–8*. These show that the overall difference between the groups is principally due to the 'general discomfort' component and, to a lesser extent, the 'nausea and vomiting' and 'activity limitations' components.

TABLE 13 Anti-reflux medication use after the first year

	Randomised participants				Preference participants			
	Surgery		Medical		Surgery		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Known whether or not taking medication at end of first year, *n*	154	104	165	156	232	205	181	178
Group taking anti-reflux drugs at end of the first year								
Taking anti-reflux drugs at end of the first year, *n*	51	10	140	137	42	20	154	152
Taking any anti-reflux drug at end of, *n*/*N* (%)								
Second year	37/42 (88)	7/8 (88)	111/119 (93)	110/116 (95)	27/34 (79)	14/17 (82)	117/130 (90)	116/128 (91)
Third year	34/41 (83)	8/9 (89)	101/115 (88)	101/114 (89)	26/33 (79)	12/14 (86)	117/134 (87)	116/132 (88)
Fourth year	34/41 (83)	7/9 (78)	94/112 (84)	93/110 (85)	20/28 (71)	9/13 (69)	106/119 (89)	105/117 (90)
Fifth year	33/39 (85)	8/9 (89)	89/101 (88)	88/99 (89)	20/29 (69)	10/12 (83)	105/117 (90)	103/115 (90)
Group not taking anti-reflux drugs at end of the first year								
Not taking anti-reflux drugs at end of first year, *n*	103	94	25	19	190	185	27	26
Taking any anti-reflux drug at end of, *n*/*N* (%)								
Second year	10/86 (12)	7/78 (9)	12/23 (52)	11/19 (58)	16/169 (9)	14/165 (8)	15/26 (58)	15/25 (60)
Third year	14/91 (15)	10/83 (12)	11/19 (58)	11/19 (58)	20/163 (12)	20/161 (12)	12/25 (48)	12/24 (50)
Fourth year	16/85 (19)	14/79 (18)	7/17 (41)	7/17 (41)	20/140 (14)	20/139 (14)	13/23 (57)	13/22 (59)
Fifth year	20/88 (23)	16/81 (20)	9/18 (50)	9/17 (53)	28/147 (19)	27/146 (18)	13/19 (68)	12/18 (67)

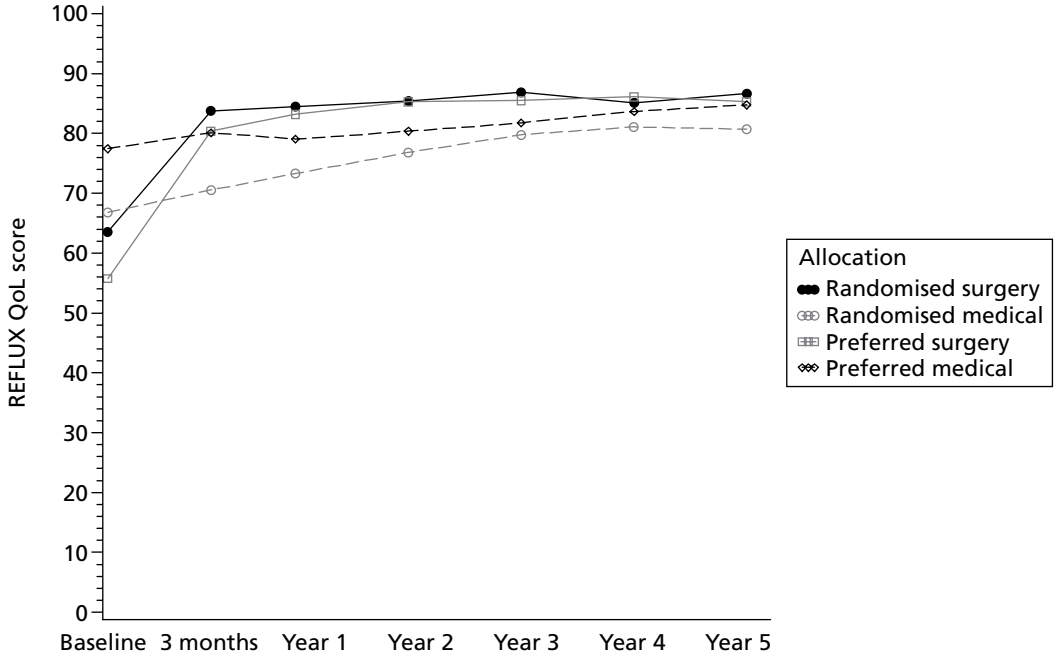


FIGURE 3 Mean REFLUX QoL score at baseline and at follow-up points up to 5 years (score range 0–100; the higher the score, the better the patient felt).

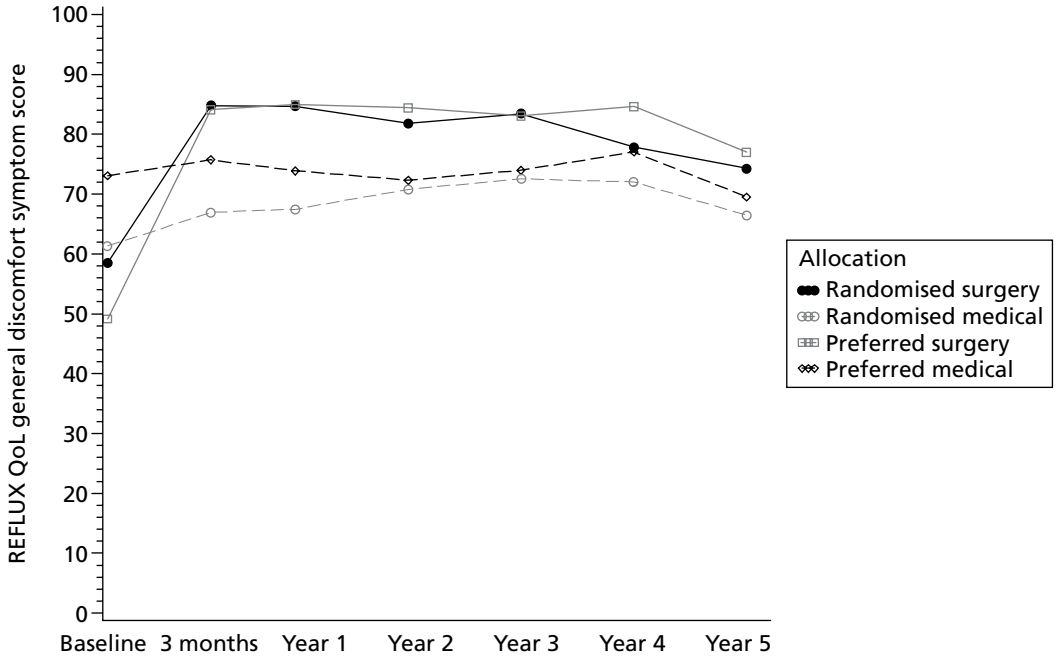


FIGURE 4 Mean REFLUX QoL general discomfort symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

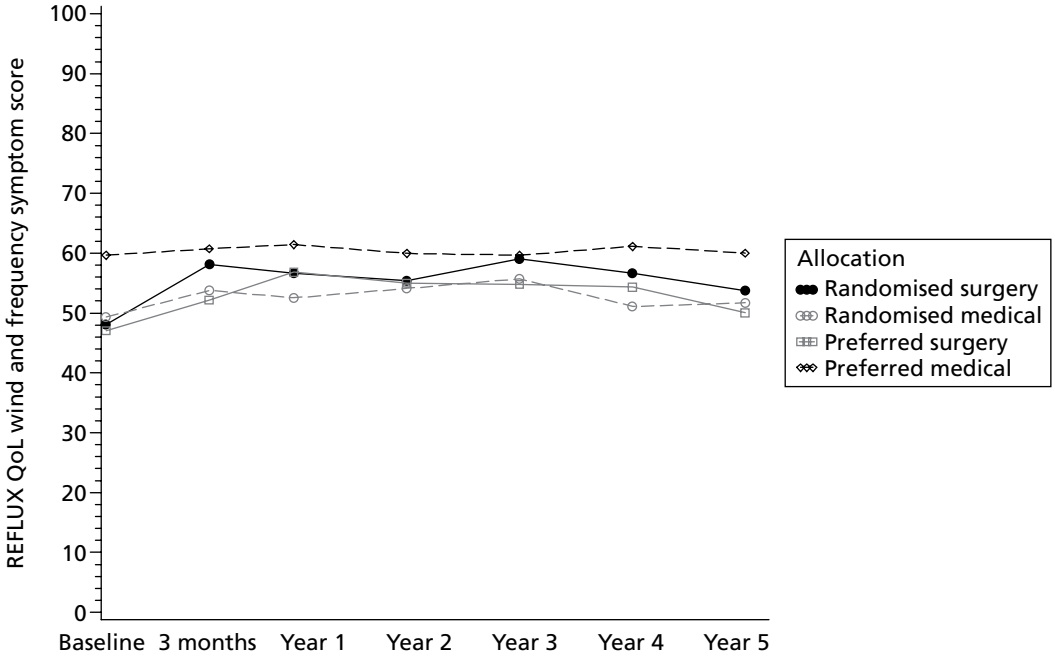


FIGURE 5 Mean REFLUX QoL wind and frequency symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

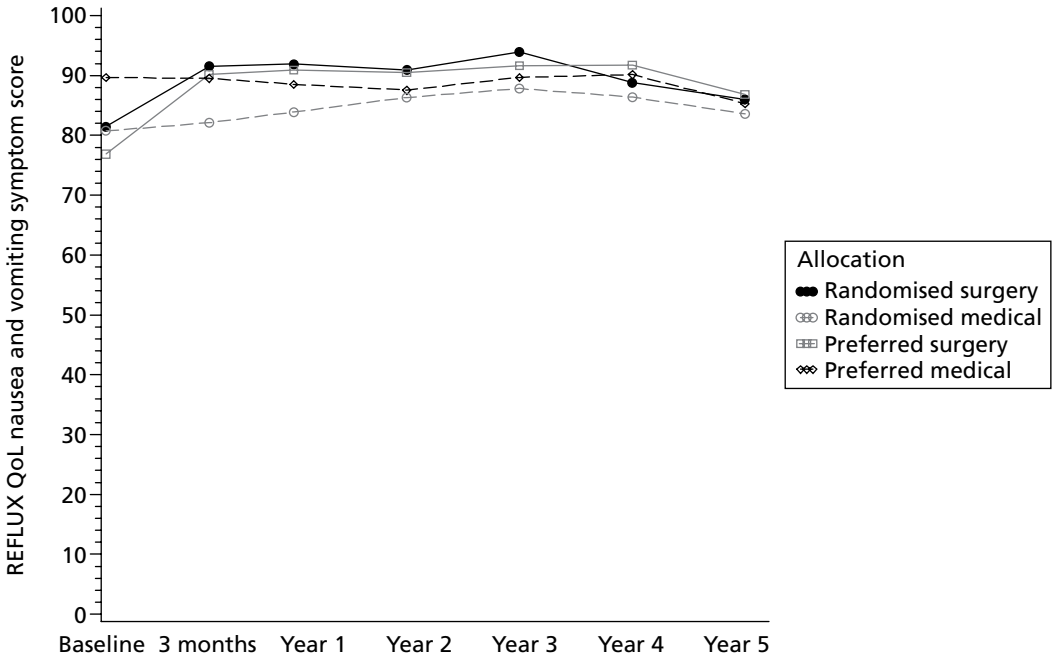


FIGURE 6 Mean REFLUX QoL nausea and vomiting symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

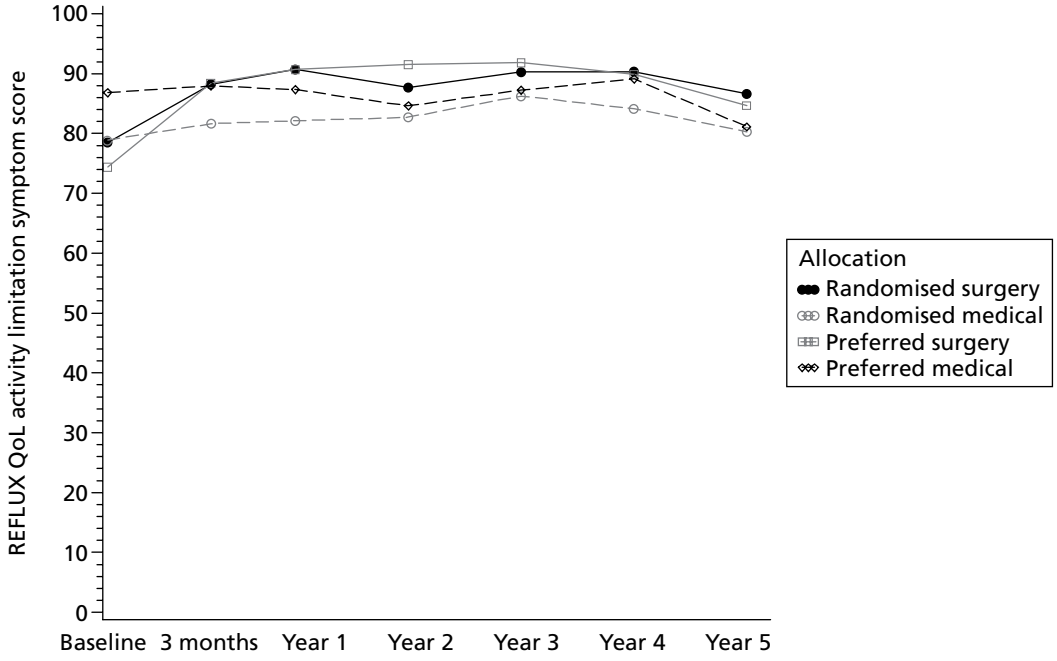


FIGURE 7 Mean REFLUX QoL activity limitation symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

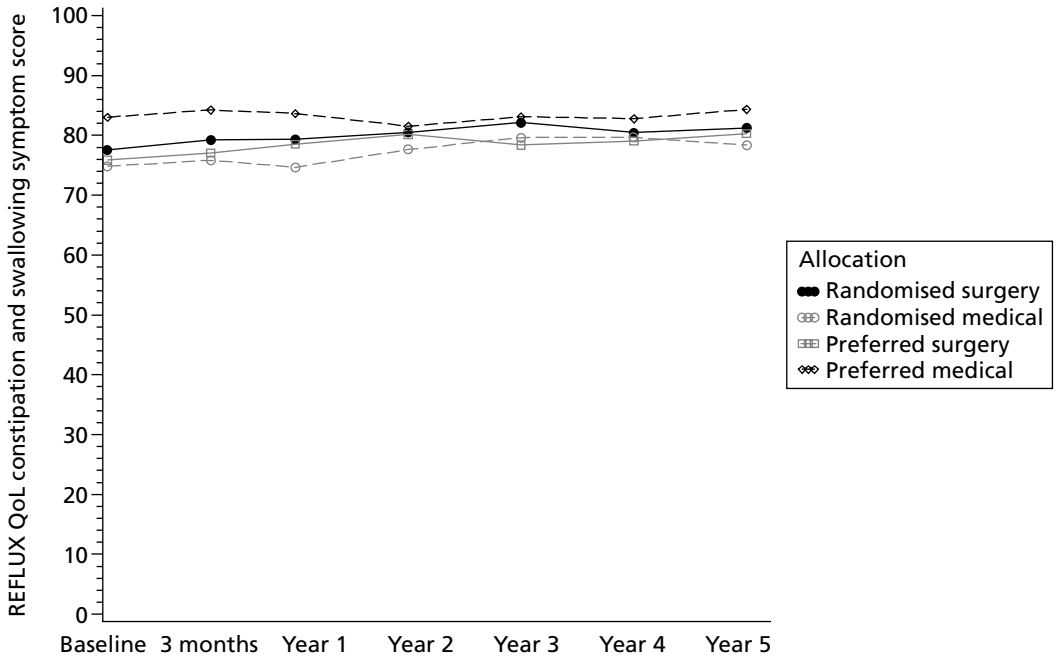


FIGURE 8 Mean REFLUX QoL constipation and swallowing symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

Short Form questionnaire-36 items

The pattern of SF-36 scores, both for the composite physical and mental scores and for the individual dimensions (*Figures 9–16*), was similar to that seen for the REFLUX score, although more compact. Differences narrowed over the 5 years of follow-up, with the 'general health' dimension showing the clearest differences between the surgery and the medical management groups.

European Quality of Life-5 Dimensions

Figure 17 graphically displays the EQ-5D scores over the course of the follow-up period. The pattern is similar to that seen for the REFLUX score although differences are less marked and only clearly seen over the first 3 years.

Use of health services

Table 14 shows use of health services for the randomised groups. The larger number of overnight hospital admissions in the medical group largely reflected admissions for surgery; as described above, 14 participants allocated to medical management had fundoplication after the first year. However, seven participants in the medical group compared with one in the surgical group had admissions for a non-surgery-related reason (data not shown).

Numbers of day-case hospital admissions were similar in the two groups. The larger number of visits to or from a GP for a reflux-related reason in the randomised medical group reflected both more individuals attending their GPs and a higher frequency of visits for those who sought GP care.

Individual symptoms of gastro-oesophageal reflux disease or its treatment

Table 15 shows the frequency with which participants reported symptoms of GORD or its treatment at 3 and 5 years of follow-up for the randomised groups. At both 3 and 5 years, heartburn was reported by a higher proportion of participants in the randomised medical group than in the randomised surgical group. In addition, a higher proportion of participants in the randomised medical group reported more frequent heartburn than in the randomised surgical group. At both time points, a higher proportion of participants

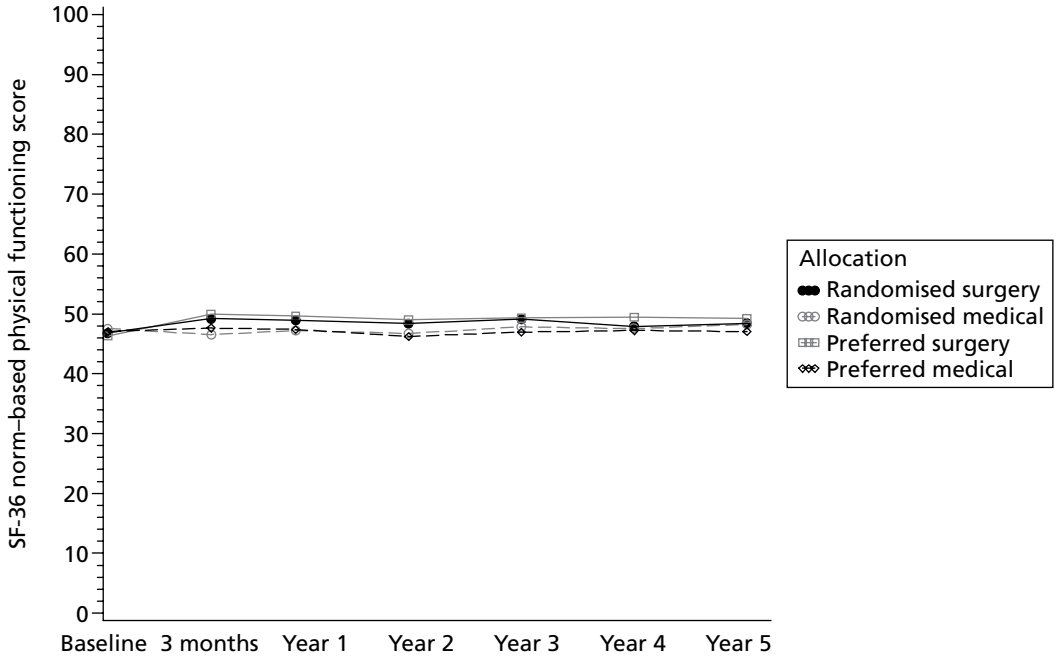


FIGURE 9 Mean SF-36 norm-based physical functioning score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

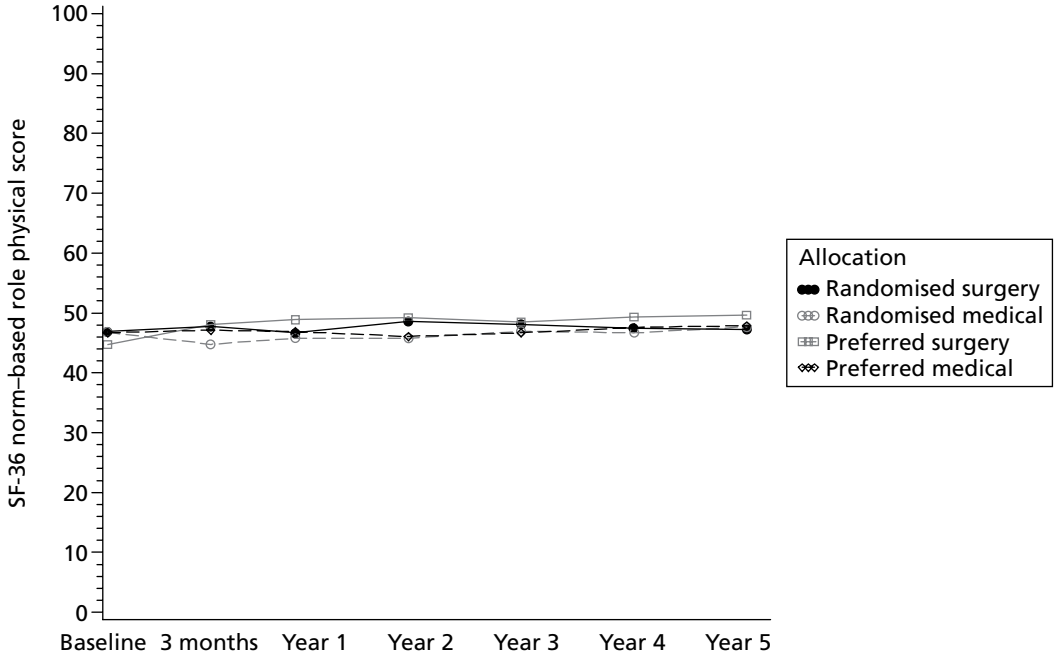


FIGURE 10 Mean SF-36 norm-based role physical score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

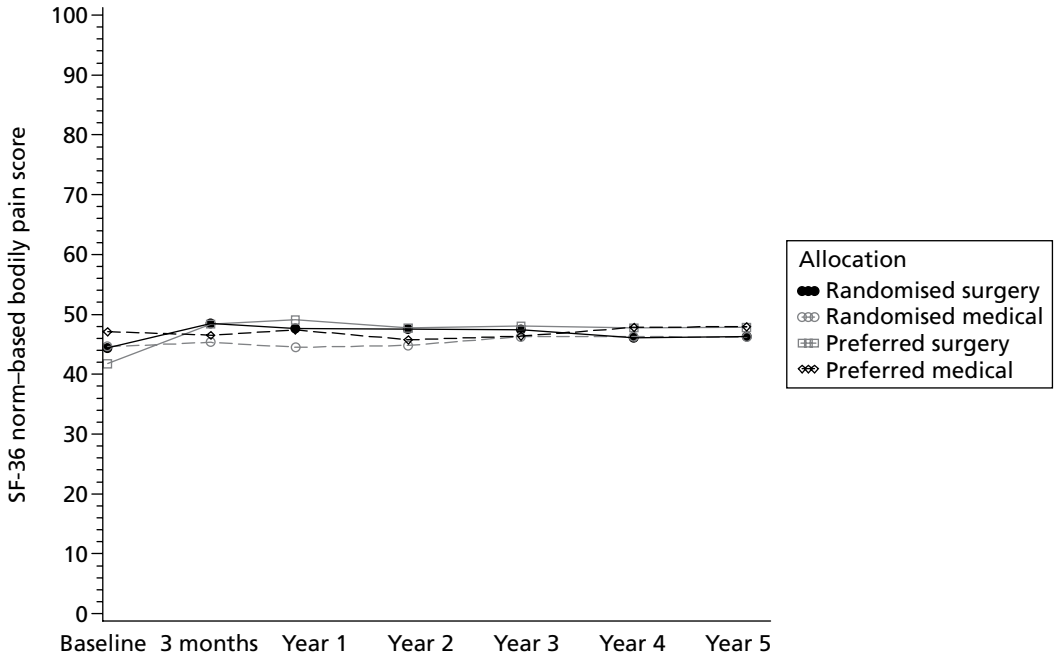


FIGURE 11 Mean SF-36 norm-based bodily pain score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

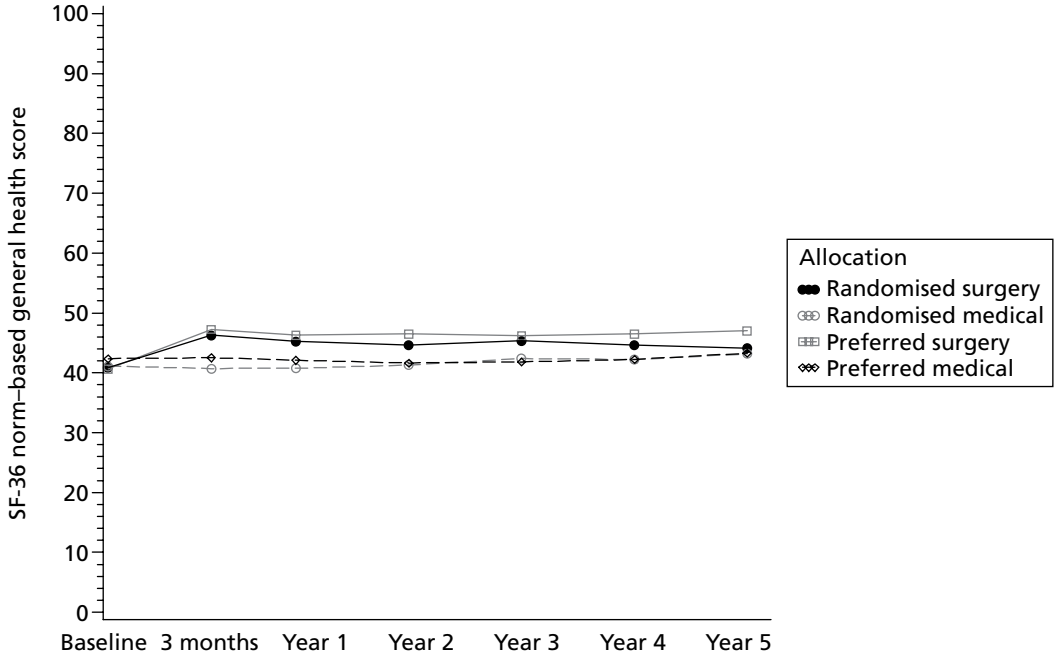


FIGURE 12 Mean SF-36 norm-based general health score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

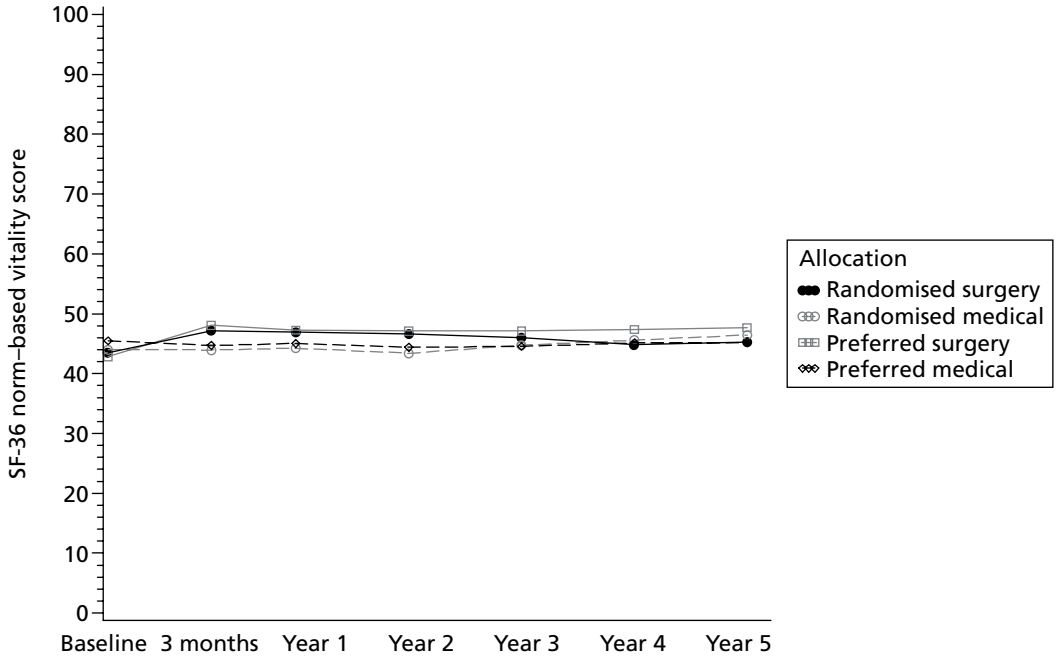


FIGURE 13 Mean SF-36 norm-based vitality score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

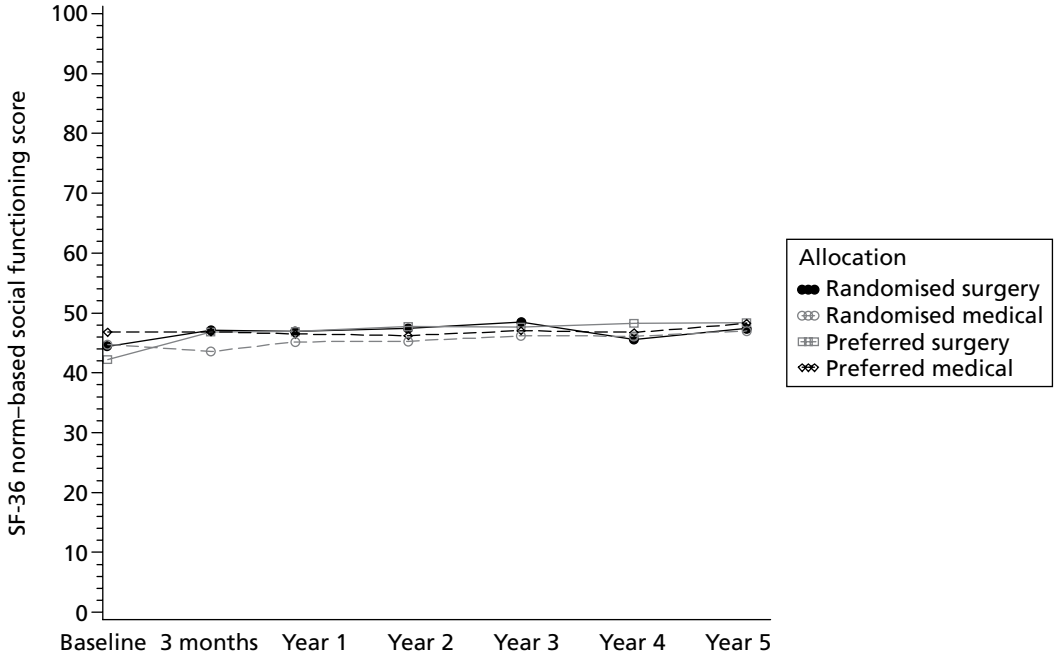


FIGURE 14 Mean SF-36 norm-based social functioning score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

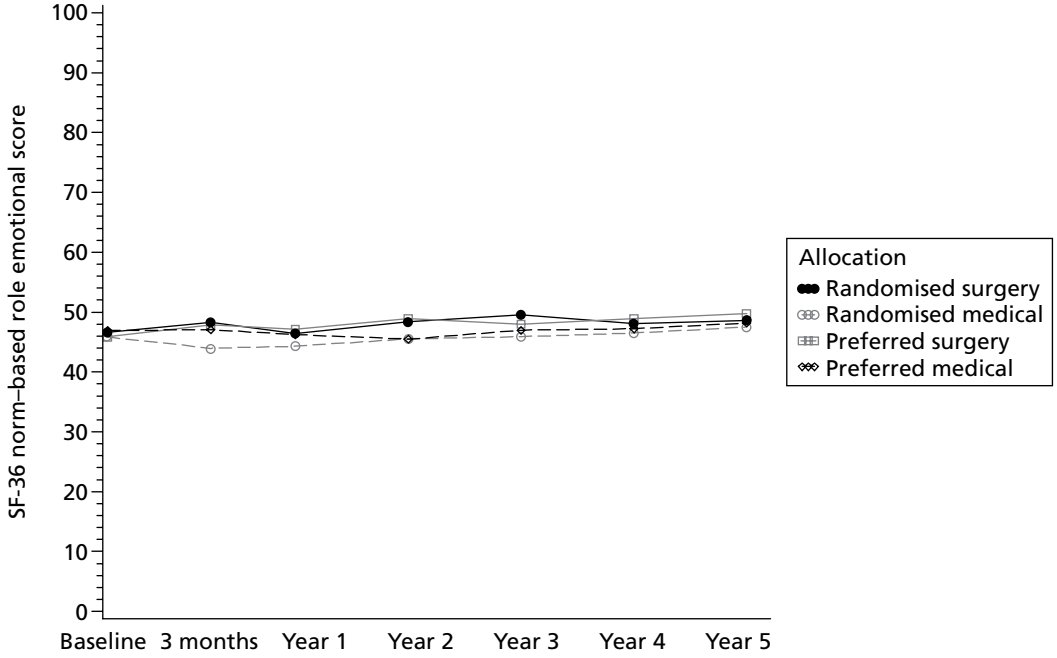


FIGURE 15 Mean SF-36 norm-based role emotional score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

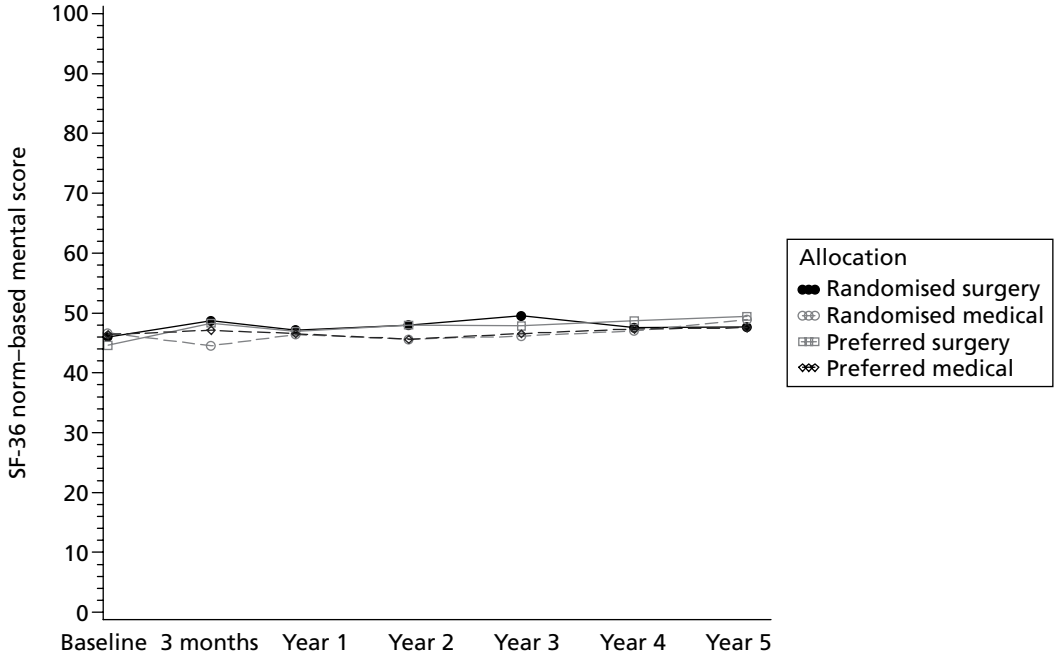


FIGURE 16 SF-36 norm-based mental score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

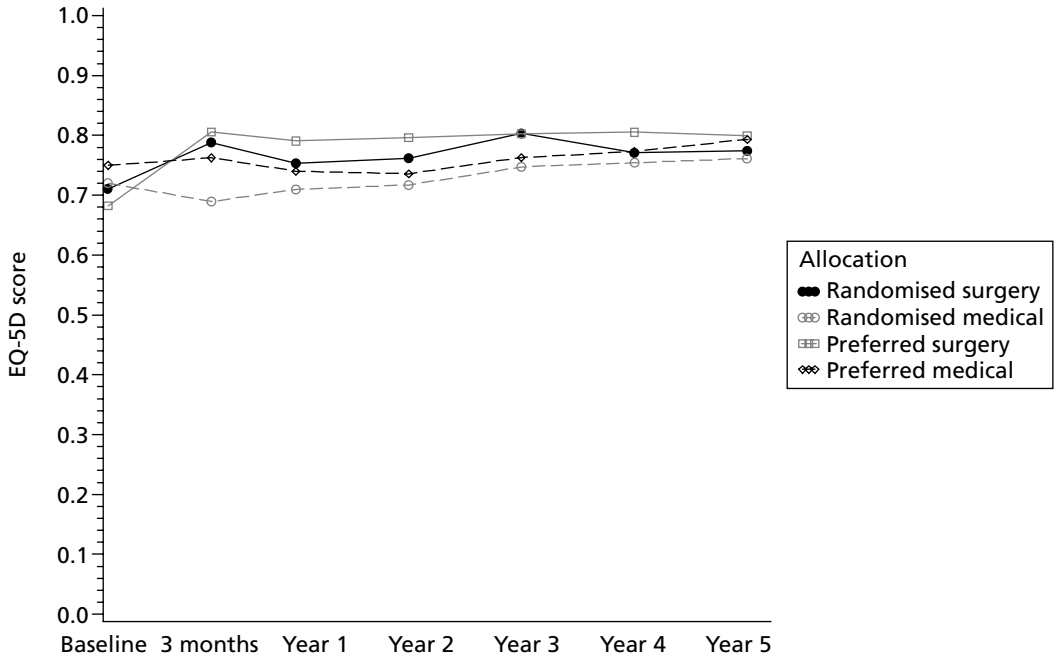


FIGURE 17 EQ-5D at baseline and follow-up points to 5 years.

TABLE 14 Use of health services

Use of health service	Year	Randomised surgical	Randomised medical
Overnight hospital admissions: reflux-related (and all reasons), *n*	1	4 (8)	2 (8)
	2	1 (8)	2 (10)
	3	2 (6)	9 (10)
	4	2 (2)	9 (10)
	5	0 (1)	8 (11)
Day hospital admissions: reflux related (and all reasons), *n*	1	22 (40)	24 (53)
	2	5 (23)	4 (24)
	3	4 (4)	6 (10)
	4	12 (13)	9 (11)
	5	4 (7)	11 (14)
Visits to and from the GP: reflux related (and all reasons), *n*	1	110 (394)	103 (376)
	2	34 (269)	115 (373)
	3	38 (381)	99 (386)
	4	55 (422)	126 (469)
	5	36 (404)	119 (370)

in the randomised medical management group also reported regurgitation symptoms and burping/belching than in the randomised surgical group. At both 3 and 5 years, the proportions who reported no difficulty swallowing and no wind from the lower bowel were similar between the randomised surgical and the randomised medical groups. There was also little difference between the groups at each time point in the proportion of participants who reported a feeling of wanting to be sick but being physically unable to do so.

Statistical analyses

Primary outcome

The pre-chosen primary outcome was the REFLUX QoL score after 5 years of follow-up. The differences between groups with corresponding 95% CIs are shown in *Table 16*. Two types of analysis are presented for the randomised participants – ITT and adjusted treatment received. *Table 16* also displays the impact of including adjustment for baseline score and randomised group*baseline score interaction terms.

Intention to treat

For the ITT analysis there was a mean difference of 6.4 between the groups in favour of surgery when only the minimisation variables were adjusted for (95% CI 1.6 to 11.2; $p = 0.009$). A repeated measures analysis across the 5 years gave a difference of 8.1 (95% CI 4.4 to 11.7). This was not the most parsimonious model – there was strong evidence of an interaction effect between randomised group and baseline REFLUX QoL score (interaction term was –0.23, 95% CI –0.43 to –0.03; $p = 0.023$). This implied that as baseline REFLUX QoL score increased the treatment effect decreased. Estimating the treatment difference at the trial baseline mean REFLUX QoL score of 65.2 resulted in a trial effect size of 8.5 (95% CI 3.9 to 13.1; $p < 0.001$). If the average patient had a lower mean REFLUX QoL score at baseline of 56.0, the effect size increased to 10.6 (95% CI 5.3 to 15.8). If the patient had a higher baseline score of 78.0, the treatment effect size decreased to 5.5 (95% CI 0.6 to 10.4). All results, however, showed strong evidence of increases in REFLUX QoL scores favouring surgery.

TABLE 15 Frequency of GORD symptoms at 3 and 5 years

GORD symptom	3 years		5 years	
	Randomised surgery	Randomised medical	Randomised surgery	Randomised medical
Frequency of heartburn, *n* (%)				
None at all	77 (58.8)	46 (34.8)	65 (58.6)	28 (26.4)
One to three times per week	44 (33.6)	64 (48.5)	38 (34.2)	64 (60.4)
More than three times per week	10 (7.6)	22 (16.7)	8 (7.2)	14 (13.2)
Frequency of regurgitation, *n* (%)				
None at all	102 (77.3)	83 (61.9)	89 (75.4)	71 (63.4)
One to three times per week	27 (20.5)	47 (35.1)	26 (22.0)	37 (33.0)
More than three times per week	3 (2.3)	4 (3.0)	3 (2.5)	4 (3.6)
Frequency of difficulty swallowing, *n* (%)				
None at all	100 (75.8)	102 (76.1)	91 (77.1)	82 (74.5)
One to three times per week	30 (22.7)	27 (20.1)	25 (21.2)	25 (22.7)
More than three times per week	2 (1.5)	5 (3.7)	2 (1.7)	3 (2.7)
Frequency of wind from the bowel, *n* (%)				
None at all	19 (14.4)	20 (15.0)	14 (11.9)	14 (12.7)
One to three times per week	37 (28.0)	35 (26.3)	27 (22.9)	30 (27.3)
More than three times per week	76 (57.6)	78 (58.6)	77 (65.3)	66 (60.0)
Frequency of burping/belching, *n* (%)				
None at all	53 (40.2)	33 (24.8)	46 (39.3)	27 (24.5)
One to three times per week	39 (29.5)	48 (36.1)	40 (34.2)	37 (33.6)
More than three times per week	40 (30.3)	52 (39.1)	31 (26.5)	46 (41.8)
Frequency of wanting to be sick but being physically unable to, *n* (%)				
None at all	116 (87.9)	110 (83.3)	101 (85.6)	92 (82.9)
One to three times per week	15 (11.4)	17 (12.9)	15 (12.7)	16 (14.4)
More than three times per week	1 (0.8)	5 (3.8)	2 (1.7)	3 (2.7)

TABLE 16 Primary outcome: REFLUX QoL scores after 5 years of follow-up

REFLUX QoL score	Randomised participants					
	ITT			Adjusted treatment received		
	Mean difference[a]	95% CI	*p*-value	Mean difference[a]	95% CI	*p*-value
Adjusted for minimisation variables	6.4	1.6 to 11.2	0.009	9.4	1.7 to 17.0	0.017
Adjusted for minimisation variables and baseline REFLUX QoL score	7.6	3.0 to 12.2	0.001	10.6	3.3 to 17.9	0.004
Adjusted for minimisation variables, baseline score and treatment*baseline REFLUX QoL score interaction	8.5	3.9 to 13.1	<0.001	11.5	4.2 to 18.7	0.002

a Difference is surgery group minus medical group.

Adjusted treatment received

The adjusted treatment received analyses attempted to mitigate the effect of non-compliance with the allocated treatment and hence provide an estimate of 'efficacy'.[40] As expected, this approach gave a larger difference, but with wider CIs (9.4, 95% CI 1.7 to 17.0; $p = 0.017$).

Preference groups

The preference for surgery participants reported considerably worse mean REFLUX QoL scores at baseline than the preference for medicine participants (55.8 vs 77.5) (see *Table 7*). Despite starting from a much lower baseline score, at follow-up, the REFLUX QoL score slightly favoured the surgical group using an ITT analysis (difference = 0.61; 95% CI –3.44 to 4.66; $p = 0.767$) and an adjusted treatment received analysis (difference = 0.10; 95% CI –4.77 to 4.97; $p = 0.967$). The differences were not, however, statistically significant.

Secondary outcomes

The secondary outcomes were the health status measures (EQ-5D, SF-36) and REFLUX symptom score at times equivalent to 3 months and then annual follow-up after surgery, and REFLUX QoL (at time points other than 5 years, when it was the primary end point). Analyses of these outcomes are shown in *Tables 17–22*.

REFLUX symptom score

There were statistically significantly higher REFLUX QoL scores at all time points, albeit with some diminution over time in the surgical group (see *Figure 3*). Although symptom category scores favoured surgery across all domains at all time points, the most marked and sustained difference was in 'general discomfort'.

Short Form questionnaire-36 items

The SF-36 scores in all domains also favoured the surgical group at all time points. Differences decreased over time and this was reflected in most *p*-values being <0.05 up to 3 years, whereas at year 5 this applied to only 'norm-based general health' and 'norm-based role emotional'.

European Quality of Life-5 Dimensions

Differences in EQ-5D had a similar pattern to differences in REFLUX QoL and SF-36 scores – differences all favoured the surgical group but tended to narrow such that scores at years 2 and 3 were statistically significantly different, but at later time points they were not. Variability tended to increase over time. Despite the general narrowing of the EQ-5D difference over time, at year 5 it was actually the same as that at 12 months after surgery but with wider CIs.

Adjusted treatment received

As would be expected, all (with a small number of exceptions) the adjusted treatment received analyses had larger differences than the corresponding ITT analyses (around 25–50% higher), but with wider CIs.

Subgroup analyses

Removal of data from the single largest clinical centre (Aberdeen)

No formal exploration of centre effects was undertaken because of the small numbers of participants recruited in many of the clinical centres. However, a sensitivity analysis removing the data from the Aberdeen centre, the centre where the largest number of participants were recruited, did not significantly change the conclusions (adjusted difference in REFLUX score at 60 months = 5.43, 95% CI 0.96 to 9.90).

Partial compared with total wrap procedure

In an observational analysis, there was no evidence of a difference between a total wrap procedure and a partial wrap procedure. The difference in the REFLUX QoL score between these procedures at time equivalent to 5 years post surgery was –1.0 (95% CI –5.4 to 3.7; $p = 0.649$).

TABLE 17 Secondary outcomes at a time equivalent to 3 months after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	15.0	10.5 to 19.4	<0.001	20.7	13.9 to 27.5	<0.001
REFLUX symptom score						
General discomfort symptom score	19.2	14.9 to 23.6	<0.001	26.0	19.6 to 32.4	<0.001
Wind and frequency symptom score	4.6	0.5 to 8.6	0.027	5.1	–1.0 to 11.3	0.101
Nausea and vomiting symptom score	8.8	5.8 to 11.9	<0.001	12.4	7.7 to 17.1	<0.001
Activity limitation symptom score	7.1	3.2 to 11.0	<0.001	9.1	3.2 to 15.1	0.003
Constipation and swallowing symptom score	2.0	–1.9 to 6.0	0.318	2.1	–3.9 to 8.2	0.486
SF-36 score						
Norm-based physical functioning	3.1[b]	1.3 to 4.9	0.001	4.4[b]	1.5 to 7.2	0.003
Norm-based role physical	2.7	0.5 to 4.9	0.018	3.4	–0.04 to 6.8	0.053
Norm-based bodily pain	3.2[b]	1.1 to 5.3	0.003	4.1[b]	0.9 to 7.2	0.012
Norm-based general health	5.8[b]	3.8 to 7.8	<0.001	7.8[b]	4.8 to 10.7	<0.001
Norm-based vitality	3.0	0.9 to 5.1	0.006	3.9	0.7 to 7.1	0.018
Norm-based social functioning	3.6	1.3 to 5.8	0.002	4.6	1.1 to 8.1	0.010
Norm-based role emotional	3.3	0.7 to 5.8	0.012	4.1	0.2 to 8.0	0.042
Norm-based mental health	4.2[b]	2.1 to 6.2	<0.001	5.5[b]	2.4 to 8.6	0.001
EQ-5D, mean (SD)	0.099[b]	0.048 to 0.150	<0.001	0.129[b]	0.051 to 0.207	0.001

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 18 Secondary outcomes at a time equivalent to 12 months after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	14.0	9.6 to 18.4	<0.001	19.4	13.0 to 25.8	<0.001
REFLUX symptom score						
General discomfort symptom score	18.3	13.8 to 22.9	<0.001	26.1	19.6 to 32.5	<0.001
Wind and frequency symptom score	4.9	0.8 to 9.1	0.019	6.7	0.6 to 12.8	0.033
Nausea and vomiting symptom score	7.8	4.6 to 10.9	<0.001	11.5	7.0 to 16.0	<0.001
Activity limitation symptom score	8.4	5.2 to 11.7	<0.001	12.0	7.3 to 16.7	<0.001
Constipation and swallowing symptom score	3.5	–0.5 to 7.5	0.085	5.0	–0.9 to 10.9	0.099
SF-36 score						
Norm-based physical functioning	2.3[b]	0.6 to 4.0	0.007	3.4[b]	0.9 to 5.9	0.008
Norm-based role physical	0.9	–1.1 to 3.0	0.383	1.2	–1.8 to 4.3	0.434
Norm-based bodily pain	3.4[b]	1.4 to 5.5	0.001	5.1[b]	2.1 to 8.0	0.001
Norm-based general health	4.8[b]	2.7 to 6.8	<0.001	7.0[b]	4.0 to 10.0	<0.001
Norm-based vitality	2.5	0.4 to 4.6	0.018	3.7	0.6 to 6.8	0.019
Norm-based social functioning	2.3	0.1 to 4.5	0.040	3.3	0.04 to 6.6	0.047
Norm-based role emotional	1.8	–0.8 to 4.4	0.177	2.7	–1.1 to 6.5	0.168
Norm-based mental health	1.0[b]	–1.0 to 3.1	0.312	1.5[b]	–1.5 to 4.5	0.324
EQ-5D	0.047[b]	–0.004 to 0.097	0.07	0.068[b]	–0.006 to 0.142	0.072

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 19 Secondary outcomes at a time equivalent to 2 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	11.4	6.8 to 16.0	<0.001	15.7	8.5 to 22.9	<0.001
REFLUX symptom score						
General discomfort symptom score	13.08	7.99 to 18.17	<0.001	17.66	9.82 to 25.50	<0.001
Wind and frequency symptom score	3.74[b]	–1.06 to 8.53	0.126	5.67[b]	–2.05 to 13.38	0.149
Nausea and vomiting symptom score	6.34	2.85 to 9.83	<0.001	9.48	4.04 to 14.92	0.001
Activity limitation symptom score	7.02	3.38 to 10.65	<0.001	10.03	4.25 to 15.80	0.001
Constipation and swallowing symptom score	3.29[b]	–1.11 to 7.68	0.142	4.98[b]	–2.09 to 12.05	0.167
SF-36 score						
Norm-based physical functioning	2.73	0.83 to 4.63	0.005	4.27	1.21 to 7.34	0.007
Norm-based role physical	3.11	0.99 to 5.22	0.004	4.69	1.27 to 8.10	0.007
Norm-based bodily pain	3.64	1.51 to 5.77	0.001	5.46	2.04 to 8.88	0.002
Norm-based general health	4.13	1.91 to 6.35	<0.001	5.96	2.39 to 9.54	0.001
Norm-based vitality	3.48	1.20 to 5.76	0.003	5.38	1.66 to 9.09	0.005
Norm-based social functioning	2.74	0.30 to 5.19	0.028	3.79[b]	–0.14 to 7.72	0.059
Norm-based role emotional	2.03[b]	–0.80 to 4.85	0.159	3.06[b]	–1.49 to 7.61	0.187
Norm-based mental health	2.33	0.08 to 4.59	0.043	3.86	0.22 to 7.49	0.038
EQ-5D	0.068[b]	0.005 to 0.131	0.036	0.098[b]	–0.003 to 0.199	0.057

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 20 Secondary outcomes at a time equivalent to 3 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	9.0	4.9 to 13.1	<0.001	12.9	6.3 to 19.5	<0.001
REFLUX symptom score						
General discomfort symptom score	11.86	6.84 to 16.88	<0.001	16.25	8.37 to 24.14	<0.001
Wind and frequency symptom score	4.98[b]	–0.26 to 10.22	0.063	15.95[b]	8.03 to 23.87	<0.001
Nausea and vomiting symptom score	6.69	3.65 to 9.73	<0.001	9.71	4.98 to 14.44	<0.001
Activity limitation symptom score	4.61	0.99 to 8.22	0.013	6.37	0.58 to 12.15	0.031
Constipation and swallowing symptom score	2.62[b]	–1.51 to 6.76	0.212	6.51[b]	0.73 to 12.29	0.027
SF-36 score						
Norm-based physical functioning	2.61	0.56 to 4.67	0.013	3.83	0.52 to 7.14	0.023
Norm-based role physical	1.82[b]	–0.43 to 4.07	0.113	3.82[b]	0.52 to 7.12	0.024
Norm-based bodily pain	2.33	0.24 to 4.42	0.029	3.74	0.36 to 7.12	0.030
Norm-based general health	3.69	1.50 to 5.87	0.001	5.21	1.70 to 8.73	0.004
Norm-based vitality	2.29[b]	–0.23 to 4.81	0.075	5.29[b]	1.77 to 8.81	0.003
Norm-based social functioning	3.27	0.87 to 5.68	0.008	4.81	0.93 to 8.69	0.015
Norm-based role emotional	4.03	1.50 to 6.57	0.002	6.89	2.77 to 11.01	0.001
Norm-based mental health	4.60	2.29 to 6.91	<0.001	7.39	3.65 to 11.14	<0.001
EQ-5D, mean (SD)	0.070[b]	0.015 to 0.126	0.013	0.108	0.016 to 0.201	0.022

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 21 Secondary outcomes at a time equivalent to 4 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	8.3	3.2 to 13.4	0.001	11.6	3.5 to 19.8	0.005
REFLUX symptom score						
General discomfort symptom score	8.81	3.49 to 14.13	0.001	11.48	3.11 to 19.84	0.007
Wind and frequency symptom score	5.98	0.70 to 11.26	0.027	9.55	0.95 to 18.14	0.030
Nausea and vomiting symptom score	2.93[b]	–1.00 to 6.86	0.143	3.25[b]	–3.01 to 9.51	0.307
Activity limitation symptom score	4.38	0.64 to 8.12	0.022	5.95[b]	–0.03 to 11.93	0.051
Constipation and swallowing symptom score	0.26[b]	–4.21 to 4.74	0.908	0.54[b]	–6.72 to 7.80	0.884
SF-36 score						
Norm-based physical functioning	2.14	0.00 to 4.28	0.050	3.10[b]	–0.36 to 6.55	0.079
Norm-based role physical	1.36[b]	–1.23 to 3.96	0.302	2.42[b]	–1.79 to 6.62	0.259
Norm-based bodily pain	1.72[b]	–0.57 to 4.02	0.140	2.59[b]	–1.13 to 6.31	0.172
Norm-based general health	4.02	1.61 to 6.44	0.001	5.74	1.84 to 9.63	0.004
Norm-based vitality	0.17[b]	–2.25 to 2.60	0.888	0.28[b]	–3.66 to 4.22	0.890
Norm-based social functioning	1.26[b]	–1.60 to 4.12	0.387	1.92[b]	–2.72 to 6.56	0.416
Norm-based role emotional	1.79[b]	–1.28 to 4.85	0.253	2.77[b]	–2.21 to 7.75	0.274
Norm-based mental health	1.55[b]	–1.03 to 4.12	0.238	1.85[b]	–2.31 to 6.00	0.382
EQ-5D	0.036[b]	–0.020 to 0.091	0.212	0.052[b]	–0.039 to 0.142	0.265

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 22 Secondary outcomes at a time equivalent to 5 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX symptom score						
General discomfort symptom score	11.82	6.50 to 17.14	<0.001	15.59	7.52 to 23.66	<0.001
Wind and frequency symptom score	3.34[b]	–1.98 to 8.66	0.218	5.12[b]	–3.50 to 13.73	0.243
Nausea and vomiting symptom score	4.97	1.53 to 8.41	0.005	7.32	2.04 to 12.60	0.007
Activity limitation symptom score	5.97	2.03 to 9.91	0.003	8.27	2.03 to 14.52	0.010
Constipation and swallowing symptom score	2.54[b]	–2.09 to 7.18	0.281	4.11[b]	–3.40 to 11.62	0.282
SF-36 score						
Norm-based physical functioning	2.01[b]	–0.26 to 4.28	0.082	3.35[b]	–0.33 to 7.03	0.074
Norm-based role physical	0.57[b]	–2.10 to 3.24	0.674	1.14[b]	–3.20 to 5.47	0.606
Norm-based bodily pain	1.52[b]	–0.90 to 3.94	0.218	1.65[b]	–2.25 to 5.54	0.406
Norm-based general health	2.76	0.21 to 5.31	0.034	3.79[b]	–0.29 to 7.88	0.068
Norm-based vitality	0.37[b]	–2.23 to 2.98	0.777	0.19[b]	–4.03 to 4.41	0.928
Norm-based social functioning	1.72[b]	–1.05 to 4.49	0.221	2.36[b]	–2.13 to 6.84	0.301
Norm-based role emotional	2.67	0.07 to 5.27	0.044	4.56	0.34 to 8.79	0.034
Norm-based mental health	0.59[b]	–1.96 to 3.14	0.650	0.40[b]	–3.72 to 4.51	0.849
EQ-5D	0.047[b]	–0.013 to 0.108	0.126	0.069[b]	–0.029 to 0.167	0.168

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

Discussion

Follow-up to 5 years after laparoscopic surgery described here provides clear evidence of sustained improvement in GORD symptoms, as judged by the REFLUX QoL scores. Differences between the groups as randomised did tend to diminish over the course of the study; nevertheless, the analyses at 5 years (the primary end point) showed highly statistically significant results with effect sizes of the order of 0.6 of a SD.

This report concentrates on the data collected annually at a time equivalent to between 2 and 5 years post surgery. Data were collected through self-complete postal questionnaires, backed up by postal and telephone reminders and occasional completion of the questionnaire over the telephone. The response rate did drop over time, from 90% at 1 year to around 70% at 5 years. The principal reason for not obtaining a follow-up questionnaire was a loss of contact, such as following a home move; the second most common reason was a decision by a participant to decline further follow-up. The category of 'non-responder' accounted for only around 8% of those without a follow-up questionnaire. Response analysis showed that responders at 5-year follow-up had a higher mean age, had been prescribed anti-reflux medication for a shorter period of time at recruitment and had higher QoL at baseline. However, the characteristics of responders and non-responders at 5 years were similar across the two randomised groups.

Randomised trials, such as the REFLUX trial, that compare surgery with medical management are challenging to mount because of the stark contrast between the treatments compared. As described in the previous report of this study, recruitment was not easy and it is to the credit of the many staff in the 21 centres involved in the trial that this was accomplished successfully. A second challenge was that, after randomisation, a sizable proportion of participants did not receive the treatment to which they had been allocated – again, reflecting the contrasts in the treatments. We explored the impact of this in a number of ways.

Figure 18 shows the results of a supplementary analysis of the group randomly allocated surgery stratified by whether or not they actually had surgery. It shows that those who had surgery started from a lower REFLUX QoL baseline score (had worse symptoms) than those who did not undergo surgery, and then had a sharp rise in score following the operation such that their scores were consistently higher than those who did not actually have fundoplication. To put this another way, the improvement seen among those who had surgery was greater than that in the randomised group overall.

Figure 19 shows a similar supplementary analysis of the group allocated medical management stratified by whether or not they in fact had surgery in the first year. This shows that those who had fundoplication (the lowest line) had more severe symptoms of GORD (low REFLUX QoL scores) at the time of trial entry, worse even than the preference surgical group. In contrast, those solely managed medically had relatively high baseline scores. Scores among those randomised to medical management who had surgery improved markedly over the course of the follow-up, such that by years 4 and 5 the scores in the two strata were similar. This indicates that much of the narrowing of the scores in the ITT groups over the 5 years can be explained by surgery in the randomised medical group.

We assessed more formally the extent to which surgery in the randomised medical management group might have affected the results by undertaking adjusted treatment received analyses. We decided to base these on treatment status at the first year follow-up point. We chose this partly to be consistent with our previous report of the results up to 1 year and partly because we considered that those who had surgery after that time point were likely to be highly selected. To put this another way, we were concerned that a PP analysis up to 5 years would be particularly prone to bias. The adjusted treatment received analyses, as expected, indicated larger effects of surgery – with differences in score around 25–50% higher. As illustrated by the preference groups in this study, the proportion of those recommended surgery and willing to have it who subsequently go on to have fundoplication is likely to be higher in everyday practice. Hence, we would argue that the results of the adjusted treatment received analyses are likely to

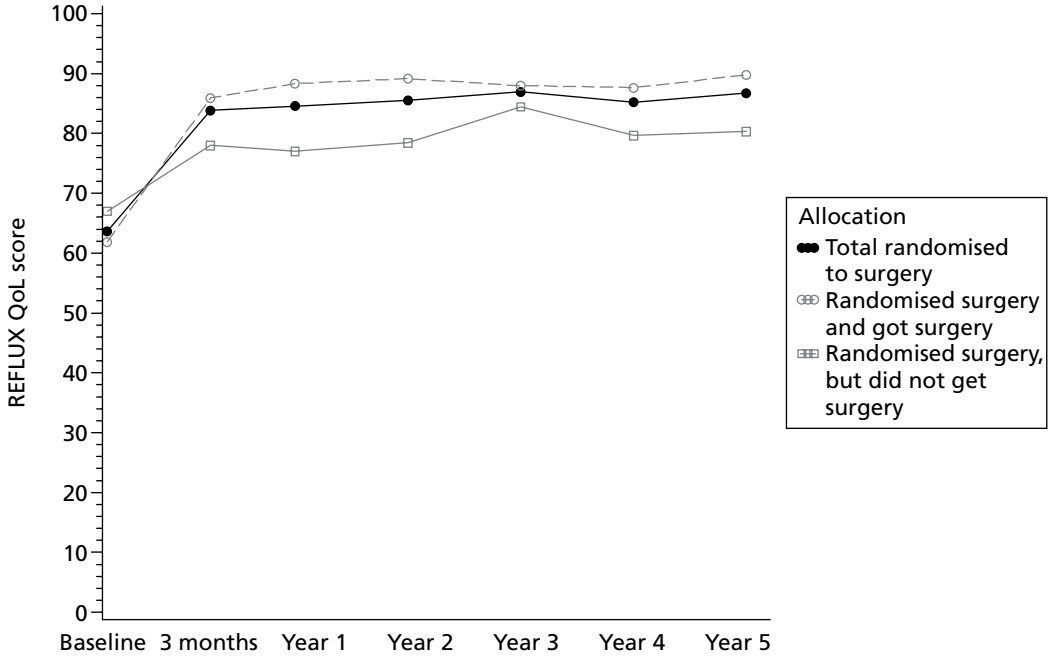


FIGURE 18 Mean REFLUX QoL scores for (a) all randomised to surgery, (b) those randomised to surgery who had fundoplication and (c) those randomised to surgery who did not have surgery.

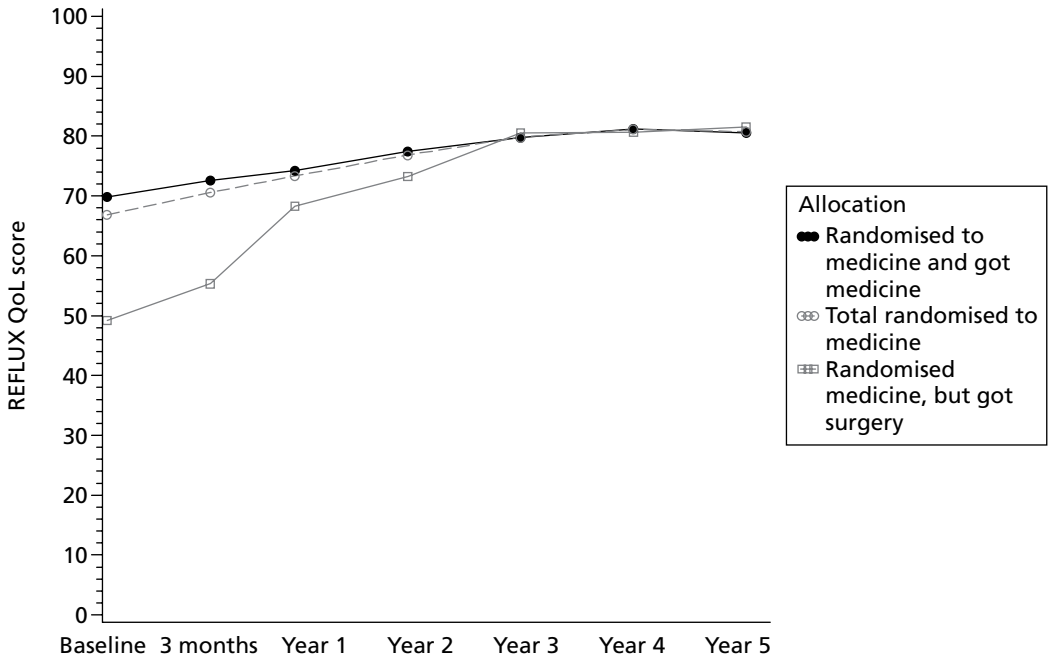


FIGURE 19 Mean REFLUX QoL scores for (a) all randomised to medical management, (b) those randomised to medical management who did not have surgery and (c) those randomised to medical management who had fundoplication.

provide a better estimate of the benefits of a policy of laparoscopic fundoplication as would apply in the health service.

The principal concern about laparoscopic fundoplication is possible risks associated with the surgery. We described intra- and postoperative surgical outcomes in our previous report.[1] Among the 329 patients in the randomised surgical and preference surgical groups who had fundoplication in the first year,

there were no major surgical complications. Two patients (0.6%; 95% CI 0.1% to 2.2%) required conversion to an open procedure; eight (2.4%; 95% CI 1.2% to 4.7%) had a visceral injury; and one (0.3%; 95% CI <0.1% to 1.7%) had a blood transfusion. Three were admitted to a high-dependency unit, but none to an intensive care unit. The 5-year follow-up provides information about longer-term risks. We are aware of seven deaths among trial participants; however, none has an apparent link to the trial. Twelve (3.3%) of the total of 364 participants who had a fundoplication had a late complication: four were oesophageal dilatations/stricture dilatations, three had repairs of incisional hernias and five were a heterogeneous group of other complications (see *Table 12*). Sixteen (4.4%) of those who had fundoplication required further surgery (see *Table 11*): five reconstruction of the same wrap, six conversion to another type of wrap, three repair of hiatus hernia only and two reversal of fundoplication. These, albeit uncommon, complications need to be taken into account when surgery is being considered.

Proton pump inhibitor use in the randomised medical group was consistently around 80%, although these participants were not always the same people at each follow-up. In our questionnaire, we chose to ask about anti-reflux drug use over the preceding 2 weeks as we thought that a recollection over a longer period would be unreliable. Nevertheless, taking of PPIs seems to be dynamic (patients stopping and restarting) and rates of use at any time over a longer period would likely have been higher. We did observe more visits to GPs in the medical groups for reflux-related reasons during the 5 years of follow-up but are not able to say whether this was due to routine reassessments or because symptom control was less stable or inadequately controlled in the medical group.

The pattern of PPIs used did change over the course of the study. At baseline, the commonest PPI was lansoprazole, but omeprazole superseded this over the course of the trial. Much of this change occurred in the first year and hence could be a consequence of the review of medical management that was part of the trial management for those randomised to medical management.

The larger number of overnight hospital admissions in the randomised medical management group was largely, but not totally, explained by the minority who went on to have surgery; as discussed in *Chapter 5* describing the economic evaluation, this was the principal driver of extra resource use by the medical group during the longer-term follow-up.

Despite the methodological challenges alluded to above, the study, through the data presented here, has successfully addressed the first of the objectives of this longer-term follow-up: to assess whether or not short-term clinical benefits, principally in terms of symptom control, are sustained – they are, albeit attenuated. In the next chapter we consider the REFLUX trial in the context of the three other randomised trials that have been conducted worldwide comparing laparoscopic fundoplication with medical management, and assess whether or not the results of the REFLUX trial are consistent with those of the other trials.

Chapter 4 Comparison of the REFLUX trial with other randomised trials of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease

Introduction

The REFLUX trial is one of four randomised trials that have compared laparoscopic surgery with medical management of GORD. Although the REFLUX trial has similarities to the other trials, its design is the most pragmatic[42] and this is reflected in significant differences in comparison with the other trials. The characteristics of the four trials are summarised in some detail in *Appendix 5*; key similarities and differences in characteristics between the REFLUX trial and the other trials will be highlighted here. This overview draws heavily on the relevant Cochrane review,[43] two of whose authors are authors of this report, but incorporates reports published since the Cochrane review, identified primarily through an updated search using a similar strategy to the one described in the Cochrane review.

The three comparable trials

The Anvari *et al.* trial[44–46] is a publicly funded single-centre trial conducted in Canada, led by upper gastrointestinal surgeons. It is the smallest of the four trials (104 randomised). The two intervention policies were standardised and the surgery was undertaken by only four surgeons (*Table 23*). Reflecting this, nearly all participants – unlike in the REFLUX trial – were managed in the way allocated. Like the REFLUX trial, its primary outcome was a GORD-related QoL instrument (the GERSS or Gastro-Esophageal Reflux Symptom Score), and HRQoL was measured with the same instruments as in REFLUX (SF-36 and EQ-5D). The first report described the trial up to 12 months after surgery,[44] and recent papers have reported 3-year results[45] and an economic evaluation.[46] At 3 years, participants in the medical group were offered surgery and a large proportion (42%) accepted; hence, although further follow-up is reported to be ongoing, it will be of limited usefulness in comparing laparoscopic surgery with medical management.

The LOng-Term Usage of esomeprazole versus Surgery for treatment of chronic GERD (LOTUS) trial[47–50] is the largest of the four trials (554 randomised). The study was funded by a pharmaceutical company, AstraZeneca, and the reports all include authors based in the company. The trial involved 39 centres in 11 European countries and was led by an upper gastrointestinal surgeon. The trial is described as 'not designed as a superiority or equivalence trial but, rather, was an exploratory study to estimate the efficacy of laparoscopic anti-reflux surgery and PPI treatment in PPI responders'. Unlike in the REFLUX trial, all participants had shown response to PPI treatment in a run-in phase, and both clinical management policies were strictly standardised (see *Table 23*).

The method by which the total fundoplication approach was standardised has been described in detail.[50] In the medically managed group, the only PPI used was esomeprazole, initially at the standard dose of 20 mg. Both the surgical and medically treated patients were followed up by the investigators at 6-monthly intervals and symptoms were assessed using the Gastrointestinal Symptoms Rating Scale (GSRS) questionnaire. In the medically treated group, esomeprazole could be increased to 40 mg once a day and then to 20 mg twice a day if symptom control was insufficient. Another key difference from the REFLUX trial was that the primary outcome measure was 'treatment failure'. A single definition of treatment failure could not be used for both trial groups; rather, this was specifically defined for each group (including

TABLE 23 Surgical procedure/experience in the four trials[a]

Trial	Surgeon experience	Crural repair	Gastric division	No. of surgeons participating
Anvari[44–46]	>50 procedures performed	Not reported	Short vessels divided	4
LOTUS[47–50]	>40 procedures performed and current workload ≥20 per annum	Protocol specified posterior repair	Protocol specified division	40 trained
Mahon[51–53]	'Experienced'	Yes, all patients	Short vessels divided	2
REFLUX[1–3]	>50 procedures performed	Surgeon discretion	Surgeon discretion	Not reported

a Adapted from Wileman *et al.*[43]

in the medical group need for escalation of medication and in the surgical group, need for regular medication). The concern is that the thresholds for these may not reflect similar levels of GORD. A GORD-specific QoL instrument (Quality of Life in Reflux and Dyspepsia or QOLRAD[54]) was among the secondary outcomes but was given relatively little emphasis in the reporting of the trial. No HRQoL instruments were used and there was no economic evaluation. Although the main analysis was said to be carried out on an ITT basis, it seems that the 40 people allocated surgery who did not receive it were excluded from analyses. Results were first reported after 3 years' follow-up[47] and recently 5-year data have been published.[48]

The Mahon *et al.* trial[51–53] was a two-centre UK trial led by and involving two upper gastrointestinal surgeons. It is not clear how the main trial was funded but supplementary funds were provided by Jansen Pharmaceuticals 'for physiological studies' and by Ethicon Endo-Surgery for the economic analysis.[52] In total, 217 people were randomised; the sequence was 'computerised' but the randomisation process and extent of concealment were not described. The two surgeons used a similar Nissen fundoplication method (see *Table 23*) and there was the option of four different PPI regimens depending on what PPI a participant had been taking prior to the trial. A range of outcome measures were reported and these included a gastrointestinal symptom score (GSRS) and a HRQoL measure [Psychological General Well-Being Index (PGWI)].[55] All those allocated to medical management were offered surgery after 1 year (and apparently this was made clear to potential participants before trial entry) and the majority [54/94 (57%)] then had surgery. The 1-year follow-up was thus essentially the end of this randomised trial, even though a further follow-up has been reported.[53]

Gastro-oesophageal reflux disease-related quality-of-life and symptom scores

Data available for each of the trials that describe GORD QoL or symptom scores at 1, 3 and 5 years' follow-up are summarised in *Tables 24–26*. Although it is not possible to combine data because different instruments (or subscales of instruments) were used in the trials, the results are consistent.

At 1 year there are eligible data from all four trials (see *Table 24*). In each case there are highly statistically significant differences all favouring the surgically managed groups. As mentioned above, the randomised element of the Mahon *et al.* trial[51–53] ended at 1 year but data at 3 years are available for the other three trials (see *Table 25*). Again, all favour the surgical group and this was statistically significant in both the LOTUS[47–50] and the REFLUX[1–3] trials.

Only the LOTUS and (now) the REFLUX trial have reported 5-year follow-up. GORD-related QoL scores significantly favour the surgical groups in both trials (see *Table 26*).

TABLE 24 Gastro-oesophageal reflux disease-related QoL and symptom scores at 1 year

Trial	Surgical		Medical		Mean difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
GERSS	52	8.3 (8.4)	52	13.6 (9.5)	–5.3 (–8.7 to –2.0)	0.002
***LOTUS*[47–50]**						
QOLRAD						
Vitality	203	6.84 (0.52)	220	6.42 (0.92)	0.42 (0.28 to 0.56)	<0.001
Food and drink	203	6.78 (0.60)	220	6.34 (0.98)	0.44 (0.28 to 0.60)	<0.001
Sleep	203	6.87 (0.49)	220	6.53 (0.76)	0.34 (0.22 to 0.46)	<0.001
Physical/social	203	6.93 (0.36)	220	6.72 (0.52)	0.21 (0.12 to 0.30)	<0.001
GSRS						
REFLUX dimension	248	1.18 (0.44)	266	1.66 (0.88)	–0.48 (–0.60 to –0.36)	<0.001
***Mahon*[51–53]**						
GSRS	80	37.0 (5.4)	86	35.0 (7.3)	2.00 (0.003 to 3.94)	0.003
***REFLUX*[1–3]**						
REFLUX QoL	178	84.6 (17.9)	179	73.4 (23.3)	14.0 (9.6 to 18.4)	<0.001

TABLE 25 Gastro-oesophageal reflux disease-related QoL and symptom scores at 3 years

Trial	Surgical		Medical		Mean difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
GERSS	49	6.21 (8.66)	44	9.05 (10.40)	–2.84 (–6.77 to 1.09)	0.166
***LOTUS*[47–50]**						
QOLRAD						
Vitality	181	6.90 (0.31)	189	6.53 (0.85)	0.37 (0.24 to 0.50)	<0.001
Food and drink	181	6.85 (0.40)	189	6.38 (0.91)	0.47 (0.33 to 0.61)	<0.001
Sleep	181	6.92 (0.33)	189	6.53 (0.82)	0.39 (0.26 to 0.52)	<0.001
Physical/social	181	6.94 (0.25)	189	6.74 (0.58)	0.20 (0.11 to 0.29)	<0.001
Mahon*[51–53] – *trial terminated at 1 year						
***REFLUX*[1–3]**						
REFLUX QoL	132	87.0 (15.0)	134	79.7 (20.1)	9.0 (4.9 to 13.1)	<0.001

Health-related quality of life

No general HRQoL measure has been reported for the LOTUS trial.[47–50] Data for the other three trials are shown in *Tables 27–29*. The SF-36 was used in the Anvari *et al.* trial[44–46] as it was in the REFLUX trial.[1–3] Unfortunately, it is reported only as the two summary component scores, physical (PCS) and mental (MCS), plus the 'general health' domain score. For comparability, in *Tables 27* and *28* the same score formats

TABLE 26 Gastro-oesophageal reflux disease-related QoL and symptom scores at 5 years

Trial	Surgical		Medical		Difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46] – no data available**						
***LOTUS*[47–50]**						
QOLRAD						
Vitality	160	6.86 (0.44)	179	6.49 (0.99)	0.37 (0.20 to 0.54)	<0.001
Food and drink	160	6.80 (0.51)	179	6.47 (0.80)	0.33 (0.18 to 0.48)	<0.001
Sleep	160	6.89 (0.47)	179	6.61 (0.72)	0.28 (0.15 to 0.41)	<0.001
Physical/social	160	6.94 (0.23)	179	6.75 (0.51)	0.19 (0.10 to 0.28)	<0.001
***Mahon*[51–53] – trial terminated at 1 year**						
***REFLUX*[1–3]**						
REFLUX QoL	127	86.7 (13.8)	119	80.7 (20.3)	6.42 (1.61 to 11.23)	0.009

are shown for the REFLUX trial but it should be borne in mind that the eight domain scores shown in *Chapter 3* for the REFLUX trial are more informative.

At 1 year, in both trials, the PCS and MCS favour the surgical group, although only the difference in the PCS in the REFLUX trial[1–3] is statistically significant. Both trials showed marked differences in the 'general health' domain score. There was also a statistically significant difference favouring surgery in the Mahon *et al.* trial[51–53] (based on the PGWI).

Although EQ-5D data were collected in the Anvari *et al.* trial,[44–46] they were not reported in a way that allows interpretation. At baseline, scores were markedly lower in the surgery group [mean 0.68 (SD 0.28) vs 0.76 (SD 0.21)] and the reason for this imbalance is not clear. At 1 year the equivalent results were 0.79 (SD 0.23) compared with 0.81 (SD 0.19), that is, still lower in the surgery group. As shown in *Table 27*, in the REFLUX trial,[1–3] the mean 1-year EQ-5D score was higher in the surgery group ($p = 0.07$).

At 3 years, the report of the Anvari *et al.* trial mentions collection of the SF-36 'every 3 months' but the only data reported are for the 'general health' domain score. This, as in the REFLUX trial, significantly favours the surgical group (see *Table 28*). There is no mention of collection of EQ-5D data in the 3-year follow-up of the Anvari *et al.* trial. At 5 years, the only data describing generic HRQoL are from the REFLUX trial (as the LOTUS trial has not included a measure) (see *Table 29*).

Individual symptoms of gastro-oesophageal reflux disease or its management

Data describing individual symptoms are available for all trials, although only dysphagia was reported in the Mahon *et al.* trial.[51–53]

Heartburn

As would be expected from the overall GORD-related QoL and symptom scores, all three trials providing data reported less heartburn in their surgical groups. At 1 year in the Anvari *et al.* trial,[44–46] the GERSS heartburn subscore is lower in the surgical group ($p < 0.001$); in the LOTUS trial[47–50] there is clearly less heartburn in the surgical group but data are presented only graphically; and in the REFLUX trial[1–3] heartburn rates in the surgical group are around half those in the medical group. At 3 years, Anvari

TABLE 27 Health-related quality of life at 1 year

Trial	Surgery		Medical		Difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
SF-36						
PCS	52	46.4 (10.9)	52	43.9 (10.3)	3.15 (–0.94 to 7.23)	0.13
MCS	52	52.7 (10.9)	52	51.5 (9.1)	0.98 (–2.8 to 4.76)	0.61
General health domain score	52	75.4 (23.2)	52	66.4 (23.6)	12.3 (3.7 to 20.8)	0.005
***LOTUS*[47–50] – not reported**						
***Mahon*[51–53]**						
PGWB	79	106.2 (16.3)	86	100.4 (18.9)	5.8 (0.43 to 11.17), adjusted 7.1 (2.5 to 11.7)	
***REFLUX*[1–3]**						
SF-36						
PCS	150	48.0 (10.2)	161	45.1 (9.7)	3.51 (1.77 to 5.25)	<0.001
MCS	150	46.6 (12.8)	161	45.1 (13.1)	1.63 (–0.79 to 3.85)	0.195
General health domain score	178	45.2 (11.1)	179	40.7 (11.2)	4.8 (2.7 to 6.8)	<0.001
EQ-5D	178	0.75 (0.25)	179	0.71 (0.27)	0.047 (–0.004 to 0.097)	0.07

TABLE 28 Health-related quality of life at 3 years

Trial	Surgery		Medical		Difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
SF-36						
PCS – not reported						
MCS – not reported						
General health domain score	49	78.50 (19.76)	44	71.41 (21.73)	12.19[a] (2.65 to 21.72)	0.0124
***LOTUS*[47–50] – not reported**						
***Mahon*[51–53] – trial terminated at 1 year**						
***REFLUX*[1–3]**						
SF-36						
PCS	128	47.2 (9.9)	127	46.6 (10.0)	1.43 (–0.45 to 3.32)	0.136
MCS	128	48.9 (10.6)	127	45.6 (12.6)	4.05 (1.57 to 6.52)	0.001
General health score	132	45.3 (10.0)	134	42.4 (11.8)	3.69 (1.50 to 5.87)	0.001
EQ-5D	132	0.803 (0.231)	134	0.747 (0.262)	0.070 (0.0015 to 0.126)	0.013

a Presumably adjusted.

TABLE 29 Health-related quality of life at 5 years

	Surgery		Medical			
Trial	*n*	Mean (SD)	*n*	Mean (SD)	Difference (95% CI)	*p*-value
Anvari[44–46] – no data available						
LOTUS[47–50] – not reported						
Mahon[51–53] – trial terminated at 1 year						
REFLUX[1–3]						
SF-36						
PCS	113	46.1 (9.9)	109	46.1 (10.5)	1.47 (–0.84 to 3.79)	0.211
MCS	113	47.8 (11.7)	109	47.9 (11.7)	1.27 (–1.36 to 3.90)	0.343
General health domain score	117	44.1 (10.3)	111	43.2 (11.5)	2.76 (0.21 to 5.31)	0.034
EQ-5D	127	0.774 (0.259)	119	0.761 (0.282)	0.047 (–0.013 to 0.108)	0.126

et al.[44–46] report significantly more heartburn-free days in the surgical group ($p = 0.008$); in the LOTUS trial,[47–50] less heartburn in the surgical group is shown graphically and the *p*-value is reported as <0.001; and in the REFLUX trial[1–3] 51% of the randomised surgical group compared with 75% of the randomised medical management group report any heartburn (see *Table 15*). At 5 years, data are available only from the LOTUS and REFLUX trials. In LOTUS,[47–50] 8% in the surgery group compared with 16% in the medical group are reported to have heartburn, 'although there was no significant difference in the severity of heartburn ($p = 0.14$)'. In the REFLUX trial,[1–3] 41% in the surgery group compared with 74% in the medical group reported any heartburn (see *Table 15*).

Regurgitation

Again, as would be expected from the overall GORD-related QoL and symptom scores, all three trials providing data reported less regurgitation in the surgical groups. At 1 year in the Anvari *et al.*[44–46] trial, the GERSS regurgitation subscore is significantly lower in the surgical group ($p = 0.002$); in the LOTUS trial,[47–50] graphical presentation clearly indicates less regurgitation in the surgical group, although no figures are reported; and in the REFLUX trial,[1–3] regurgitation rates in the surgical group are half those in the medical group. At 3 years, information is available only for the LOTUS and REFLUX trials and both report lower rates in the surgical groups. At 5 years in the LOTUS trial, 2% in the surgical group compared with 13% in the medical group ($p<0.001$) have regurgitation, and in the REFLUX trial 25% in the surgical group compared with 37% in the medical group report any regurgitation.

Dysphagia

As mentioned in *Chapter 1*, dysphagia following both open fundoplication and laparoscopic fundoplication has been reported. At 1 year, Anvari *et al.*[44–46] report a higher GERSS dysphagia subscore in the surgical group but this was not statistically significant ($p = 0.8$); in the LOTUS trial[47–50] there were more reports of dysphagia in the surgical group but data were presented only graphically; in the Mahon *et al.* trial,[51–53] dysphagia persisting beyond 3 months was reported in 5 out of 104 (4.8%) having surgery; and in the REFLUX trial,[1–3] rates of 'difficulty swallowing' were the same in the two randomised groups. At 3 and 5 years, information is available only from the LOTUS and REFLUX trials. In the LOTUS trial there is more dysphagia in the surgical group ($p<0.001$) at both time points: at 5 years 11% in the surgical group report dysphagia compared with 5% in the medical group. In the REFLUX trial, one further participant had undergone oesophageal dilatation (see *Table 12*), but the numbers reporting difficulty swallowing were the same in the two randomised groups (see *Table 15*, e.g. any difficulty swallowing 24.2% vs 23.9%).

Flatulence

Flatulence has also been reported as more common after both open and laparoscopic fundoplication. Information is available only from the LOTUS[47–50] and REFLUX[1–3] trials. In the LOTUS trial, flatulence was more common in the surgery group than in the medical management group at 1, 3 and 5 years. At 5 years, the rates are 57% in the surgical group and 40% in the medical group ($p < 0.001$). In the REFLUX trial, rates of 'wind from the lower bowel' are not statistically significantly different between the groups [more than three times per week: 65.0% in the randomised surgical group vs 59.4% in the randomised medical group at 1 year; 57.6% vs 58.5% at 3 years; and 65.3% vs 60.0% at 5 years (see *Table 15* for more detail)].

Other symptoms

In the LOTUS trial,[47–50] 'bloating' was reported more commonly in the surgical group (40% vs 28% at 5 years). In contrast, 'bloating/trapped wind' was reported less commonly in the surgical group in the REFLUX trial[1–3] (at 1 year: 72.1% vs 82.4%). A particular concern following fundoplication is an inability to vomit despite wanting to. In the REFLUX trial we attempted to address this through a question on 'frequency of wanting to be sick but being physically unable to' and found no difference between the groups (see *Table 15*).

Surgical complications

Like all procedures involving surgery under general anaesthesia, laparoscopic fundoplication carries risks. *Table 30* summarises intra and early postoperative complications reported in the four trials.

Conversion to an open procedure

The decision to convert from a laparoscopic to an open approach is usually indicative of difficulties experienced during the procedure. Rates varied from 0% in the Anvari *et al.* trial[44–46] to 2.4% in the LOTUS trial[47–50] (see *Table 30*).

Intraoperative complications

In the Mahon *et al.*[51–53] and REFLUX[1–3] trials combined, the 10 intraoperative complications reported (overall rate 2.3%) were injuries to the spleen ($n = 3$), liver ($n = 3$), pleura ($n = 3$) and oesophagus ($n = 1$). In the LOTUS trial[47–50] it was unclear whether intraoperative complications occurred or whether they were incorporated within all postoperative complications; however, the report noted that 29 participants encountered a variety of operative difficulties that were described as 'trivial'.

TABLE 30 Intra- and early postoperative events in the four trials[a]

Trial	*n* having operation	Conversion, *n* (%)	Intraoperative complications, *n* (%)	Postoperative complications, *n* (%)
Anvari[44–46]	51	0 (0.0)	0 (0.0)	7 (13.7)
LOTUS[47–50]	248	6 (2.4)	Unclear	7 (2.8)
Mahon[51–53]	109	1 (0.9)	4 (3.7)	6 (5.5)
REFLUX[1–3]				
Randomised	111	2 (1.8)	2 (1.8)	1 (0.9)
Preference	218	0 (0.0)	4 (1.8)	2 (0.9)

a Adapted from Wileman *et al.*[43]

Early postoperative complications

In the Anvari *et al.* trial,[44–46] seven (14%) participants had postprandial bloating, two of whom were treated with a single dilatation of the wrap. No details are given of the postoperative complications in the LOTUS trial. In the Mahon *et al.* trial[51–53] there were three wrap migrations, two respiratory tract infections and one case of a sutured nasogastric tube. In the REFLUX trial,[1–3] one participant in the randomised group and two in the preference group were admitted to a high-dependency unit immediately after the surgical procedure.

Reoperations

By the time of the 3-year follow-up in the Anvari *et al.* trial,[44–46] 4 of 51 (7.8%) participants had undergone a second fundoplication operation. Four (3.7%) in the Mahon *et al.* trial[51–53] required reoperation within 3 months of their first fundoplication, one of whom had a gastric resection because of necrosis. It is not clear if anyone in the LOTUS trial[47–50] had a reoperation. As shown in *Table 11*, in the REFLUX trial,[1–3] 5 of the 112 (4.5%) randomised to surgery who actually had a fundoplication had a second reflux-related operation, and this applied to 16 (4.4%) of the total 364 participants in the study who had a laparoscopic fundoplication.

Other late postoperative complications

Dilatation of the wrap was reported for two (3.9%) people in the Anvari *et al.* trial[44–46] and four (3.7%) in the Mahon *et al.* trial.[51–53] It is not stated whether or not dilatation occurred in the LOTUS trial.[47–50] In the REFLUX trial,[1–3] two (1.8%) participants in the randomised surgical group (plus two in the preference surgical group – giving an overall rate of 1.1%) had stricture dilatation or food disimpaction (see *Table 12*). There were three cases (0.8%) of repair of incisional hernia in the REFLUX trial – all in the preference group – but this complication was not mentioned in the other trials' reports. There were no deaths in any of the trials associated with surgical or medical management.

Surgery-related mortality

No perioperative deaths were reported among the 771 people in the four trials who had fundoplication surgery.

Discussion

Of the four trials, the REFLUX trial is the most pragmatic in design. It involved a large proportion of UK centres where laparoscopic anti-reflux surgery is undertaken and the surgery was undertaken by NHS upper gastrointestinal surgeons within these centres, all of whom had experience of carrying out the procedure. The exact method of fundoplication was left to the discretion of the surgeon, so he or she was comfortable with the approach. After surgery and, in the medically treated patients, after optimisation of their PPI medication, care of the participants was the responsibility of GPs. The principal measure of outcome was a patient-reported disease-specific QoL measure. Unlike the other trials, the REFLUX trial was coordinated from an accredited trials unit, local recruitment was led by gastroenterologist/gastrointestinal surgeon partnerships rather than by gastrointestinal surgeons alone, and the trial was publicly funded through the HTA programme rather than by industry.

In respect of potential benefits of surgery, the four trials appear to be consistent. All show significantly better relief of GORD symptoms for as long as the length of their current follow-up. (Surprisingly, the LOTUS trial report[48] does not draw attention to this but, judged on data describing the QOLRAD reported in an e-table, there are significant differences between the groups in all dimensions of this instrument, favouring surgery.) Data available describing the principal symptoms of GORD (heartburn and regurgitation) show large differences, again favouring surgery. Only limited data are available from generic QoL measures, and much of this is from the REFLUX trial; although differences are less marked than for the GORD-related QoL instruments, they are consistent with benefit from surgery.

The four trials are broadly consistent in respect of intraoperative and early postoperative complications: a small number of operations are converted to an open procedure, a small number of laparoscopic procedures have associated visceral injuries, a small number of people have problems postoperatively and a small number require dilatation of the wrap. The REFLUX trial suggests that 4.5% have reoperations and the other trials are broadly consistent with this. None of the trials had a reported perioperative death. Data from the Finnish Registry[56] suggest a mortality of 0.1%, but this is based on a single case among 1162 people who had laparoscopic fundoplication; furthermore, the registry included all cases of fundoplication and hence went beyond the sorts of patients recruited to the REFLUX trial.

The other trials, particularly the LOTUS trial, show higher rates of dysphagia and flatulence following laparoscopic fundoplication than in the medically managed group. As mentioned above, a small number of participants in the REFLUX trial did have a dilatation procedure, presumably because of difficulty swallowing, but this was not reflected in responses to the REFLUX questionnaire, suggesting that there were only a few isolated cases of dysphagia following surgery in this trial. Similarly, there were no significant differences in flatulence in the REFLUX trial.

Hence, taking all four trials together, it is now possible to give a clear picture of most of the potential benefits and risks of laparoscopic fundoplication, at least up to 5 years. There are, however, differing resource implications of surgery and medical management. In the next chapter we explore whether or not the benefits of surgery in patients with established GORD requiring long-term PPI therapy for reasonable control and suitable for either clinical policy (average age around 45 years) are sufficient to outweigh any differences in costs.

Chapter 5 Economic analysis

The economic evaluation aimed to determine the cost-effectiveness of laparoscopic fundoplication compared with continued medical management in patients with GORD symptoms that are reasonably controlled by medication and who are judged suitable for both surgical and medical management. The analysis entailed three components:

1. systematic review of existing cost-effectiveness evidence
2. within-trial (5-year) economic analysis
3. validation of within-trial analysis and exploration of the need for a longer-term model.

Systematic review of existing cost-effectiveness evidence

The aim of this systematic review is to identify any existing cost-effectiveness studies that compare laparoscopic fundoplication with medical management for GORD. A previous HTA report included a review of the evidence available from 1995 to December 2005 and identified three relevant studies (described below).[1] The updated search focuses on the period from December 2005 to April 2011. The methods used to identify studies and the results of the systematic search are discussed in the sections below.

Methods

The following data sets were searched to identify published evidence: MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (1948 to present), EMBASE (1996 to week 15, 2011), Cochrane Database of Systematic Reviews (CDSR) and the NHS Centre for Reviews and Dissemination databases [Database of Abstracts of Reviews of Effects (DARE), NHS Economic Evaluation Database (NHS EED), HTA]. The search strategy incorporated broad reflux-related search terms as used in a recent Cochrane Review.[57] The search also focused on identifying health-related and GORD-specific QoL evidence.

Studies were considered relevant for inclusion in the review if they were published in English and were full health economic evaluations (cost-effectiveness, cost-utility or cost–benefit analysis) comparing costs and outcomes associated with laparoscopic fundoplication and medical management. For the purpose of this study laparoscopic fundoplication includes both complete and partial wrap procedures. Publications outside the above criteria were excluded from this review. Details of the updated search strategy are presented in *Appendix 6*.

Results

A total of 3662 references were identified from the searches (MEDLINE: 1640, EMBASE: 1825, CDSR: 44, DARE: 56, NHS EED: 85, HTA: 12). Titles and/or abstracts were reviewed and studies that satisfied all inclusion criteria were included in the review. Papers describing five additional studies were obtained for inclusion. These were published between 2007 and 2011 and were related to the UK and Canadian settings. Of the total of eight studies, five are linked to three of the randomised trials described in *Chapter 4*: Anvari *et al.*,[44–46] Mahon *et al.*[51–53] and the REFLUX trial,[1] the long-term follow-up of which is the topic of this report. There is no economic evaluation in the LOTUS trial.[48] Three of the studies were based on the REFLUX trial. These were published as part of the earlier HTA report[1] and in two journal articles.[3,5] Summaries of the two within-trial economic evaluations are presented in *Appendix 7*. Below is a brief description of the eight reports – the five linked to the three randomised trials are considered first, followed by the three studies based on observational data.

Economic analyses based on clinical trials

Economic evaluation based on the Anvari et al. *trial*[46]

This was an economic evaluation conducted alongside the Anvari *et al.* trial described in *Chapter 4*. Laparoscopic fundoplication was compared with PPI for patients with chronic GORD. The follow-up period was 3 years and the analysis was conducted from a societal perspective. Cost-effectiveness was reported in terms of cost per QALY gained.

Three generic preference-based questionnaires were administered during the trial: Health Utilities Index Mark 3 (HUI3), EQ-5D and Short Form questionnaire-6 dimensions (SF-6D). Although these instruments have been valued by large general public samples, they differ in the attributes used for their descriptive system and the method of valuation applied. The EQ-5D has been valued using time trade-off whereas the SF-6D and HUI3 use the standard gamble. Utility scores showed an improvement in patients' HRQoL in both groups across the three utility instruments; however, the degree of improvement varied according to the utility instrument used. The base-case analysis (using the HUI3 instrument), after adjustment for baseline differences, indicated that, over the 3 years, laparoscopic fundoplication patients experienced a 0.109 gain in QALYs compared with PPI patients. The ICER for laparoscopic fundoplication patients was around C$29,400 (£19,000) per QALY gained. An increased ICER of C$76,300 (£49,300) was obtained using the EQ-5D as the HRQoL measure.

Economic evaluation based on the Mahon et al. *trial*[52]

This study looked at the cost-effectiveness of laparoscopic fundoplication compared with maintenance PPI medication for severe GORD based on the Mahon *et al.* randomised trial described in *Chapter 4*. Results based on the 12-month follow-up were extrapolated using other published data sets. Costs and outcomes for up to 12 months were obtained from a sample of patients in the trial (the first 100) and resource use was quantified using data from hospital records and GPs' notes. The incremental cost of laparoscopic fundoplication compared with PPI therapy per additional patient returned to a physiologically normal acid score (<13.9) at 3 months was £5515 (95% CI £3655 to £13,400) and the incremental cost per point improvement in combined gastrointestinal and psychological well-being score at 12 months was £293 (90% CI £149 to £5250). The authors concluded that laparoscopic surgery would break even compared with medical management after 8 years and would be cost saving thereafter.

Economic evaluation based on the REFLUX trial[1,3,5]

Bojke *et al.*[5] present a preliminary cost-effectiveness analysis conducted before the availability of the 1-year REFLUX trial results. The analysis compared the cost-effectiveness of surgery (laparoscopic fundoplication) with long-term medical management (PPIs) for GORD disease in an average 45-year-old man. A lifetime (30 years) Markov model that adopted the perspective of the NHS was developed. Effectiveness data were obtained from a fixed-effect meta-analysis that synthesised data from multiple sources. QALYs were estimated using utility scores (measured by the EQ-5D instrument) derived from a subset of UK patients included in the REFLUX trial. Over a lifetime, expected costs associated with surgery (£5014) were higher than expected costs associated with PPI (£4890). Expected QALYs associated with surgery (13.04) were greater than QALYs associated with PPIs (12.36). The incremental cost per QALY gained (ICER) for surgery compared with medical care was £180. The estimated probability that surgery was cost-effective at the threshold of £30,000 per QALY was 0.639. The authors highlighted important areas for further research, such as the HRQoL of patients on PPIs or post surgery.

The within-trial cost-effectiveness analysis, comparing laparoscopic fundoplication with medical management 1 year post surgery, was described in full in the 2008 report of the REFLUX trial.[1] The analysis was conducted on an ITT basis from a NHS perspective. HRQoL was assessed at baseline and at 3 and 12 months' follow-up using the EQ-5D. Cost-effectiveness was reported in terms of the difference in mean QALYs between the treatment groups. This difference was estimated using ordinary least squares (OLS) regression, adjusting for baseline differences in EQ-5D between individuals. The estimated difference in mean costs between the groups was £1280 (95% CI £1054 to £1468). The HRQoL of patients randomised

to surgery tended to improve on average by 0.066 more QALYs (95% CI 0.023 to 0.107) than in the medical management group. The estimated mean ICER was around £19,000. At a threshold of £30,000 per QALY, the probability of surgery being cost-effective was 0.86.

Epstein *et al.*[3] developed a Markov model using 12-month data from the REFLUX trial and other sources in order to extrapolate the cost-effectiveness of laparoscopic fundoplication compared with medical management over the longer term (lifetime). Cost-effectiveness was reported in terms of the cost per QALY gained from surgery. The analysis was conducted from a NHS perspective. Under base-case assumptions, surgery had an additional mean cost of £847 and additional mean QALYs of 0.37 over the lifetime of the patients. The incremental cost per additional QALY gained was around £3000. At a threshold of £20,000 per QALY, the probability that surgery was cost-effective was around 0.74.

Economic analyses based on observational data

Economic evaluation based on Romagnuolo et al.[58]

This study is based on observational data and compares the cost-effectiveness of maintenance regimens of omeprazole and laparoscopic fundoplication within the Canadian medical system. The effectiveness, HRQoL and resource-use data were derived from studies published between 1985 and 2000. Outcomes were expressed as QALYs and costs were estimated from the perspective of a provincial health ministry. A two-stage Markov model (healing and maintenance phases) was used to estimate costs and utilities using a time horizon of 5 years. Laparoscopic fundoplication was the most cost-effective option at 3.3 years of follow-up and was cost saving at 5 years. These results were sensitive to the price of omeprazole. QALYs did not differ significantly between treatment groups.

Economic evaluation based on Arguedas et al.[59]

This study, also based on observational data, compared the cost-effectiveness of laparoscopic fundoplication and medical management in patients with severe reflux oesophagitis. Outcomes were quantified using QALYs with model inputs derived from the published literature. A Markov simulation model was used to extend a previous analysis to a 10-year time horizon. Procedure and hospitalisation costs were estimated using Medicare reimbursement rates from the authors' institution. Medical therapy was associated with a total cost of $8798 and 4.59 QALYs, whereas the surgery was more expensive ($10,475) and less effective (4.55 QALYs). The authors concluded that medical therapy dominated surgery.

Economic evaluation based on Comay et al.[60]

This is a cost-effectiveness analysis, based on observational data, principally concerned with assessing an endoscopic therapy (Stretta procedure) compared with PPIs and laparoscopic fundoplication in the management of GORD. The Strettra procedure is out of the scope of our analysis; however, the data on costs and QALYs provided by the authors allow us to better understand QoL related to these technologies and make comparisons with other authors' estimates. The authors constructed a Markov model that tracked patients over a period of 5 years. Analysis was undertaken from the Canadian Ministry of Health perspective. A literature review for published studies before 2004 was carried out to derive effectiveness and utility data. Symptom-free months and QALYs were used to measure benefit. PPI was the dominant strategy, producing more symptom-free months at lower costs than the other strategies. Laparoscopic fundoplication was associated with higher costs and generated more QALYs. The discounted mean QALYs over 5 years were 4.6487 for laparoscopic fundoplication and 4.6357 for PPI. The ICER for laparoscopic fundoplication compared with PPI was C$384,692 (£240,470). This is unlikely to be considered cost-effective.

Conclusions

The different outcomes used make it difficult to compare the results of the various studies analysed here. For those studies quantifying the benefits associated with the two treatments using QALYs, the results differ depending on the type of analysis conducted. Although the trial-based results suggest that there

is good short- and medium-term evidence indicating that surgery may well represent a cost-effective alternative intervention, the model-based studies are not so optimistic.

The ICER for surgery ranged from £180 to £49,000 per QALY gained. However, the limitations of the studies included in this review suggest that we should be cautious when interpreting these results. The decision model developed as part of the REFLUX trial extrapolated from data at 12 months and was based on the assumption that the treatment effect of surgery (in terms of impact on HRQoL) remains constant over the lifetime of patients. However, as would be expected, the results of the sensitivity analysis suggested that surgery was less cost-effective when the beneficial effect of surgery was limited to 5 years (increasing the ICER to £11,300) and when HRQoL was worse in those for whom surgery failed (increasing the ICER to £11,310 when considering very high rates of surgical failure).

The value of conducting additional research to reduce any uncertainty in the REFLUX model was demonstrated. The expected value of perfect information (EVPI) is the maximum amount that a decision-maker should be willing to pay to eliminate all uncertainty that arises because of imprecision in the parameters of the model. The value of information analysis suggested that further research could be worthwhile. At a threshold of £30,000, the per-patient EVPI was £15,106.

Within-trial economic evaluation

Follow-up data from the REFLUX trial up to 5 years after surgery are now available. These economic data represent the longest follow-up of randomised patients currently available. These data can help to inform the question regarding the sustainability of initial improvement in HRQoL following surgery. This section describes the updating of the cost-effectiveness analysis using these data to reduce the level of uncertainty about the cost-effectiveness of surgery and thus its role in the NHS.

Overview

Differences in mean costs and QALYs at 5 years (based on data collected within the REFLUX trial) were used to derive an estimate of the cost-effectiveness of laparoscopic surgery (laparoscopic fundoplication) and continued medical management. The extent of missing data throughout the trial follow-up is significant; therefore, the base case consists of the multiple imputed data set following ITT analysis. A separate scenario – complete-case analysis, in which patients with any missing data are excluded – was employed for ITT and PP for 1-year analyses. Costs and QALYs were evaluated on the basis of costs falling on the NHS and Personal Social Services expressed in UK pounds sterling at a 2010 price base. All analysis and modelling were undertaken in Stata/SE 11.1 (StataCorp LP, College Station, TX, USA).

Methods

Patient population

As described in earlier chapters, the patient population in the REFLUX trial was patients with GORD whose symptoms required medication for reasonable control and for whom either surgery or continued medical management appeared to be an acceptable treatment option. A policy of offering relatively early laparoscopic fundoplication was compared with the alternative policy of continued medical management. The analysis used data only from the randomised trial component of the REFLUX trial (i.e. not from the preference groups). As described in *Chapter 3*, 357 patients were randomised to either surgical treatment ($n = 178$) or medical management ($n = 179$) and patients were followed for up to 5 years.

Health-care resource use

Health-care resource-use data were collected prospectively as part of the clinical report forms and patient questionnaires at 3 and 12 months and 2, 3, 4 and 5 years. Patient questionnaires at 3 and 12 months collected information for the previous 3 and 9 months respectively. In addition, a questionnaire at 12 months recorded resource use for the whole of the first year (see following section on costs). Patient

questionnaires from the second year onwards collected information for the previous 12 months on hospital admissions (day and overnight admissions) and GP visits, and data on medication for the previous 2 weeks. Clinical report forms collected data on surgery and perioperative complications of surgery.

Costs

The cost for each individual patient in the trial was calculated by multiplying his or her use of health-care resources by the associated unit costs (*Table 31*). Discount was applied from year 2. Unit costs were all sourced from published data (see *Table 31*). Total costs include the costs of surgery, GP visits, hospital admissions and medication. Incremental costs (laparoscopic fundoplication vs medical management) for each year and per category of resource use, according to ITT allocation, were calculated using OLS regression.

The questionnaires asked for details of anti-reflux medication taken in the previous 2 weeks: name, dose and number of tablets/capsules. The cost of anti-reflux medication during these 2 weeks was calculated by multiplying the prices published in the Drug Tariff for December 2010[61] for each medicine by the number of tablets taken. Yearly medication costs are calculated using the area under the curve method,[62] which assumes linear interpolation between follow-up points. The costs of reflux-related inpatient, outpatient and day-case visits were derived from the *NHS Reference Costs 2009–10*,[63] in which the relevant codes were weighted by activity level.

For the base-case analysis, total costs included the costs of surgery, complications due to surgery, reoperations, reflux-related prescribed medication, reflux-related visits to and from the GP and reflux-related hospital inpatient, outpatient and day visits. For the sensitivity analysis, all GP visits and all hospital admissions are included in the calculation of total costs (see *Incremental analysis* for more details on sensitivity analysis). Costs of hospital admissions and GP visits were obtained by multiplying the relevant unit costs by the numbers of admissions and visits reported by the patients respectively. Patients themselves classified how many visits and admissions were reflux related in relation to the total number of visits. There is a possibility that patients may not have fully understood the clinical consequences of GORD; hence, they may misclassify the reason for a consultation. If such misclassification is different across treatment groups, estimates of incremental costs may be biased.

For the first year of the trial, data on resource use were collected at 3 months and 12 months, and for the whole year using an additional questionnaire. To make the most efficient use of the data available for the first year of the trial, resource use at 1 year was estimated as the greater of the area under the curve between the first and second questionnaire and the 12-month health-care survey. This is in line with the procedure employed for the earlier publication evaluating the REFLUX trial.[1]

The cost of surgery included the costs of (1) presurgical procedures (endoscopy, pH monitoring and manometry), (2) the surgery team, (3) operative complications, (4) hospital stay, (5) capital costs and overheads and (6) consumables. The cost of reoperations was assumed to be equivalent to the mean cost of the first surgery. The cost of reflux-related visits to and from the GP was assumed to be equivalent to the average cost of visits to and from the GP.[64]

Quality-adjusted life-years

Health outcomes were expressed in terms of QALYs. HRQoL was assessed in the REFLUX trial at baseline and 3 months and then yearly until 5 years using the EQ-5D.[65,66] The EQ-5D is a standardised and validated generic instrument for the measurement of HRQoL. It has five dimensions: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort, and anxiety and depression. Each dimension has three possible responses (no problems, moderate problems or severe problems), creating 245 mutually exclusive health states. Each of these health states has been valued in a large UK population study using the time trade-off method, in which 1 corresponds to perfect health (thus the maximum value possible) and 0 corresponds to death.[65,66]

TABLE 31 Unit costs employed to calculate the costs of reflux-related health-care use

Health-care activity	Resource	Cost (£)	Source
Laparoscopic fundoplication surgery	Endoscopy	218.52	Grant *et al.*[1] (inflated to 2009–10 prices using Curtis[64])
	pH test	81.85	
	Manometry	76.94	
	Operation cost per minute	6.36	
	Capital cost per surgery	11.71	
	Consumables	1080.96	
	High-dependency unit per night	797.86	
	General ward per night[a]	282.78	*NHS Reference Costs 2009–10* (excess bed stay)[63]
Hospital admissions[b]	Overnight admission due to surgery	2108.22	Mean surgery cost
	Overnight admission due to complications[a]	1534.76	*NHS Reference Costs 2009–10* (elective inpatient)[63]
	Day case	559.00	*NHS Reference Costs 2009–10* (day case)[63]
	Outpatient	221.98	*NHS Reference Costs 2009–10* (outpatient)[63]
GP use	Visit from GP	120.00	Curtis[64]
	Visit to GP	36.00	

Health-care activity	Resource	Cost (£)	Source
Medication			
PPI	Omeprazole 10 mg, 28 capsules	1.92	Drug Tariff December 2010[61]
	Omeprazole 20 mg, 28 capsules	1.81	
	Omeprazole 40 mg, 7 capsules	1.95	
	Lansoprazole 15 mg, 28 capsules	1.44	
	Lansoprazole 30 mg, 28 capsules	2.23	
	Pantoprazole 20 mg, 28 tablets	1.79	
	Pantoprazole 40 mg, 28 tablets	2.82	
	Rabeprazole 10 mg, 28 tablets	11.56	
	Rabeprazole 20 mg, 28 tablets	19.55	
	Esomeprazole 20 mg, 28 tablets	18.50	
	Esomeprazole 40 mg, 28 tablets	25.19	
H_2RA	Ranitidine 150 mg, 60 tablets	1.97	
	Ranitidine 300 mg, 30 tablets	2.17	
	Famotidine 20 mg, 28 tablets	4.40	
	Famotidine 40 mg, 28 tablets	5.55	
	Nizatidine 150 mg, 30 capsules	12.04	
	Nizatidine 300 mg, 30 capsules	15.34	
	Cimetidine 400 mg, 60 tablets	7.61	
	Cimetidine 800 mg, 30 tablets	21.63	
Prokinetic	Domperidone 10-mg tablets[c]	1.53	
	Metoclopramide 10 mg, 28 tablets	1.01	

a Average of the relevant cost for diagnoses codes FZ24B, FZ24C, FZ24D, FZ25A, FZ25B, FZ27A, FZ27B, FZ27C, FZ27D, FZ28A, FZ28B, FZ28C, FZ29Z, FZ30Z, FZ31D, FZ31E and FZ31F, weighted by activity levels.

b Costing hospital admissions included the cost of the individual procedures (endoscopy, pH test, manometry, barium meal) using the unit cost data used for costing surgery.

c Average cost of 30-tablet and 100-tablet pack.

DOI: 10.3310/hta17220

QALYs for each patient were calculated as the area under the curve following the trapezium rule,[67] which assumes linear interpolation between follow-up points. Incremental mean QALYs between treatment groups were estimated with and without adjustment for baseline utility, using OLS regression.

Discounting

Costs and outcomes from year 2 were discounted using a 3.5% annual discount rate, in line with current guidelines.[65,68]

Missing data and multiple imputation

Given the extent of missing data, the multiple imputed data set is presented as the base case. This was created using all available data and multiple imputation with chained equations.[69] Mean imputation was used to predict missing data at baseline,[70] as randomisation should ensure equal distribution of potentially confounding variables. Complete-case analysis refers to only those patients who returned all questionnaires and completed all EQ-5D profiles.

Missing or inconsistent answers to questions on resource use were dealt with as follows. For medication use, patients were asked at each follow-up questionnaire whether or not they were using prescribed medication for reflux and, if so, to indicate the name, strength and the number of tablets taken in the past 2 weeks. It was evident from preliminary analyses that the answers to the first question were not necessarily consistent with the answers to the second question. Therefore, the following rule was applied for the costing of drugs: (1) if the patient provided the name, strength and number of tablets taken, he/she was assumed to be taking medication; (2) if the patient did not specify either a drug or the number of tablets taken, he/she was considered not to be taking medication; (3) if the patient specified a particular drug but no dosage, the missing data were imputed as the median of all other patients on that medication. Similarly, missing answers to the questions regarding GP visits and hospital admissions were assumed to indicate that no visits or admissions occurred. Because of the nature of the questionnaire, it is reasonable to assume that absence of an answer indicates no use of services.

Multiple imputation[71] was the statistical technique chosen to deal with missing cost and HRQoL data because of non-returned questionnaires and incomplete EQ-5D profiles, using the user-defined programme 'ice' in Stata 11.1. Multiple imputation presents three major advantages over standard ad hoc methods for dealing with missing data (such as mean imputation and last value carried forward): (1) it makes full use of all of the available data, (2) it incorporates uncertainty associated with the missing data and (3) it ensures unbiased estimates and standard errors as long as data are MAR.[69] [Little and Rubin[72] defined three missing data mechanisms: (1) MCAR if the probability of data being unobserved is independent of both observed and unobserved values; (2) MAR if the probability of data being unobserved is dependent on the observed values but independent of unobserved ones and (3) missing not at random (MNAR) if the probability of data being unobserved is dependent on unobserved values.]

Multiple imputation follows three steps. First, regression models are used to predict plausible values for the missing observations from the observed values. A random component is included to reflect the uncertainty around the predictions. These values are then used to fill in the gaps in the data set. This process is repeated *m* number of times (*m* being the number of imputations), creating *m* number of imputed data sets. Second, each data set is analysed independently using complete-case methods. Third, the estimates obtained from each imputed data set are combined to generate mean estimates of costs and QALYs, variances and CIs using Rubin's rules,[73] in such a way that the uncertainty around the predicted values is fully taken into account.[69,74] Because the REFLUX trial has missing data for both costs and EQ-5D scores, multiple imputation using chained equations (MICE) was employed. For MICE, each variable is predicted with its own regression model. Each imputed data set is created by running the regression models over several cycles, in which each variable informs the prediction of the other variables.[69,74] To obtain overall estimates of mean and incremental costs and QALYs across all of the imputed data sets, the 'mim' command was used.[75] Semi-parametric bootstrapping in Stata 11.1 was employed to estimate the

probability that surgery is cost-effective, while maintaining the correlation between costs and QALYs (see *Incremental analysis* for more details).[76]

Plausible prediction of the missing data depends on the appropriate specification of the regression models used in MICE.[74] If a model is misspecified, the distribution of imputed values may not resemble that of the observed values, and thus the estimates of treatment effect may be biased.[69] The regression model specified will depend on the type and distribution of the variable to be predicted.[70] The variables required for the economic evaluation are costs for each year and EQ-5D scores at each time point. Both are continuous variables and neither is normally distributed; EQ-5D scores in the REFLUX trial are bounded between –0.594 and 1,[66] and costs are bounded at zero and tend to present a positive skew. Two approaches to deal with non-normality with MICE have been suggested in the literature:[69] (1) transformation towards normality and (2) predictive mean matching. [In predictive mean matching the missing observation is imputed with an observed value from an individual with a similar linear predictor.[70] Consequently, the distribution of imputed values tends to closely match the distribution of the observed values.[69]] Using the REFLUX data set none of the transformation approaches (Box–Cox,[77] log-transformation and log-transformation of non-zero values with generation of an indicator variable[78]) were successful in transforming the data distribution to normality. As a result, predictive mean matching was the strategy employed to ensure that the distribution of imputed values closely resembled the distribution of observed values. All known covariates thought to be associated with the missingness mechanism, costs and EQ-5D scores were included in the prediction equations: EQ-5D scores at each follow-up point, costs at each year, allocation, BMI, age and sex. A total of 100 imputations ($m = 100$) was used to ensure efficient and reproducible estimates.[69]

Multiple imputation provides unbiased estimates of treatment effect if data are MAR. Whether or not data are MAR is an untestable assumption by definition, as unobserved values are unknown. Departure from the MAR assumption may have implications for decision-making if the results from the cost-effectiveness analysis differ from those of the base case. Sensitivity analysis was used to test the impact on the cost-effectiveness results if data were MNAR, that is, if patients with worst outcomes or greater costs were more likely to have missing data.[70,79] Four scenarios were tested. In scenario (1), all patients with missing data had their total QALYs reduced by 10%, 20%, 30%, 40% and 50%. Conversely, in scenario (2), for all patients with missing data costs were increased by the same proportions (10%, 20%, 30%, 40% and 50%). In scenario (3), only surgery patients with missing data had their QALYs reduced. In scenario (4), costs were increased only for patients undergoing surgery.

Incremental analysis

The cost-effectiveness of surgery was evaluated by comparing the costs and QALYs incurred in the surgery arm with the costs and QALYs in the medical management arm at 5 years of follow-up, using conventional decision rules and estimating ICERs as appropriate.[80] If one intervention is associated with greater mean QALYs and lower mean costs it is deemed cost-effective by dominance. The ICER is calculated if either treatment arm does not dominate. The ICER summarises the additional costs associated with one intervention over another and relates this to the additional benefits. This ICER is then compared with a threshold for the cost per QALY. The National Institute for Health and Care Excellence (NICE) uses a threshold cost per QALY of around £20,000–30,000 to determine whether or not an intervention represents good value for money in the NHS.[65] Consequently, if the ICER is < £20,000, laparoscopic fundoplication could be considered potentially cost-effective. ICERs between £20,000 and £30,000 per QALY are considered borderline and an ICER > £30,000 is not typically considered cost-effective.

The ICER can be re-expressed using the net monetary benefit (NMB). The NMB of an intervention is the value of the health benefits gained from a particular intervention compared with standard care in monetary terms, minus the incremental costs of the intervention. The translation of health benefits into the monetary scale was made using a cost-effectiveness threshold of £20,000. This is the threshold commonly use by NICE (this corresponds to 1 QALY being valued at £20,000). Therefore, the NMB provides a measure of the gain (or loss) in resources of investing in a particular intervention when those resources could have

been used elsewhere.[81] The NMB of laparoscopic fundoplication and medical management were calculated and used to demonstrate the influence of trial duration on the estimates of cost-effectiveness of surgery.

As discussed previously, the multiple imputed data set was used as the base case for the cost-effectiveness analysis because of the large proportion of data lost for the complete-case analysis. Because total costs and total QALYs are cumulative quantities, any missing data at any of the follow-up points will result in that patient being dropped from a complete-case analysis. The cost-effectiveness results using the complete case are presented for comparison. Complete-case analysis will provide unbiased estimates only if the data are MCAR, that is, the probability of data being unobserved is independent of both observed and unobserved values. Multiple imputation ensures unbiased estimates if the data are MAR (the probability of data being unobserved is dependent on the observed values but independent of unobserved ones). Because unobserved values are unknown, the missing data mechanism and hence the validity of either assumption is untestable. Nevertheless, multiple imputation presents two advantages. First, it requires a less stringent assumption for ensuring unbiased estimates. Second, if data are MCAR, both complete-case and multiple imputation estimates will be unbiased whereas, if data are MAR, complete-case analysis will be biased.

Analysis of uncertainty for incremental analysis

Sensitivity analysis is used to explore and quantify any uncertainty in the cost-effectiveness results. Three types of sensitivity analysis were undertaken: structural, scenario and probabilistic sensitivity analysis. Structural and scenario sensitivity analyses were carried out on the complete-case data set. Probabilistic sensitivity analysis was carried out in both the complete case and the multiple imputation data set.

Structural sensitivity analysis consisted of a PP analysis that classified patients according to treatment compliance at 1 year of follow-up, that is, whose management at 1 year was consistent with their original random allocation. Consequently, the PP data set consisted of the patients randomised to surgery who actually had surgery, and of the patients randomised to medical management who did not undergo surgery at 1 year. Patients randomised to medical management who had surgery might differ from those randomised to medical management who were managed medically without surgery, for several reasons. A patient's condition might have worsened, prompting surgery, or patients might have changed their preferences and wish to be taken off medication. The latter implies that, had they been screened for the study at the point in time when they had surgery, they would not have been eligible for the study. These patients would have had a preference and would not have accepted randomisation. The condition itself is complex because of its recurrent and cyclical nature (patients suffering from reflux have punctual exacerbations, which can lead them to change their preferences and request surgery). Therefore, the reasons for not complying with randomisation are likely to be a combination of the two motives (worsening of condition and change in preference). PP was chosen because it was thought to be more similar to clinical practice, where patients can experience a wait for surgery and change their preferences during this period. Any switching of treatment after 1 year is assumed to be because of a change in clinical status, which would preclude inclusion in the clinical trial.

The base-case analysis included only the costs of reflux-related GP visits and hospitalisations. Two alternative costing scenarios were tested in sensitivity analysis: including either all GP visits or all hospital use, regardless of whether they had been classified as reflux or non-reflux related.

Probabilistic sensitivity analysis attempts to quantify the joint effect of uncertainty around the costs and QALYs. Semiparametric bootstrapping was used to estimate the probability that each intervention is cost-effective for a range of cost-effectiveness threshold values. In bootstrapping, the original data are sampled with replacement to create a new data set, in order to calculate estimates of treatment effect. Repeating this process a large number of times results in a vector of replicated statistics, which ultimately provide an empirical estimate of the CIs around mean incremental costs and QALYs. The probability of an intervention being the most cost-effective is the conventional method of presenting the uncertainty around the cost-effectiveness results. The CIs around the ICER are not presented because they are difficult to interpret and

are not easy to use: a negative ICER can indicate that an intervention dominates (because it is associated with more benefits and lower costs than its comparator) or it is dominated (because it is associated with fewer benefits and higher costs).[76]

Validation

Several procedures were used to ensure the validity of the analysis. First, two statistical analysis codes (written in Stata) were developed in parallel and their results compared. Second, the code was developed by one analyst and checked independently by another. Third, the results were cross-checked in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) for a sample of the data set. Lastly, selected results were represented graphically and examined for face validity. The validity of the imputation strategy was explored by (1) analysing the data for predictors of missingness,[70] (2) comparing the distributions of the observed and imputed values graphically[70] and (3) estimation of Monte Carlo errors.[69] *Appendix 8* describes the validation process in more detail.

Results

Patient population

Complete-case analysis consisted of the patients who returned all questionnaires and completed all EQ-5D profiles. Overall, there are 172 patients in the complete-case analysis (88 randomised to medical management and 84 randomised to surgery). *Table 32* shows the numbers of questionnaires returned (includes those with some missing data) and the numbers of completed questionnaires returned for each year. As expected, the number of questionnaires returned in each year of follow-up decreases with time. The return of questionnaires does not follow a monotonic pattern, that is, patients who did not return the questionnaire for one particular year may have returned a questionnaire in subsequent years. Therefore, the number of patients in the complete-case analysis is lower than the number of completed questionnaires in year 5. The large number of patients not included in the complete-case analysis because of missing data strengthens the rationale for using the multiple imputation data sets in the base case.

Health-care resource use

Table 33 summarises yearly health-care resource use in the two trial arms according to ITT analysis. During the first year of the trial, 111 patients randomised to surgery and 10 patients randomised to medical management underwent laparoscopic fundoplication. The 111 patients who were randomised to and received surgery constituted the surgery group in the PP analysis. The 169 patients who were randomised to medical management and did not undergo surgery during the first year of follow-up constituted the medical management group in the PP analysis. In the subsequent years of follow-up there were 15 patients who underwent surgery (one patient who had been randomised to surgery and 14 patients

TABLE 32 Numbers of questionnaires returned and completed questionnaires returned and corresponding proportions per trial arm, according to ITT analysis

	Questionnaires returned, *n* (%)		Completed questionnaires,[a] *n* (%)	
Year	Surgery	Medical management	Surgery	Medical management
1	154 (87)	164 (92)	134 (75)	147 (82)
2	128 (72)	142 (79)	121 (68)	134 (75)
3	132 (74)	134 (75)	112 (63)	119 (66)
4	126 (71)	129 (72)	114 (64)	118 (66)
5	127 (71)	119 (66)	115 (65)	113 (63)
Number of patients in complete-case analysis			88 (49)	84 (47)

a Completed questionnaires means that all of the questions on health-care resource use and EQ-5D were filled in.

TABLE 33 Health-care resource use per year per trial arm, according to ITT analysis

Health-care resource	Year	Reflux-related reasons, *n*		All reasons, *n*	
		Surgery (*n* = 178)	Medical management (*n* = 179)	Surgery (*n* = 178)	Medical management (*n* = 179)
Laparoscopic fundoplication (first year)	1	111	10	N/A	N/A
Hospital overnight admissions (excluding surgery in the first year)	1	4	2	8	8
	2	1	2	8	10
	3	2	9	6	10
	4	2	9	2	10
	5	0	8	1	11
	Total	9	30	25	49
Hospital day admissions	1	22	24	40	53
	2	5	4	23	24
	3	4	6	4	10
	4	12	9	13	11
	5	4	11	7	14
	Total	47	54	87	112
Visits to and from GP	1	110	103	394	376
	2	34	115	269	373
	3	38	99	381	386
	4	55	126	422	469
	5	36	119	404	370
	Total	273	562	1870	1974
Number of patients on reflux-related medication	1	58	148	N/A	N/A
	2	48	124	N/A	N/A
	3	51	113	N/A	N/A
	4	51	106	N/A	N/A
	5	56	98	N/A	N/A

N/A, not applicable.

who had been randomised to medical management). These patients are included in the overnight hospital admissions category. Patients randomised to medical management reported more hospital and GP visits than the surgery patients over the 5 years of follow-up.

Table 34 shows the costs associated with health-care use according to ITT analysis for all available cases (see *Appendix 9* for corresponding table for PP). All available cases uses data from all questionnaires returned at each time point. Per annum costs and costs per category refer to all available data, that is, to all participants who returned the questionnaire for that particular year or for that particular category. Therefore, the sum of the costs per category is different from the sum of the costs per annum. Similarly, total costs for complete-case analysis do not correspond to the sum of the costs per category or to the sum of the costs per annum because complete case is a subset of all available data because of

the non-monotone missing data pattern. Total costs for complete-case analysis refer to the patients who returned all questionnaires and completed all EQ-5D profiles (84 surgery patients and 88 medical management patients).

Patients randomised to medical management accumulate lower costs than patients randomised to surgery. *Table 34* indicates that surgery patients accrued a large proportion of the total costs in the first year, and accumulated lower costs during the remaining 4-year follow-up than the medical management group. In contrast, the costs accrued by medical management patients are evenly distributed across the duration of the trial. These results suggest that the cost trend in medical management patients is steeper than in surgery patients; hence, that cumulative costs in medical management patients tend to increase at a greater rate than in surgery patients. Costs associated with surgery were the major cost driver for the surgery group. Costs associated with reflux-related medication were significantly greater for the medical management group than for the surgery group. Costs associated with admissions to hospital and GP visits were not statistically significantly different between the two groups. Surgery during years 2–5 is accounted for in the overnight hospital admissions. There were a few crossovers from medical management to surgery from year 2; hence, the difference in costs associated with overnight hospital admissions between the two treatment groups is small. These results suggest that patients undergoing surgery in subsequent years are not a major cost driver in determining the cost-effectiveness of surgery.

Quality-adjusted life-years

Table 35 summarises the EQ-5D scores reported at each follow-up point for all available cases (see *Appendix 9* for the corresponding table for PP). All available cases uses data from all questionnaires returned at each time point. The surgery group appears to have better HRQoL than the medical

TABLE 34 Costs associated with resource use for all available cases, discounted from year 2 at 3.5%, according to ITT analysis

Returned questionnaires in each year			Mean (SD) resource-use cost (£)		Incremental mean cost (cost surgery – cost medical management) (95% CI[a]) (£)
Surgery	Medical management	Year	Surgery	Medical management	
154	164	1	2500.75 (1697.99)	559.62 (1006.81)	1941.13 (1621.43 to 2260.83)
128	142	2	94.15 (317.63)	150.96 (356.57)	–56.81 (–138.08 to 24.46)
132	134	3	94.35 (340.33)	276.41 (894.16)	–182.05 (–345.87 to –18.24)
126	129	4	111.41 (394.00)	303.50 (1337.26)	–192.09 (–436.56 to 52.28)
127	119	5	58.38 (178.58)	234.03 (629.33)	–175.65 (–290.26 to –61.03)
Cost category					
Surgery in year 1[b]			1734.05 (1407.58)	164.31 (644.63)	1569.74 (1342.05 to 1797.42)
Reflux-related hospital night admissions			343.82 (1176.05)	302.34 (818.41)	41.47 (–247.48 to 330.42)
Reflux-related hospital day admissions			221.67 (633.61)	250.35 (631.37)	–28.68 (–209.24 to 151.87)
Reflux-related GP visits			127.18 (178.96)	200.13 (462.53)	–72.95 (–173.26 to 27.35)
Medication			121.34 (265.05)	365.70 (517.05)	–244.35 (–361.82 to –126.89)

a CIs estimated using OLS regression.

b Only surgery occurring during the first year of the trial was included here. Laparoscopic fundoplication occurring in the subsequent years of the trial has been included in reflux-related hospital admissions. For the 10 medical management patients who had surgery in the first year of the trial, the average (SD) cost was £2679 (£126). For the 111 surgery patients who had surgery in the first year of the trial, the average (SD) cost was £2798 (£501).

TABLE 35 Health-related quality of life (EQ-5D) for all available cases according to ITT analysis

Completed questionnaires returned at each time point			Mean (SD) EQ-5D		
Surgery ($n = 178$[a])	Medical management ($n = 179$[a])	Follow-up	Surgery	Medical management	Difference in EQ-5D (surgery – medical management) (95% CI)[b,c]
171	173	Baseline	0.7107 (0.2581)	0.7201 (0.2545)	–0.0094 (–0.0638 to 0.0445)
149	153	3 months	0.7881 (0.2328)	0.6894 (0.3012)	0.0987 (0.0376 to 0.1597)
152	164	Year 1	0.7537 (0.2468)	0.7097 (0.2715)	0.0440 (–0.0136 to 0.1016)
122	138	Year 2	0.7619 (0.2718)	0.7172 (0.3127)	0.0447 (–0.0273 to 0.1167)
129	132	Year 3	0.8034 (0.2312)	0.7474 (0.2621)	0.0560 (–0.0043 to 0.1163)
125	127	Year 4	0.7713 (0.2438)	0.7544 (0.2719)	0.0169 (–0.0472 to 0.0810)
124	117	Year 5	0.7743 (0.2590)	0.7612 (0.2815)	0.0131 (–0.0555 to 0.0817)

a *n* refers to the number of patients originally randomised to each trial arm.
b CIs estimated using OLS regression.
c Unadjusted for baseline EQ-5D.

management group, despite starting from a lower baseline EQ-5D on average (0.7201 in the medical management group and 0.7107 in the surgery group). The difference in HRQoL between the two treatment groups decreased with time. This may be due to patients randomised to medical management undergoing surgery throughout the follow-up period and/or to diminishing treatment effect over time.

Comparison of costs and quality-adjusted life-years between multiple imputation and complete case

Table 36 shows the comparison of the total costs per year between the complete-case data set and the multiple imputation results. Complete case includes only those participants who returned all questionnaires and fully completed the EQ-5D questionnaires. The similarity of both the means and the CIs provides some reassurance of the validity of the multiple imputation model. The distribution of costs and EQ-5D scores in the imputed data sets matches reasonably well the distribution of the original data (see *Appendix 8* for details). Furthermore, the Monte Carlo errors are <15% of the coefficient and CI estimates, suggesting that 100 imputations are sufficient to ensure reproducibility and statistical efficiency.[69]

For both the complete-case and multiple imputation data sets, the participants randomised to laparoscopic fundoplication accrued greater costs but also reported greater HRQoL than participants randomised to continued medical management. The 95% CI for mean incremental QALYs crosses zero for the unadjusted for baseline estimates, whereas it remains above zero for the adjusted values. This result reflects the baseline imbalance in mean utility between treatment groups. Therefore, these results strongly indicated that surgery is associated with a greater QALY improvement than medical management. The sum of the differences in EQ-5D for the ITT groups does not correspond to the incremental mean QALYs because of the effect of discounting.

Cost-effectiveness

The results of the incremental analysis suggest that laparoscopic fundoplication is a cost-effective strategy for GORD patients eligible for the REFLUX trial (*Table 37*). The results for the complete-case analysis concur with those for the multiple imputation data set; across adjusted and unadjusted ICER for baseline EQ-5D, ICERs range between £5468 and £8410, well below conventional cost-effectiveness thresholds of £20,000 and £30,000 per additional QALY. For both data sets (complete case and multiple imputation), the probability of surgery being the more cost-effective intervention is >0.82 for incremental analysis

TABLE 36 Comparison between the complete-case and multiple imputation data sets for costs and HRQoL, according to ITT allocation

	Incremental mean cost (cost surgery–cost medical management) (95% CI[a]) (£)		Difference in mean EQ-5D (surgery–medical management) (95% CI[a])		
Year	Complete case	Multiple imputation	Follow-up	Complete case	Multiple imputation
N/A	N/A	N/A	Baseline	–0.0388 (–0.1083 to 0.0308)	–0.0091 (–0.0615 to 0.0433)
N/A	N/A	N/A	3 months	0.0848 (0.0122 to 0.1573)	0.0825 (0.0232 to 0.0142)
1	2197.14 (1779.67 to 2614.61)	1958.46 (1617.31 to 2299.62)	Year 1	0.0519 (–0.0198 to 0.1237)	0.0407 (–0.0150 to 0.0963)
2	–139.14 (–237.02 to –41.26)	–44.58 (–129.68 to 40.52)	Year 2	0.0467 (–0.0356 to 0.1289)	0.0445 (–0.0218 to 0.1108)
3	–193.44 (–361.94 to –24.93)	–127.42 (–306.36 to 51.52)	Year 3	0.0508 (–0.0195 to 0.1211)	0.0454 (–0.0150 to 0.1057)
4	–37.66 (–173.02 to 97.70)	–144.61 (–374.78 to 85.56)	Year 4	0.0324 (–0.0395 to 0.1044)	0.0260 (–0.0356 to 0.0875)
5	–165.12 (–304.74 to –25.50)	–123.90 (–236.56 to –11.24)	Year 5	–0.0095 (–0.0871 to 0.0680)	0.0294 (–0.0358 to 0.0945)
Total cost	1661.78 (1130.00 to 2193.55)	1517.95 (1006.49 to 2029.41)	Total QALYs (unadjusted)	0.1976 (–0.0857 to 0.4810)	0.1948 (–0.0356 to 0.4251)
Monte Carlo error[b]		10.16	Monte Carlo error[b]		0.0034
			Total QALYs (adjusted)[c]	0.3039 (0.0928 to 0.5150)	0.2160 (0.0205 to 0.4115)
			Monte Carlo error[b]		0.0035

N/A, not applicable (due to costs being calculated using the area under the curve method).

a 95% CI estimated with OLS regression and the 'mim' command.

b Monte Carlo error – SD across repeated runs of the same imputation procedure with the same data.[69]

c Total QALYs adjusted for baseline EQ-5D.

TABLE 37 Incremental analysis for the ITT analysis at 5 years of follow-up for the complete-case and multiple imputation data sets

Data set	Adjustment for baseline EQ-5D?	Incremental mean costs (£) (95% CI)	Incremental mean QALYs (95% CI)	ICER (£)	Probability cost-effective at £20,000 per QALY[a]	Probability cost-effective at £30,000 per QALY[a]
Complete case	No – unadjusted QALYs	1661.78 (1130.00 to 2193.55)	0.1976 (–0.0857 to 0.4810)	8409.82	0.828	0.866
	Yes – adjusted QALYs	1661.78 (1130.00 to 2193.55)	0.3039 (0.0928 to 0.5150)	5468.36	0.989	0.996
Multiple Imputation	No – unadjusted QALYs	1517.95 (1006.49 to 2029.41)	0.1948 (–0.0356 to 0.4251)	7792.35	0.861	0.906
	Yes – adjusted QALYs	1517.95 (1006.49 to 2029.41)	0.2160 (0.0205 to 0.4115)	7027.55	0.932	0.962

a Probability of intervention being cost-effective calculated with semiparametric bootstrapping.

unadjusted for baseline EQ-5D and >0.93 once incremental QALYs are adjusted for baseline EQ-5D. In the ITT analysis the ICER is higher for the multiple imputed data than for the complete case if QALYs are adjusted for baseline EQ-5D, but lower if QALYs are unadjusted. This might reflect the effect of having baseline EQ-5D in the prediction model, which would preclude the need for adjustment.

Figure 20 shows how the NMB associated with laparoscopic fundoplication increases with the duration of the trial. This reflects the increase in costs associated with the medical group, which offsets the initial investment made in laparoscopic fundoplication in the surgery group.

Structural sensitivity analysis: per-protocol status for the complete case

Structural sensitivity analysis consisted of PP status at 1 year for the complete case. In the PP analysis patients are classified according to the treatment actually received at 1 year of follow-up. The PP group consists of 111 patients who were randomised to surgery and who actually had surgery during the first year of the trial and 169 patients who were randomised to medical management and who did not undergo surgery during this time period. However, complete-case data exist only for 84 medical management patients and 66 laparoscopic fundoplication patients. *Appendix 9* presents detailed results for costs and HRQoL according to PP analysis. As expected, patients who actually had surgery have higher costs than patients who did not undergo surgery, regardless of their randomisation. *Table 38* summarises the incremental results of the PP analysis. Similar to the ITT analysis, the surgical policy is likely to be cost-effective at conventional (NICE) thresholds for cost-effectiveness. The incremental costs are higher and the incremental QALYs lower for the PP analysis (for surgery compared with medical management) than for the ITT analysis if no adjustment is made for baseline imbalances in EQ-5D. Therefore, the ICER is also greater (surgery is less cost-effective than suggested by the ITT analysis). Once total QALYs are adjusted for baseline EQ-5D, however, the incremental mean QALYs increase substantially and the ICER is reduced. Nevertheless, the adjusted ICER in the ITT analysis is lower than that in the PP analysis by around £2000.

Scenario sensitivity analysis: all general practitioner and all hospital costs for complete case

The results of the scenario analyses strengthen the case for the surgical policy (*Table 39*). For scenario 1, replacing reflux-related GP costs by all GP costs, the ICER increased slightly in relation to the base case. Nevertheless, the ICER remains well below conventional thresholds and the probability of surgery being cost-effective is >0.83, for both adjusted and unadjusted analyses. In scenario 2, replacing reflux-related

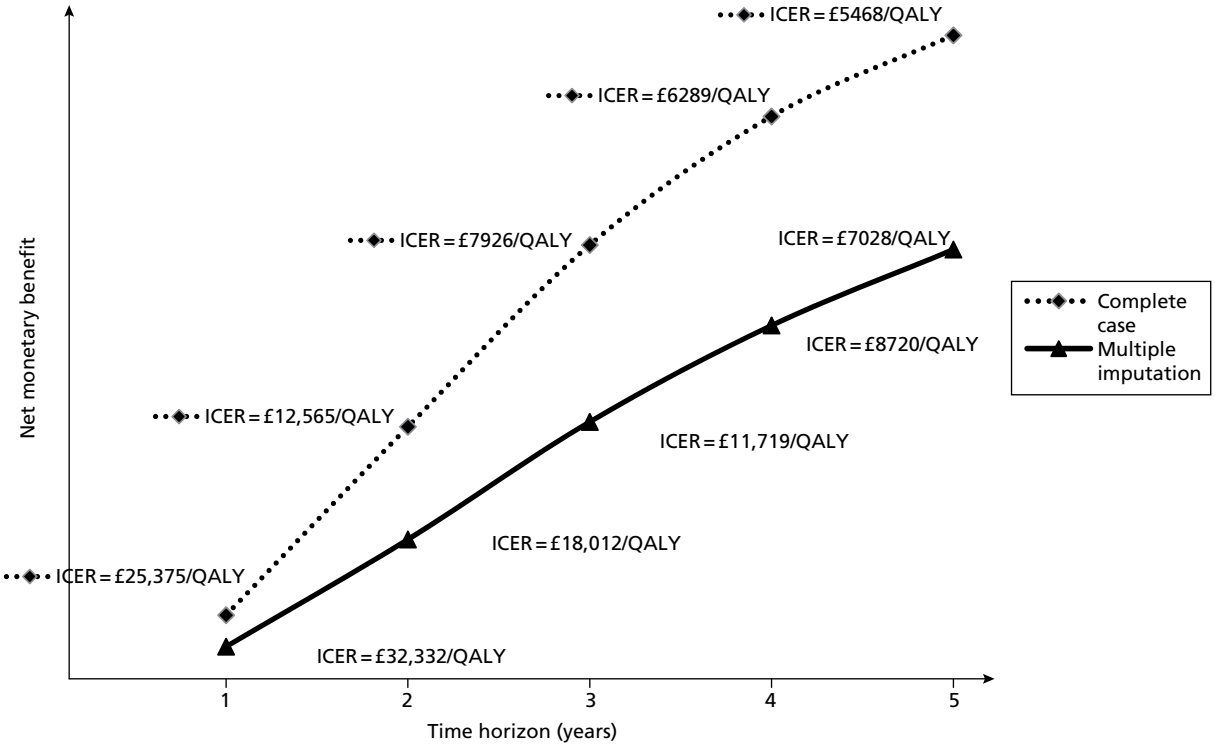


FIGURE 20 Net monetary benefit (incremental QALYs × £20,000 per QALY − incremental costs) over the duration of the REFLUX trial for the multiple imputation and complete-case data sets (QALYs adjusted by baseline EQ-5D).

hospital costs by all hospital costs, medical management was 'dominated' by the surgical policy because of this intervention being associated with greater benefits in terms of QALYs and lower costs. For this scenario the probability of surgery being cost-effective was > 0.93.

Sensitivity analysis for the multiple imputation model: departure from missing at random assumption

The multiple imputation procedure assumes that the individuals who completed and returned all questionnaires are similar to the individuals who did not, conditional on their observed characteristics (MAR assumption).[69,79] However, this may not be the case: patients who did not return a questionnaire may have experienced worse HRQoL and accrued higher health service costs, or vice versa. Sensitivity analysis on the multiple imputation model tested how sensitive the cost-effectiveness results are to the MAR assumption. *Figure 21* represents the change in NMB adjusted for baseline EQ-5D as costs and QALYs are varied in patients with missing data. The origin, marked as 'base case', refers to the incremental results from the multiple imputed data set (ICER = £7028 per additional QALY). The right quadrant plots NMB after increasing the total costs in steps of 10% for patients for whom there was missing data, for both treatment groups and for surgery-allocated patients. The left quadrant plots NMB after decreasing total QALYs in similar fashion. Positive values for NMB indicate that surgery is cost-effective; negative values indicate that surgery is not cost-effective for a threshold of £20,000 per additional QALY.

The cost-effectiveness of surgery is relatively insensitive to any increase in costs; the NMB changes little if costs are increased for patients with missing data in both treatment groups and if costs are increased just for surgery-allocated patients with missing data. A similar result is observed for the reduction in total QALYs for all patients with missing data. In contrast, the cost-effectiveness of surgery is highly sensitive if it is assumed that surgery-allocated patients with missing data experience lower HRQoL than patients with complete data. A 10% decrease in QALYs for patients randomised to surgery with missing data results in NMB decreasing to negative values. This scenario shows that missing data can have an impact on the results under certain conditions. It is impossible to empirically confirm or refute the scenario from

TABLE 38 Incremental analysis for the PP analysis at 5 years of follow-up for the complete-case data set

Adjustment for baseline EQ-5D	Incremental mean costs (95% CI) (£)	Incremental mean QALYs (95% CI)	ICER (£)	Probability cost-effective at £20,000 per QALY[a]	Probability cost-effective at £30,000 per QALY[a]
Unadjusted QALYs	2323.77 (1799.90 to 2847.65)	0.1782 (–0.1316 to 0.4879)	13,043.90	0.672	0.747
Adjusted QALYs	2323.77 (1799.90 to 2847.65)	0.3200 (0.0837 to 0.5562)	7262.85	0.957	0.983

a Probability of intervention being cost-effective was calculated with semiparametric bootstrapping.

TABLE 39 Incremental analysis for the scenario sensitivity analysis at 5 years of follow-up for the complete-case data set

Sensitivity analysis	Adjustment for baseline EQ-5D?	Incremental mean costs (95% CI) (£)	Incremental mean QALYs (95% CI)	ICER (£)	Probability cost-effective at £20,000 per QALY[a]	Probability cost-effective at £30,000 per QALY[a]
Scenario 1: all GP costs	No – unadjusted QALYs	1685.60 (1103.97 to 2267.23)	0.2125 (–0.0748 to 0.4998)	7932.23	0.826	0.863
	Yes – adjusted QALYs	1685.60 (1103.97 to 2267.23)	0.3191 (0.1061 to 0.5321)	5282.36	0.987	0.994
Scenario 2: all hospital costs	No – unadjusted QALYs	–262.72 (–860.08 to 334.65)	0.2125 (–0.0748 to 0.4998)	Medical management dominated	0.930	0.928
	Yes – adjusted QALYs	–£262.72 (–860.08 to 334.65)	0.3191 (0.1061 to 0.5321)	Medical management dominated	0.999	0.999

a Probability of intervention being cost-effective calculated with semiparametric bootstrapping.

the data in the trial, but it could be considered an extreme case. It seems improbable in practice that surgical patients with poor quality of life are less likely to respond to follow-up questionnaires than similar participants undergoing medical management.

Conclusion

The results of the within-trial economic analysis suggest that laparoscopic fundoplication is the more cost-effective option for the management of the sorts of patients suffering from GORD who were eligible for the REFLUX trial. The ICER for the ITT approach in the complete case was between £5468 and £8410 per additional QALY, and for the multiple imputed data set was between £7028 and £7792 per additional QALY, depending on whether QALYs are unadjusted or adjusted for baseline. Adjusted results are likely to be more reflective of the improvement in HRQoL associated with surgery. The probability of surgery being cost-effective was >0.80 for all analyses. The results are robust to the scenario analyses testing assumptions regarding resource-use and missing data mechanism apart from when surgery-allocated patients with missing data were assumed to experience lower HRQoL than other patients. In all scenarios the ICERs were similar to the base case ICERs and well below NICE cost-effectiveness thresholds.

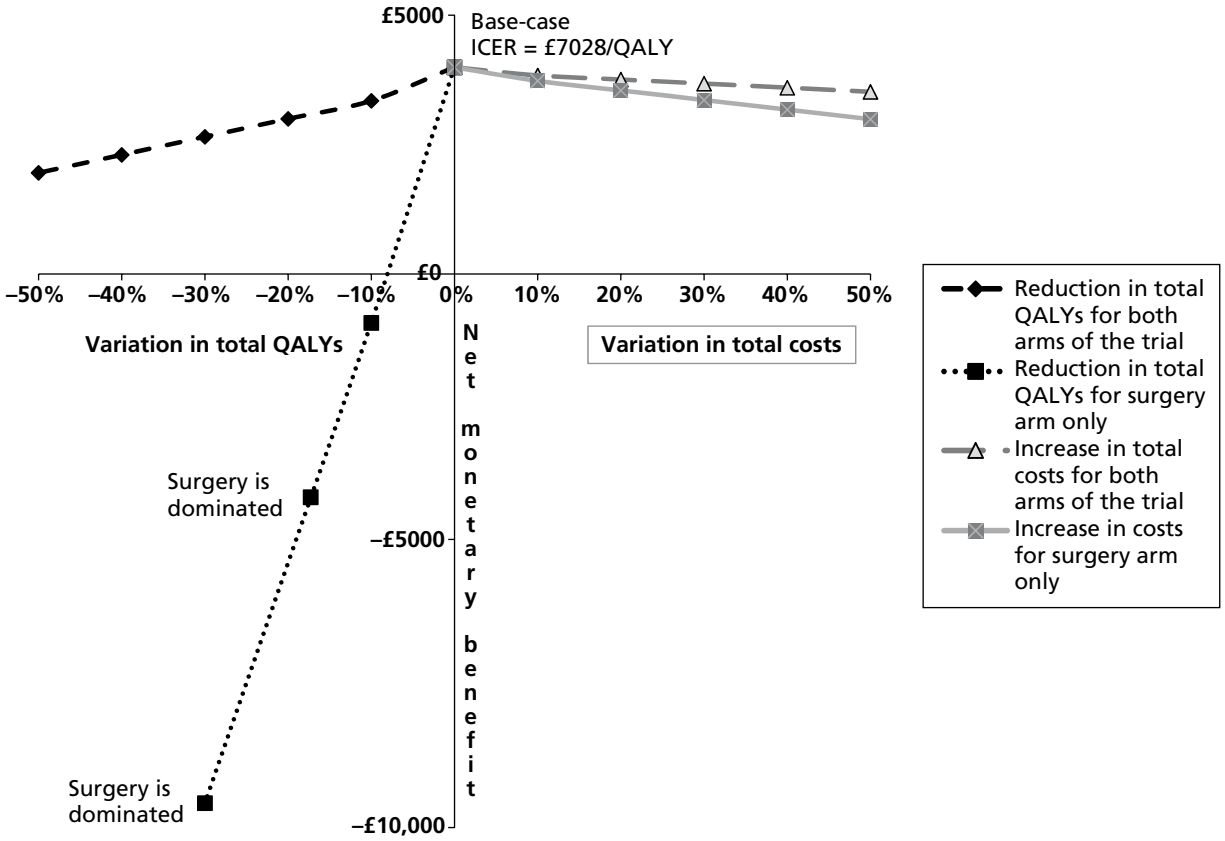


FIGURE 21 Net monetary benefit (incremental QALYs adjusted for baseline EQ-5D × £20,000 per QALY – incremental costs) over variation in total costs and total QALYs in the multiple imputed data set.

Validation of within-trial (5-year) analysis and exploration of the need for a long-term model

Introduction

The within-trial analysis found that surgery was cost-effective over a 5-year time horizon. A sufficient condition for surgery to be unambiguously cost-effective over a longer term is that, in each year after 5 years, HRQoL is lower and costs are the same or increasing faster in the medical group than in the surgical group. The results from both the multiple imputation and the complete-case analysis suggest that surgery is likely to be a cost-effective alternative over the longer term. Based on the ITT analyses undertaken so far, it is unlikely that mean HRQoL in patients who had surgery will become lower than that in patients on medical management after 5 years, and it is also very unlikely that mean annual costs incurred by surgery patients will exceed those incurred by medical management patients. If these results are robust, then there is no need to develop an economic model to extrapolate the 5-year results over a longer time horizon. Surgery would simply become more cost-effective over time.

This section develops a statistical model to investigate whether or not the results obtained in the within-trial economic analysis are robust to alternative assumptions and methods, and uses the results to consider whether or not the evidence supports this sufficient condition over the longer term.

Methods

Overview

The aim of this analysis was to estimate the difference in costs and the difference in HRQoL (measured with the EQ-5D) between the surgical and medical management randomised groups and describe how this difference evolves over time. A simple way of doing this would be to estimate the difference in costs and outcomes at each time point independently. The results of this analysis were shown in *Table 34* (for costs) and *Table 35* (for EQ-5D). These showed that costs were greater in the surgical group in the first year but greater in the medical group thereafter. EQ-5D tended to be higher in the surgical group in years 4 and 5 but the CIs crossed zero. There are two main limitations of this simple analysis:

1. The outcomes at each time point are unlikely to be independent. If the outcomes at one time point are correlated with those at other time points this analysis may lead to biased estimates of standard errors.
2. The analysis does not take account of missing data. If missing data are not MCAR then this analysis may lead to biased estimates of the mean of the coefficients.

The multiple imputation accounts for the correlation of responses from the same individuals and for the missing data (see *Table 36*). However, the validity of this analysis depends on the correct specification of the equations used to impute the missing data. Moreover, other regression-based methods are available for handling missing data in longitudinal studies, principally mixed models, and results may be sensitive to the methods used. This section uses a mixed model to handle the missing data and compares predicted outcomes with those using multiple imputation.

Mixed models

A mixed model is a regression-based method for handling continuous data that is measured at more than one time point during follow-up. It allows estimation of treatment effects under the assumption that the data are MAR, that is, dropout may depend on intermediate values. Analysing each time period separately assumes that dropouts are MCAR, a stronger assumption. A mixed model uses all of the observed data. Individuals who dropped out after providing intermediate data contribute to the estimation of the final outcomes. This analysis has the same aims as multiple imputation but uses a different method and with different assumptions. Therefore, it can also be viewed as a sensitivity analysis to test the robustness of the multiple imputation.

The mixed model can be written as:

Where for an individual i,

$Y_i = \alpha + \beta R_i + X_i + e_i$, $e_i \sim \text{MVN}(0, \Sigma)$
R_i = randomised group
Y_i = vector of all outcomes (at times 1…T)
X_i = vector of covariates

The variance–covariance matrix Σ is unstructured, that is, no prior assumptions are made about the values of the correlations. Separate models are fitted for costs and for EQ-5D. Baseline values of the EQ-5D are included as an 'outcome' (i.e. at $t = 1$). Dummies representing time points 1 to T were included as covariates X_i. Treatment effects are included as time*randomised group interactions although no treatment effect at baseline is allowed. No other covariates are included in the model.

Results

Costs

Figure 22 shows the difference in costs (excluding initial surgery) in years 1–5. Mean costs are greater in the medical management arm of the trial after the first year and the CIs only just cross zero. These results are very similar to those of *Table 34*.

European Quality of Life-5 Dimensions

Figure 23 shows the difference in EQ-5D at 3 months and in years 1–5. Mean HRQoL tends to be greater in the surgical group during the trial, although the CIs cross zero in some periods. These results are very similar to those of *Table 35*.

Conclusion

The results of the mixed model (taking account of correlations and missing data) are very similar to those of the complete-case analyses (which assumed that data at different time points were independent) and the multiple imputation (see *Table 36*). All of these analyses show that follow-up costs are significantly greater in the medical management arm of the trial (because of greater reflux-related hospital admissions, GP visits and use of medication). The analyses also show that surgery tends to be more effective, in terms of HRQoL, than medical management over the 5-year follow-up. Although this treatment difference appears to weaken over time, there is no reason to expect that surgery will become less effective with a longer follow-up. Consequently, the evidence suggests that the cost-effectiveness of laparoscopic fundoplication will not diminish if measured over a longer follow-up time. Nevertheless, there is uncertainty surrounding these conclusions because of the large proportion of missing data.

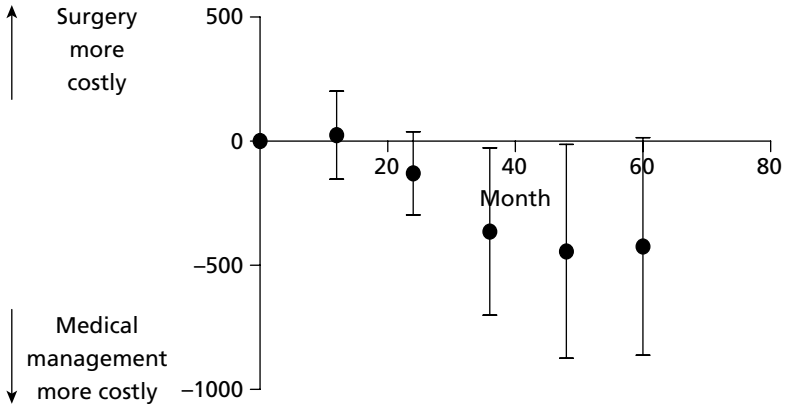


FIGURE 22 Difference in costs (£) excluding initial surgery (mean, 95% CI).

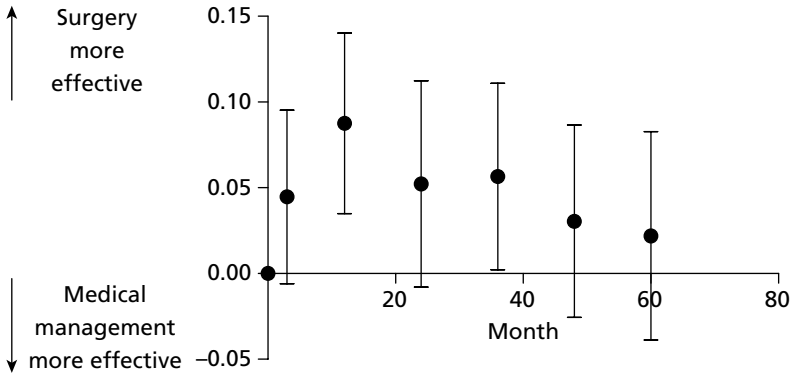


FIGURE 23 Difference in EQ-5D, adjusted for baseline (mean, 95% CI).

Discussion

The results of the cost-effectiveness analysis strongly suggest that a policy of offering laparoscopic fundoplication to people with GORD who require long-term PPI treatment for symptom control is more cost-effective than continuing to manage them with PPIs (with selective use of surgery if symptoms are poorly controlled), assuming that the cost-effectiveness thresholds used by NICE (£20,000–30,000 per QALY) are appropriate for the NHS. Surgery represents a greater initial investment and lower medium-term costs, whereas costs associated with medical management remain relatively constant or slightly increase over time. The difference in HRQoL achieved with surgery is sustained over 5 years, although the results indicate that mean EQ-5D scores for surgery and medical management tended to converge (as discussed in *Chapter 3*, in part this reflects later surgery in patients randomised to medical management). The ICER favours surgery when incremental QALYs are both adjusted and unadjusted for baseline EQ-5D. Nevertheless, adjusted incremental QALYs are likely to be a more reliable estimate of treatment effect as they account for differences in baseline utility. Patients randomised to medical management reported higher baseline utility than patients randomised to surgery. Failure to adjust for these baseline differences could result in a biased ICER, as discussed elsewhere.[62] The results from the multiple imputed data set are likely to be more accurate than the results from the complete-case analysis because of the large number of patients with incomplete data (>50%). Therefore, multiple imputation was chosen for the base case. Nevertheless, the results are similar across the data sets and laparoscopic surgery is the more cost-effective intervention for both.

There is little uncertainty regarding the cost-effectiveness results once adjustment for EQ-5D at baseline is performed. The probability of surgery being cost-effective ranged between 0.932 and 0.999 for the base case and across the scenarios tested. Furthermore, it is clear from the results of the scenario analysis that the base-case results are robust to alternative costing assumptions. The PP analysis is used to test whether or not the ITT analysis is potentially misleading because of the dilution of treatment effect (some patients randomised to surgery did not have surgery and some patients randomised to medical management actually had surgery). The PP analysis has the advantage of mimicking clinical practice and could be thought to be more relevant to decision-makers. However, the PP analysis is not without its limitations. First, and as with any PP analysis, it is sensitive to selection bias because of breaking randomisation. Second, the PP analysis may still underestimate the effect of surgery because patients having surgery between 2 and 5 years are counted as medical management. Third, the PP analysis is actually a subset of the ITT groups, which further reduces the data set. For these reasons, the ITT results are likely to be more reliable. It is important to characterise any uncertainty in the analysis as failure to do so can result in inaccurate estimates of cost-effectiveness, particularly when costs and benefits are highly skewed.[82] In addition, any analyses of uncertainty can help to illustrate where caution should be exercised when interpreting the results of a cost-effectiveness analysis. The results of the sensitivity analyses suggest that the uncertainty is likely to be driven by HRQoL. If QALYs for randomised surgery patients with missing data are reduced, surgery may no longer be cost-effective.

For the within-trial analysis no assumptions are needed about the longer-term effectiveness and costs associated with surgery and medical management. However, the within-trial analysis has some disadvantages. First, it does not account for any differences in costs and QALYs that may be expected over the longer term (>5 years), which could be due to differences in recurrence/relapse, medication use, NHS service utilisation or HRQoL. Second, it uses data only from the REFLUX trial and does not consider other sources of evidence. Third, only a limited range of sensitivity analyses was possible. Finally, the large proportion of missing observations required an assumption regarding the mechanism of missing data, which may have some impact on the cost-effectiveness estimates. The exploration of the need for a longer-term model aimed to address the first limitation of the within-trial analysis. A mixed model was used to examine the trend in the difference in costs and the difference in QALYs between treatment groups over time.

No evidence was found to suggest that the cost-effectiveness of surgery diminishes over a longer follow-up time. Both multiple imputation and mixed models are commonly used methods to handle missing data. Multiple imputation was used in the previous section because, by imputing missing data, it naturally allows the estimation of the total cost and total QALYs for each patient in the trial. Furthermore, it can handle correlation between several outcomes (in this case costs and QALYs) as well as correlation between outcomes over time. Mixed models do not explicitly impute missing data but adjust the estimates of the differences between treatment groups at each discrete follow-up time to take account of the missing data. The approach therefore offers an alternative method to multiple imputation to examine trends in the difference in costs and the difference in QALYs between treatment groups over time. Because the analyses using multiple imputation and mixed models agree, we can have more confidence that the results are valid and that surgery is the most cost-effective intervention.

A number of other studies have quantified the cost-effectiveness of laparoscopic fundoplication and medical management. Not all of these, however, use a common metric (such as QALYs) to measure benefits. Of those studies quantifying the benefits associated with the two treatments using QALYs, the ICER for surgery ranged from £180 to £49,000. There are a number of key differences between the methodologies used in the studies, which limit the extent to which comparisons can be made between results. Importantly, not all of the studies are based on within-trial analysis; in fact, only two are: those by Grant *et al.*[1] and Goeree *et al.*[46] The remainder use modelling techniques to either extrapolate short-term trial results over the longer term or pool available evidence to generate estimates of costs and outcomes. Comparing the results from Grant *et al.*[1] with those from Goeree *et al.*[46] we can see that there are quite significant differences in the estimates of cost per QALY, from £19,000 in Grant *et al.*[1] to £49,000 in Goeree *et al*.[46] This difference is primarily driven by the difference in QALYs. In Goeree *et al.*[46] the EQ-5D score is actually lower in the surgical group than in the medical management group (this is unadjusted for baseline imbalances) whereas the HUI3 score is higher for the surgical group than the medical management group. The reason for the difference between the EQ-5D and the HUI3 scores is not discussed in the paper. The cost differences in the two studies were similar. Comparing the results from Goeree *et al.*[46] with those from the updated trial analysis we see even starker contrast between the ICERs produced (£7028 vs £49,000). Again, this is driven by the differences in EQ-5D scores observed throughout the trial period. The EQ-5D scores in the REFLUX 5-year analysis are consistently higher in the surgery group than in the medical management group, although there is a tendency for convergence towards the end of the follow-up period. Further research is required to look at why the trials produce such different results using the EQ-5D.

Other considerations

The generalisability of these findings to the GORD population in the UK is difficult to ascertain because the proportion of GORD patients meeting the entry criteria for this trial is uncertain. The surgeons participating in the trial may be more proficient in the procedure than those in actual practice. Furthermore, capacity constraints may limit the offer of the surgery policy to all potentially eligible patients.

Chapter 6 Conclusions

In the report of the first phase of the REFLUX trial[1] we concluded that, among the sorts of patients recruited to the trial, laparoscopic fundoplication 'significantly increases general and reflux related QoL measures, at least up to 12 months after surgery'. There was, however, considerable uncertainty about cost-effectiveness, largely because the follow-up period was so short. Varying plausible assumptions about the longer-term effects of surgery, particularly in terms of QALYs gained and costs of medication, led to markedly differing results. This was the basis for this second phase of the trial, in which follow-up has been extended out to a time equivalent to 5 years after surgery.

The trial has a pragmatic design and compared two policies for managing GORD, rather than directly comparing surgery with PPI therapy. This is the basis for the primary analyses being based on the ITT principle as this directly compares the policies. The first policy can be characterised as relatively early surgery for most eligible patients but with the option to take medication if considered helpful, irrespective of whether or not surgery had been performed. The second policy can be described as medical management as appropriate with 'delayed' surgery in selected cases. Hence, we have not made an assumption that those taking medication after surgery are 'failures'. In our view, although surgery may have improved symptoms, the addition of PPIs may give further improvement and hence should be considered to be a component of both policies.

In contrast to the other large randomised trial (the LOTUS trial,[48] discussed in *Chapter 4*), whose primary outcome was 'treatment failure', we chose patient-reported outcome measures as our primary and main secondary outcome measures. The advantage is that they provide a 'common currency' across the two trial policies and do not depend on clinical judgements (as 'treatment failures' do). There is, however, a concern that completion of the patient-reported outcome questionnaires may be influenced by the nature of the management received. We had a reminder of this in the early stages of our trial. The DMC noticed an imbalance in baseline scores of the first few patients randomised, but not in other descriptive characteristics. It seemed that this might have been due to completion of the form after the allocation was known (although it could still have been due to chance); once it was made a requirement that the form had to be filled in before the allocation was known, however, this discrepancy disappeared. We believe that a strength of the long-tem follow-up as reported here is that, as the time from the differentiating event (surgery or no surgery) gets increasingly long, the possibility of such reporting bias becomes remote. Protection was also provided by the partially randomised patient preference design: the randomised component was limited to patients who were uncertain which treatment to choose while those who had strong views were enrolled into the preference groups.

We designed the trial with the aim of making the management policies as similar as possible to normal NHS care. So, for example, a large number of centres were involved (both teaching and non-teaching hospitals); recruitment was based on gastroenterologist–upper gastrointestinal surgeon partnerships; surgeons chose the type of fundoplication and other aspects of the procedure; after optimisation of medical management in secondary care, all subsequent medical care was in general practice; there was no requirement for extra tests or hospital visits; and simple entry criteria identified people with chronic troublesome GORD symptoms that required anti-reflux medication for reasonable control suitable for either policy (average age 46 years). The results should, therefore, be easily generalisable to standard NHS care.

The one area in which we think the trial did not 'mimic' usual care is in the relatively low proportion of those allocated surgery who actually had surgery (62%; see *Table 10*). There are reasons for thinking that the unusual circumstances of a randomised trial comparing medical management with surgery were partly responsible for the large proportion who did not have surgery. We think the rate (84%) in the preference group is likely to be more indicative of 'normal' acceptance rates. For this reason we undertook secondary

adjusted treatment received analyses aimed to compensate for this. These analyses are likely to give a better estimate of differential effects in usual care, but because they depart from the randomised groups and hence may be prone to bias they should be treated with appropriate caution. We also explored this issue through post hoc analyses stratified by whether or not those allocated surgery actually had fundoplication. This showed (see *Figure 18*) that those who had surgery had lower baseline REFLUX scores (worse symptoms) than those who did not have an operation, but that, following surgery, their scores were consistently higher that those who did not have surgery.

Despite our best attempts to retain the cohort of participants there has been some attrition over the course of the follow-up period. The response rate of 69% at 5 years can be considered satisfactory in a study of this type and is similar to the rate in the LOTUS trial (67%).[48] The rate in the REFLUX trial reflects the decision among some participants to withdraw, but with high levels of return among those remaining. Responders did differ from non-responders but we used analysis techniques to make the most of the available data (repeated measures and imputation), and the responders in the two randomised groups were generally reassuringly similar in respect of baseline characteristics.

The new results provide clear evidence of a sustained greater improvement in GORD-related QoL in the group randomised to surgery. The results also suggest sustained benefit in respect of generic health-related measures of QoL, although the differences attenuate over time and are not statistically significant at 5 years. In these respects the REFLUX trial is in line with the results of the other three randomised trials that have compared laparoscopic surgery with medical management. The worse the symptoms at entry (the lower the score at baseline), the greater are the benefits of surgery.

By 5 years, 24 (13%) of the participants randomly allocated to medical management had undergone anti-reflux surgery. Exploratory analyses (see *Figure 19*) showed that, as a group, these 24 had low REFLUX questionnaire scores (worse symptoms) at trial entry, which subsequently improved markedly after surgery. Hence, this group is at least a contributory factor to the narrowing of differences between the randomised groups over time (see, in particular, *Figures 3* and *17*) and a reason for thinking that the ITT-based analyses comparing the two management policies are likely to underestimate the effects of surgery.

The follow-up has clarified the rates of longer-term use of PPI medication in both policies. In the randomised medical group, 87% were taking medication at 1 year, falling gradually to 82% at 5 years (see *Figure 2*). The equivalent figures in the randomised surgery group were 36% at 1 year (15% among those who had surgery) and 41% (26%) at 5 years. This was in response to a question that, to avoid problems with recall, asked just about the preceding 2 weeks (rather than the full year), and we have assumed that the 2 weeks are typical of the previous year. We know, however, that medication use is sometimes dynamic – that patients stop and start. This is apparent in *Table 13*, for example: among those in the medical group who were not taking medication at the end of the first year, 13 (68%) of the 19 respondents reported that they were taking PPIs at 5 years.

Short-term complications of surgery were described in more detail in the first report of this trial. However, the REFLUX trial is consistent with the other three trials in this respect, with small numbers having associated visceral injuries, postoperative problems and dilatation of the wrap. The longer-term follow-up has now clarified the likelihood of further surgery following a fundoplication. Overall, 4% ($n = 16$) of the total 364 in the study who had fundoplication had a subsequent reflux-related operation, of whom two had a further (i.e. third) operation. Reoperation was most often conversion to a different type of wrap or a reconstruction of the same wrap. There were only two cases of reversal of the fundoplication and neither was in the randomised comparison. In total, 3% ($n = 12$) of those who had fundoplication required surgical treatment for a complication directly related to the original surgery, including oesophageal dilatation ($n = 4$) and repair of incisional hernia ($n = 3$). As described in *Chapter 4*, although it is not possible to extract exactly comparable data, these results are broadly in line with those of the other trials.

Where the REFLUX trial results do differ from the results of the other trials, especially the LOTUS trial, is in the likelihood and extent of adverse symptoms associated with fundoplication. Dysphagia, flatulence and bloating, and inability to vomit despite wanting to have all been reported to be problematical after fundoplication. However, in the REFLUX trial, the patterns of difficulty swallowing, flatulence and wanting to vomit but being physically unable to do so were similar in the two randomised groups (see *Table 15*), with no statistically significant differences.

The economic analysis of the 5-year data from the REFLUX trial had two phases. First, a within-trial 5-year cost-effectiveness analysis was undertaken; this was followed by an exploration of the need to develop a longer-term model. Differences in mean costs and mean QALYs at 5 years were used to derive an estimate of relative cost-effectiveness. The base-case approach used multiple imputation (principally because of the extent of missing data), an ITT analysis and adjustment for baseline QALYs. As described in *Chapter 5*, complete-case and PP analyses were also undertaken, as were a range of structural, scenario and probabilistic sensitivity analyses. Costs were estimated from a health-care perspective and consideration was limited to randomised trial participants. Costs for each participant were calculated by multiplying their use of health-care resources by associated unit costs and were discounted at an annual rate of 3.5%. HRQoL was calculated from serial EQ-5D measurements. The mean (SD) costs in the first year were £2501 (£1698) in the surgical group compared with £560 (£1007) in the medical group; in each subsequent year the mean costs were around £175 higher in the medical group. The estimated incremental mean cost of the surgical policy was £1518 (95% CI £1006 to £2029) with incremental mean QALYs of 0.2160 (95% CI 0.0205 to 0.4115), giving an ICER of £7028. The probability of the surgical policy being the more cost-effective was 0.93 at a threshold of £20,000 per QALY and 0.96 at a threshold of £30,000 per QALY. The complete-case analysis gave similar results and the conclusions were robust to plausible changes in assumptions, the only exception being when surgery-allocated patients with missing data were assumed to experience lower HRQoL than other patients. A regression-based mixed-model approach was then used to explore the robustness of the findings and to gauge the likelihood that the current strong evidence for cost-effectiveness might be reversed over subsequent years. The regression-based model gave very similar results to the base-case imputation approach. Given the trends in both costs and benefits, it was concluded that it was highly unlikely that the cost-effectiveness of surgery would be reversed when extrapolated beyond 5 years.

Thus, this second phase of the REFLUX trial has accomplished what it set out to do. After 5 years' follow-up, a policy of relatively early laparoscopic fundoplication among patients for whom reasonable control of GORD symptoms requires long-term medication and for whom both surgery and medical management are suitable continues to provide better relief of GORD symptoms with associated better QoL. Although surgery carries risks, complications were rare. And despite being initially more costly, a surgical policy was found to be highly likely to be cost-effective for such patients at conventional threshold costs per QALY.

Implications for health care

Extending the use of laparoscopic fundoplication to people whose GORD symptoms require long-term medication for reasonable control and who would be suitable for surgery would provide health gain that extends over a number of years. The longer-term data reported here indicate that this is highly likely to be a cost-effective use of resources. The more troublesome the symptoms, the greater the potential benefit from surgery.

Recommendations for research

The practical implications for health services of any extension of the use of laparoscopic fundoplication depend on how many patients might seek such surgery as a consequence. Most patients taking anti-reflux medication are managed in general practice. Currently, it is uncertain how many people require long-term medication for reasonable control of their GORD symptoms, how many of these would be suitable for surgery and how many would seek it; hence, it is not clear what the most efficient provision of future care might be. We therefore recommend further research to address these issues and explore the practical and resource implications of alternative policies for laparoscopic fundoplication, which include extending its use within the NHS to the sorts of patients enrolled in the REFLUX trial.

Acknowledgements

The authors wish to thank Janice Cruden and Pauline Garden for their secretarial support and data management; Samantha Wileman and Julie Bruce for invaluable help with the overview of trials; the following researchers for their assistance in nurse co-ordination, patient recruitment and follow-up: Maureen GC Gillan, Marie Cameron, Christiane Pflanz-Sinclair, Lynne Swan; Sharon McCann, who was involved in the piloting of the practical arrangements of this trial; and Allan Walker, Daniel Barnett and Gladys McPherson for database and programming support.

Members of the (1) Trial Steering Group and (2) Data Monitoring Committee who oversaw the first phase of this study were:

1. Wendy Atkin* (Independent chairperson), John Bancewicz, Garry Barton (1999–2002), Ara Darzi, Robert Heading, Janusz Jankowski,* Zygmunt Krukowski, Richard Lilford,* Iain Martin (1997–2000), Ashley Mowat, Ian Russell, Mark Sculpher and Mark Thursz.
2. Jon Nicholl,* Chris Hawkey* and Iain MacIntyre.*

*Independent of trial.

Members of the REFLUX trial group responsible for recruitment in the clinical centres were as follows:

Aberdeen: Aberdeen Royal Infirmary	A Mowat, Z Krukowski, E El-Omar, P Phull and T Sinclair
Belfast: Royal Victoria Hospital	B Clements, J Collins, A Kennedy and H Lawther
Bournemouth: Royal Bournemouth Hospital	D Bennett, N Davies, S Toop and P Winwood
Bristol: Bristol Royal Infirmary	D Alderson, P Barham, K Green and R Mittal
Bromley: Princess Royal University Hospital	M Asante and S El Hasani
Edinburgh: Royal Infirmary of Edinburgh	A De Beaux, R Heading, L Meekison, S Paterson-Brown and H Barkell
Guildford: Royal Surrey County Hospital	G Ferns, M Bailey, N Karanjia, TA Rockall and L Skelly
Hull: Hull Royal Infirmary	M Dakkak, J King, C Royston and P Sedman
Inverness: Raigmore Hospital	K Gordon, LF Potts, C Smith, PL Zentler-Munro and A Munro
Leeds: Leeds General Infirmary	S Dexter and P Maoyeddi
Leicester: Leicester Royal Infirmary	DM Lloyd
London: St Mary's Hospital	V Loh, M Thursz and A Darzi
London: Whipps Cross Hospital	A Ahmed, R Greaves, A Sawyerr, J Wellwood and T Taylor
Poole: Poole Hospital	S Hosking, S Lowrey and J Snook
Portsmouth: Queen Alexandra Hospital	P Goggin, T Johns, A Quine, S Somers and S Toh
Salford: Hope Hospital	J Bancewicz, M Greenhalgh and W Rees
Stoke-on-Trent: North Staffordshire Hospital	CVN Cheruvu, M Deakin, S Evans, J Green and F Leslie
Swansea: Morriston Hospital	JN Baxter, P Duane, MM Rahman, M Thomas and J Williams
Telford: Princess Royal Hospital	D Maxton, A Sigurdsson, MSH Smith and G Townson
Yeovil: Yeovil District Hospital	S Gore, RH Kennedy, ZH Khan and J Knight
York: York District Hospital	D Alexander, G Miller, D Parker, A Turnbull and J Turvill

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health Directorate.

Contribution of authors

Adrian Grant (Professor, Health Services Research Trialist) was the principal grant applicant and contributed to the development of the trial protocol and the preparation of the report and had overall responsibility for the conduct of the study.

Charles Boachie (Statistician, Health Statistics) conducted the statistical analysis.

Seonaidh Cotton (Trial Co-ordinator, Health Services Research Trialist) was responsible for the day-to-day management of the trial, monitored data collection and assisted in the preparation of the report.

Rita Faria (Research Fellow, Health Economics) was involved in the cost-effectiveness section, namely conducting the economic evaluation and writing the final report.

Laura Bojke (Senior Research Fellow, Health Economics) was responsible for the cost-effectiveness section – supervising the economic evaluation and the systematic review of existing economic evidence and writing the final report.

David Epstein (Honorary Research Fellow, Health Economics) was involved in the cost-effectiveness section – conducting and supervising the economic evaluation and writing the final report.

Craig Ramsay (Professor, Health Services Research Statistician/Trialist) contributed to the grant application and the trial design and conducted the statistical analysis.

Belen Corbacho (Research Fellow, Health Economics) was responsible for the systematic review of existing health economic evidence – study selection, data extraction, validity assessment and writing the final report.

Mark Sculpher (Professor of Health Economics, Health Economics) was responsible for the economic evaluation section of the grant application and overseeing the economic evaluation.

Zygmunt Krukowski (Surgeon, Gastroenterology) advised on clinical aspects of the trial and commented on the draft report.

Robert C Heading (Honorary Professor) advised on clinical aspects of the trial design and the conduct of the trial and commented on the draft report.

Marion Campbell (Director, Health Services Research Statistician/Trialist) contributed to the development of the trial design and all aspects of the conduct of the trial and commented on the draft report.

Publications

1. Grant AM, Cotton SC, Boachie C, Ramsay CR, Krukowski ZH, Heading RC, *et al.* Minimal access surgery compared with medical management for gastro-oesophageal reflux disease: five year follow-up of a randomised controlled trial (REFLUX) *BMJ* 2013;**346**:f1908. DOI: http://dx.doi.org/10.1136/bmj.f1908 (published online on 18 April 2013).
2. Faria R, Bojke L, Epstein D, Corbacho B, Sculpher M, on behalf of the REFLUX trial group. Cost-effectiveness of laparoscopic fundoplication versus continued medical management for the treatment of gastro-oesophageal reflux disease based on long-term follow-up of the REFLUX trial. *Br J Surg* 2013; in press. DOI: http://dx.doi.org/10.1002/bjs.9190.

References

1. Grant A, Wileman S, Ramsay C, Bojke L, Epstein D, Sculpher M, *et al.* The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease – a UK collaborative study. The REFLUX trial. *Health Technol Assess* 2008;**12**(31).
2. Grant AM, Wileman SM, Ramsay CR, Mowat NA, Krukowski ZH, Heading RC, *et al.* Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ* 2008;**337**:a2664. http://dx.doi.org/10.1136/bmj.a2664
3. Epstein D, Bojke L, Sculpher MJ, REFLUX trial group. Laparoscopic fundoplication compared with medical management for gastro-oesophageal reflux disease: cost effectiveness study. *BMJ* 2009;**339**:b2576. http://dx.doi.org/10.1136/bmj.b2576
4. Macran S, Wileman S, Barton G, Russell I. The development of a new measure of quality of life in the management of gastro-oesophageal reflux disease: the Reflux questionnaire. *Qual Life Res* 2007;**16**:331–43. http://dx.doi.org/10.1007/s11136-006-9005-3
5. Bojke L, Hornby E, Sculpher M. A comparison of the cost effectiveness of pharmacotherapy or surgery (laparoscopic fundoplication) in the treatment of GORD. *Pharmacoeconomics* 2007;**25**:829–41. http://dx.doi.org/10.2165/00019053-200725100-00003
6. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R, Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;**101**:1900–20. http://dx.doi.org/10.1111/j.1572-0241.2006.00630.x
7. Locke GR, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. *Gastroenterology* 1997;**112**:1448–56. http://dx.doi.org/10.1016/S0016-5085(97)70025-8
8. Kennedy T, Jones R. The prevalence of gastro-oesophageal reflux symptoms in a UK population and the consultation behaviour of patients with these symptoms. *Aliment Pharmacol Ther* 2000;**14**:1589–94. http://dx.doi.org/10.1046/j.1365-2036.2000.00884.x
9. Heading RC. Prevalence of upper gastrointestinal symptoms in the general population: a systematic review. *Scand J Gastroenterol* 1999;**34**:3–8.
10. Davis CS, Baldea A, Johns JR, Joehl RJ, Fisichella PM. The evolution and long-term results of laparoscopic antireflux surgery for the treatments of gastroesophageal reflux disease. *JSLS* 2010;**13**:332–41. http://dx.doi.org/10.4293/108680810X12924466007007
11. Peters MJ, Mukhtar A, Yunus RM, Khan S, Pappalardo J, Memon B, *et al.* Meta-analysis of randomized clinical trials comparing open and laparoscopic anti-reflux surgery. *Am J Gastroenterol* 2009;**104**:1548–61. http://dx.doi.org/10.1038/ajg.2009.176
12. Donahue PE, Larson GM, Stewardson RH, Bombeck CT. Floppy Nissen fundoplication. *Rev Surg* 1977;**34**:223–4.
13. DeMeester TR, Bonavina L, Albertucci M. Nissen fundoplication for gastroesophageal reflux disease. Evaluation of primary repair in 100 consecutive patients. *Ann Surg* 1986;**204**:9–19. http://dx.doi.org/10.1097/00000658-198607000-00002
14. Broeders JAJL, Mauritz FA, Ahmed Ali U, Draaisma WA, Ruurda JP, Gooszen HG, *et al.* Systematic review and meta-analysis of laparoscopic Nissen (posterior total) versus Toupet (posterior partial) fundoplication for gastro-oesophageal reflux disease. *Br J Surg* 2010;**97**:1318–30. http://dx.doi.org/10.1002/bjs.7174

15. Lundell L, Miettinen P, Myrvold HE, Hatlebakk JG, Wallin L, Engstrom C, *et al.* Comparison of outcomes twelve years after antireflux surgery or omeprazole maintenance therapy for reflux esophagitis. *Clin Gastroenterol Hepatol* 2009;**7**:1292–8. http://dx.doi.org/10.1016/j.cgh.2009.05.021

16. Chiba N, De Gara CJ, Wilkinson JM, Hunt RH. Speed of healing and symptom relief in grade II to IV gastroesophageal reflux disease: a meta analysis. *Gastroenterology* 1997;**112**:1798–810. http://dx.doi.org/10.1053/gast.1997.v112.pm9178669

17. Boutet R, Wilcock M, Mackenzie I. Survey on repeat prescribing for acid suppression drugs in primary care in Cornwall and the Isles of Scilly. *Aliment Pharmacol Ther* 1999;**13**:813–17. http://dx.doi.org/10.1046/j.1365-2036.1999.00524.x

18. McCullagh M, Brown C, Bell D, Powell K. Long term acid suppressing treatment survey shows variation in practice. *BMJ* 1994;**308**:1238. http://dx.doi.org/10.1136/bmj.308.6938.1238

19. Ryder SD, O'Reilly S, Miller RJ, Ross J, Jacyna MR, Levi AJ. Long term acid suppressing treatment in general practice. *BMJ* 1994;**308**:827–30. http://dx.doi.org/10.1136/bmj.308.6932.827

20. Dibley LB, Norton C, Jones R. Non-pharmacological intervention for gastro-oesophageal reflux disease in primary care. *Br J Gen Pract* 2010;**60**:e459–65. http://dx.doi.org/10.3399/bjgp10X544050

21. Roberts SJ, Bateman DN. Prescribing of antacids and ulcer-healing drugs in primary care in the north of England. *Aliment Pharmacol Ther* 1995;**9**:137–43. http://dx.doi.org/10.1111/j.1365-2036.1995.tb00362.x

22. Bashford JNR, Norwood J, Chapman SR. Why are patients prescribed proton pump inhibitors? Retrospective analysis of the link between morbidity and prescribing in the General Practice Research Database. *BMJ* 1998;**317**:452–6. http://dx.doi.org/10.1136/bmj.317.7156.452

23. Yang Y, Metz DC. Reviews in basic and clinical gastroenterology and hepatology safety of proton pump inhibitor exposure. *Gastroenterology* 2010;**139**:1115–27. http://dx.doi.org/10.1053/j.gastro.2010.08.023

24. Sheen E, Triadafilopoulos G. Adverse effects of long-term proton pump inhibitor therapy. *Dig Dis Sci* 2011;**56**:931–50. http://dx.doi.org/10.1007/s10620-010-1560-3

25. Ogawa R, Echizen H. Drug–drug interaction profiles of proton pump inhibitors. *Clin Pharmacokinet* 2010;**49**:509–33. http://dx.doi.org/10.2165/11531320-000000000-00000

26. Henshaw RC, Naji SA, Russell IT, Templeton AA. Comparison of medical abortion with surgical vacuum aspiration: women's preferences and acceptability of treatment. *BMJ* 1993;**307**:714–17. http://dx.doi.org/10.1136/bmj.307.6906.714

27. Cooper KG, Grant AM, Garratt AM. The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. *Br J Obstet Gynaecol* 1997;**104**:1367–73. http://dx.doi.org/10.1111/j.1471-0528.1997.tb11005.x

28. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ* 1989;**299**:313–15. http://dx.doi.org/10.1136/bmj.299.6694.313

29. British Medical Association and Royal Pharmaceutical Society of Great Britain. *British national formulary*. No. 52, September 2006. London: BMA and RPS; 2006.

30. Torgerson D, Sibbald B. Understanding controlled trials: what is a patient preference trial? *BMJ* 1998;**316**:360. http://dx.doi.org/10.1136/bmj.316.7128.360

31. Dent J, Brun J, Fendrick AM, Fennerty MB, Janssens J, Kahrilas PJ, *et al.* An evidence-based appraisal of reflux disease management – the Genval Workshop Report. *Gut* 1999;**44**:S1–16.

32. Edwards PJ, Roberts IG, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, *et al.* Methods to increase response rates to postal questionnaires. *Cochrane Database Syst Rev* 2007;**2**:MR000008.

33. Brealey S, Atwell C, Bryan S. Improving response rates using monetary incentives for patient completion of questionnaires: an observational study. *BMC Med Res Methodol* 2007;**7**:12. http://dx.doi.org/10.1186/1471-2288-7-12

34. Kenyon S, Pike K, Jones D, Taylor D, Salt A, Marlow N, *et al.* The effect of a monetary incentive on return of a postal health and development questionnaire: a randomised trial. *BMC Health Serv Res* 2007;**5**:55. http://dx.doi.org/10.1186/1472-6963-5-55

35. Dirmaier J, Harfst T, Koch U, Schulz H. Incentives increased return rates but did not influence partial nonresponse or treatment outcome in a randomized trial. *J Clin Epidemiol* 2007;**60**:1263–70. http://dx.doi.org/10.1016/j.jclinepi.2007.04.006

36. Brooks R, with the EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. http://dx.doi.org/10.1016/0168-8510(90)90421-9

37. Jenkinson C, Layte R, Wright L, Coulter A. *The UK SF-36: an analysis and interpretation manual*. Oxford: Health Services Research Unit; 1996.

38. Nagelkerke N, Fidler V, Bersen R, Borgdorff M. Estimating treatment effects in randomised clinical trials in the presence of non-compliance. *Stat Med* 2000;**19**:1849–64. http://dx.doi.org/10.1002/1097-0258(20000730)19:14<1849::AID-SIM506>3.0.CO;2-1

39. White IR. Uses and limitations of randomization-based efficacy estimators. *Stat Methods Med Res* 2005;**14**:327–47. http://dx.doi.org/10.1191/0962280205sm406oa

40. Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al.* Contamination in trials of educational interventions. *Health Technol Assess* 2007;**11**(43).

41. Fielding S, Fayers P, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health Qual Life Outcomes* 2009;**7**:57. http://dx.doi.org/10.1186/1477-7525-7-57

42. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998;**316**:285. http://dx.doi.org/10.1136/bmj.316.7127.285

43. Wileman SM, McCann S, Grant AM, Krukowski ZH, Bruce J. Medical versus surgical management for gastro-oesophageal reflux disease (GORD) in adults. *Cochrane Database Syst Rev* 2010;**3**:CD003243.

44. Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, *et al.* A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for treatment of patients with chronic gastroesophageal reflux disease: one-year follow-up. *Surg Innov* 2006;**13**:238–49. http://dx.doi.org/10.1177/1553350606296389

45. Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, *et al.* A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for the treatment of patients with chronic gastroesophageal reflux disease (GERD): 3-year outcomes. *Surg Endosc* 2011;**25**:2547–54. http://dx.doi.org/10.1007/s00464-011-1585-5

46. Goeree R, Hopkins R, Marshall JK, Armstrong D, Ungar WJ, Goldsmith C, *et al.* Cost-utility of laparoscopic Nissen fundoplication versus proton pump inhibitors for chronic and controlled gastroesophageal reflux disease: a 3-year prospective randomized controlled trial and economic evaluation. *Value Health* 2011;**14**:263–73. http://dx.doi.org/10.1016/j.jval.2010.09.004

47. Lundell L, Attwood S, Ell C, Fiocca R, Galmiche J, Hatlebakk J, *et al.* Comparing laparoscopic antireflux surgery with esomeprazole in the management of patients with chronic gastro-oesophageal reflux disease: a 3-year interim analysis of the LOTUS trial. *Gut* 2008;**57**:1207–13. http://dx.doi.org/10.1136/gut.2008.148833

48. Galmiche J, Hatlebakk J, Attwood S, Ell C, Fiocca R, Eklund S, *et al.* Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD. The LOTUS randomized clinical trial. *JAMA* 2011;**305**:1969–77. http://dx.doi.org/10.1001/jama.2011.626

49. Attwood SE, Lundell L, Hatlebakk JG, Eklund S, Junghard O, Galmiche J, *et al.* Medical or surgical management of GERD patients with Barrett's esophagus: the LOTUS trial 3-year experience. *J Gastrointest Surg* 2008;**12**:1646–55. http://dx.doi.org/10.1007/s11605-008-0645-1

50. Attwood SEA, Lundell L, Ell C, Galmiche J, Hatlebakk J, Fiocca R, *et al.* Standardization of surgical technique in antireflux surgery: the LOTUS trial experience. *World J Surg* 2008;**32**:995–8. http://dx.doi.org/10.1007/s00268-007-9409-4

51. Mahon D, Rhodes M, Decadt B, Hindmarsh A, Lowndes R, Beckingham I, *et al.* Randomized clinical trial of laparoscopic Nissen fundoplication compared with proton-pump inhibitors for treatment of chronic gastro-oesophageal reflux. *Br J Surg* 2005;**92**:695–9. http://dx.doi.org/10.1002/bjs.4934

52. Cookson R, Flood C, Koo B, Mahon D, Rhodes M. Short-term cost effectiveness and long-term cost analysis comparing laparoscopic Nissen fundoplication with proton-pump inhibitor maintenance for gastro-oesophageal reflux disease. *Br J Surg* 2005;**92**:700–6. http://dx.doi.org/10.1002/bjs.4933

53. Mehta S, Bennett J, Mahon D, Rhodes M. Prospective trial of laparoscopic Nissen fundoplication versus proton pump inhibitor therapy for gastroesophageal reflux disease: seven-year follow-up. *J Gastrointest Surg* 2006;**10**:1312–17. http://dx.doi.org/10.1016/j.gassur.2006.07.010

54. Wiklund I, Bigard MA, Grace E, Talley NJ, Kamm M, Veldhuyzen van Zanten S, *et al.* Quality of life in reflux and dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD) 11. *Eur J Surg* 1998;**164**(Suppl. 583):41–9. http://dx.doi.org/10.1080/11024159850191238

55. Dimenas E. Methodological aspects of evaluation of quality of life in upper gastrointestinal diseases. *Scand J Gastroenterol* 1993;**28**(Suppl. 199):18–21. http://dx.doi.org/10.3109/00365529309098350

56. Rantanen TK, Salo JA, Sipponen JT. Fatal and life-threatening complications in antireflux surgery: analysis of 5502 operations. *Br J Surg* 1999;**86**:1573–7. http://dx.doi.org/10.1046/j.1365-2168.1999.01297.x

57. van Pinxteren B, Sigterman KE, Bonis P, Lau J, Numans ME. 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* 2010;**11**:CD002095.

58. Romagnuolo J, Meier MA, Sadowski DC. Medical or surgical therapy for erosive reflux oesophagitis: cost–utility analysis using a Markov model. *Ann Surg* 2002;**236**:191–202. http://dx.doi.org/10.1097/00000658-200208000-00007

59. Arguedas MR, Heudebert GR, Klapow JC, Centor RM, Eloubeidi M, Wilcox CM. Re-examination of the cost-effectiveness of surgical versus medical therapy in patients with gastroesophageal reflux disease: the value of long-term data collection. *Am J Gastroenterol* 2004;**99**:1023–8. http://dx.doi.org/10.1111/j.1572-0241.2004.30891.x

60. Comay D, Adam V, da Silveira EB, Kennedy W, Mayrand S, Barkun AN. The Stretta procedure versus proton pump inhibitors and laparoscopic Nissen fundoplication in the management of gastroesophageal reflux disease: a cost-effectiveness analysis. *Can J Gastroenterol* 2008;**22**:552–8.

61. *HMSO Electronic Drug Tariff for the National Health Service of England and Wales*. London: HMSO; 2010.

62. Manca A, Hawkins N. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. http://dx.doi.org/10.1002/hec.944

63. Department of Health (DoH). *NHS reference costs 2009–2010*. London: DoH; 2010.

64. Curtis L. *Unit costs of health and social care*. Canterbury: PSSRU, University of Kent; 2010.

65. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal*. London: NICE; 2008.

66. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. Discussion paper 138. York: Centre for Health Economics, University of York; 1995.

67. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technol Assess* 1999;**3**(10).

68. HM Treasury. *Green book: appraisal and evaluation in central government*. London: The Stationery Office; 2003.

69. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2010;**30**:377–99. http://dx.doi.org/10.1002/sim.4067

70. White IR, Mander A, Wason J, Bond S. Handling missing outcome data in randomised trials: a workshop for the Network of Hubs for Trials Methodology Research. Medical Research Council, Cambridge, 20 June 2011.

71. Royston P. Multiple imputation by the MICe system of chained equations. Implementation in Stata. *J Stat Softw* 2011;**45**(4). URL: www.jstatsoft.org (accessed 5 December 2012).

72. Little RJA, Rubin DB. *Statistical analysis with missing data*. New York, NY: Wiley; 1987.

73. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley; 1987. http://dx.doi.org/10.1002/9780470316696

74. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* 2007;**61**:79–90. http://dx.doi.org/10.1198/000313007X172556

75. Galati JC, Carlin JB, Royston P. MIM: Stata module to analyse and manipulate multiply imputed datasets. EconPapers. URL: http://EconPapers.repec.org/RePEc:boc:bocode:s456825 (accessed 1 July 2011).

76. Briggs AH, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999;**3**(2).

77. Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc Series B* 1964;**2**:211–43.

78. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin Trials* 2007;**4**:154–61. http://dx.doi.org/10.1177/1740774507076914

79. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011;**342**:910–12. http://dx.doi.org/10.1136/bmj.d40

80. Johannesson M, Weinstein S. On the decision rules of cost-effectiveness analysis. *J Health Econ* 1993;**12**:459–67. http://dx.doi.org/10.1016/0167-6296(93)90005-Y

81. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;**18**(Suppl.):S68–80. http://dx.doi.org/10.1177/0272989X9801800209

82. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;**26**:781–98. http://dx.doi.org/10.2165/00019053-200826090-00008

Appendix 1 Annual questionnaire

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

The REFLUX Trial

ANNUAL FOLLOW-UP QUESTIONNAIRE

A questionnaire for people participating in the REFLUX trial,
which aims to find out whether taking medication or having an operation
is the best form of treatment for gastro-oesophageal reflux disease

CONFIDENTIAL

This study is funded by the NHS Research and Development Health Technology Assessment Programme

PLEASE READ ALL THE INSTRUCTIONS BEFORE COMPLETING THE QUESTIONNAIRE

Thank you for agreeing to take part in the study. The responses you give in this questionnaire will help us find out if the treatments you get are helpful for your condition.

The information you provide will be completely confidential.

HOW TO FILL IN THE QUESTIONNAIRE

For each section please put a cross in the appropriate box like this:

Do you drive a car? **Yes** ☐

No ☒

If you make any errors while completing this questionnaire, shade out the incorrect box completely and put a cross in the correct box like this:

Do you drive a car? **Yes** ☒

No ☒

The intended answer above is No.

PLEASE USE A BLUE OR BLACK PEN TO FILL IN YOUR ANSWERS

REFLUX QUESTIONNAIRE

For the questions in section A - F, please put a cross in the box which best describes how often your symptoms have occurred and the effect they have had on your quality of life.

SECTION A - HEARTBURN

A1. In the last two weeks, how often have you experienced heartburn (a burning sensation which moves up from your chest to your throat)?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

A2. In the last two weeks, how often have you experienced any discomfort or pain in your chest?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

A3. In the last two weeks, how much has the heartburn or discomfort/pain in your chest affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

SECTION B - ACID REFLUX

B1. In the last two weeks, how often have you experienced acid reflux and/or had an acid taste in your mouth?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B2. In the last two weeks, how often have you been sick (vomited)?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B3. In the last two weeks, how often have you regurgitated (brought up) quantities of liquid or solids into your mouth?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B4. In the last two weeks, how often have you experienced a feeling of nausea (without actually being sick or regurgitating)?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B5. In the last two weeks, how often have you wanted to be sick but physically been unable to?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B6. In the last two weeks, how much have these acid reflux symptoms affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION C – WIND

C1. In the last two weeks, how often have you experienced a lot of wind from the lower bowel?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

C2. In the last two weeks, how often have you experienced a lot of burping/belching?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

Participant Study No

☐☐☐☐☐

(for completion by co-ordinating centre in Aberdeen)

C3. In the last two weeks, how often have you experienced bloatedness and/or a feeling of trapped wind, in your stomach?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

C4. In the last two weeks, how often have you experienced loud gurgling noises from your stomach?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

C5. In the last two weeks, how much have these wind problems affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION D - EATING AND SWALLOWING

D1. In the last two weeks, how often have you experienced difficulty swallowing food or have you actually choked on food?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

Participant Study No

☐☐☐☐☐

(for completion by co-ordinating centre in Aberdeen)

D2. In the last two weeks, how often have your eating habits been restricted because of your condition? Examples might be eating more slowly, having smaller portions or eating different foods.

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

D3. In the last two weeks, how much have these problems with eating affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION E – BOWEL MOVEMENTS

E1. In the last two weeks, how often have you experienced diarrhoea and/or loose stools?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

E2. In the last two weeks, how often have you experienced constipation and/or hard stools?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

Participant Study No

☐☐☐☐☐

(for completion by co-ordinating centre in Aberdeen)

E3. In the last two weeks, how often have you had a feeling of an urgent need to have a bowel movement?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

E4. In the last two weeks, how often have you had a feeling of not emptying your bowels?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

E5. In the last two weeks, how much have these bowel problems affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION F – SLEEP

F1. In the last two weeks, how often have you experienced difficulty in lying down to sleep?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most nights ☐
- Every night ☐

F2. In the last two weeks, how often have you experienced difficulty getting to sleep because of your reflux symptoms?

- ☐ Not at all
- ☐ Once a week
- ☐ Two or three times a week
- ☐ Most nights
- ☐ Every night

F3. In the last two weeks, how often have you been woken up because of your reflux symptoms?

- ☐ Not at all
- ☐ Once a week
- ☐ Two or three times a week
- ☐ Most nights
- ☐ Every night

F4. In the last two weeks, how much have these sleep related problems affected your quality of life?

- ☐ Not at all
- ☐ A little
- ☐ Moderately
- ☐ A lot
- ☐ Extremely

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

SECTION G – WORK, PHYSICAL AND SOCIAL ACTIVITIES

For the following section, please put a cross in the box which best applies to you.

G1. In the last two weeks, have your reflux symptoms affected you at work (paid or voluntary)?

- Not applicable (I do not do paid or voluntary work) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still work ☐
- Yes, I have worked less often because of my symptoms ☐
- Yes, I have not worked in the last two weeks because of my symptoms ☐
- I no longer work because of my symptoms ☐

G2. In the last two weeks, have your reflux symptoms affected your ability to perform less strenuous activities (such as going for a gentle walk, shopping or housework)?

- Not applicable (I do not perform these activities, though this is not due to my reflux symptoms) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still perform these activities as often as ever ☐
- Yes, I perform these activities less often because of my symptoms ☐
- Yes, I have not performed these activities in the last two weeks ☐
- I no longer perform these activities at all because of my symptoms ☐

G3. In the last two weeks, have your reflux symptoms affected your ability to perform strenuous activities (such as brisk walking or swimming)?

- Not applicable (I do not perform these activities, though this is not due to my reflux symptoms) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still perform these activities as often as ever ☐
- Yes, I perform these activities less often because of my symptoms ☐
- Yes, I have not performed these activities in the last two weeks ☐
- I no longer perform these activities at all because of my symptoms ☐

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

G4. In the last two weeks, have you found that your reflux symptoms have affected any of your social activities (such as going out for meals, going out for drinks or socializing with other people)?

- Not applicable (I do not perform these activities, though this is not due to my reflux symptoms) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still perform these activities as often as ever ☐
- Yes, I perform these activities less often because of my symptoms ☐
- Yes, I have not performed these activities in the last two weeks ☐
- I no longer perform these activities at all because of my symptoms ☐

G5. In the last two weeks, how much has the effect of your reflux symptoms on your work, physical or social activities affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION H – DESCRIBING YOUR OWN HEALTH TODAY

By placing a cross in one box in each group below, please indicate which statements best describe your own health state today

Mobility	I have no problems in walking about	☐
	I have some problems in walking about	☐
	I am confined to bed	☐
Self-care	I have no problems with self-care	☐
	I have some problems washing or dressing myself	☐
	I am unable to wash or dress myself	☐
Usual Activities *(e.g. work, study, housework, family or leisure activities)*	I have no problems with performing my usual activities	☐
	I have some problems with performing my usual activities	☐
	I am unable to perform my usual activities	☐
Pain/Discomfort	I have no pain or discomfort	☐
	I have moderate pain or discomfort	☐
	I have extreme pain or discomfort	☐
Anxiety/Depression	I am not anxious or depressed	☐
	I am moderately anxious or depressed	☐
	I am extremely anxious or depressed	☐

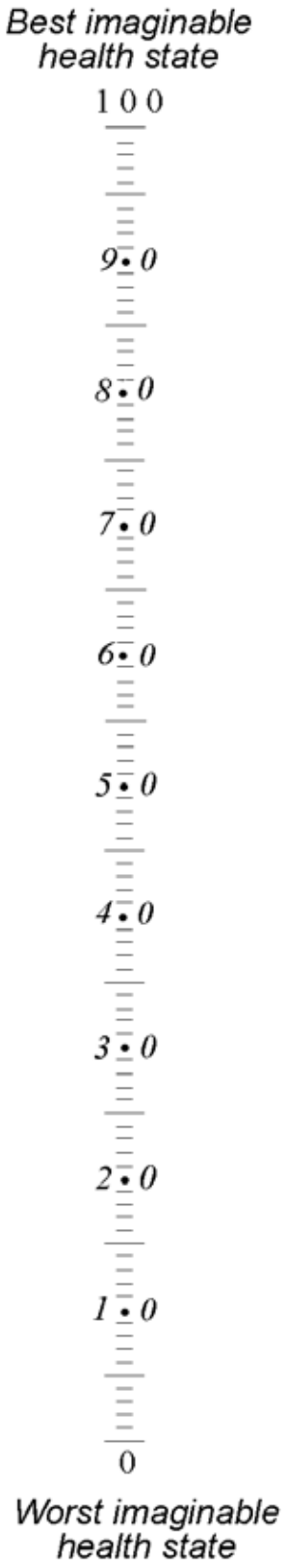
SECTION H - DESCRIBING YOUR OWN HEALTH TODAY

Please indicate on this scale how good or bad your own health state is today.

The best health state you can imagine is marked 100 and the worst health state you can imagine is marked 0.

Please draw a line from the box below to the point on the scale that best indicates how good or bad your health state is today.

Your own health state today



SECTION I – GENERAL HEALTH

Please fill in all the questions again by putting a cross in the relevant box of the answer that applies to you.

These questions ask for your views about your health and how you feel about life in general. Do not spend too much time in answering as your immediate response is likely to be the most accurate.

1. In general, would you say your health is:

Excellent	Very good	Good	Fair	Poor
☐	☐	☐	☐	☐

2. Compared to one year ago, how would you rate your health in general now?

Much better now than one year ago	Somewhat better now than one year ago	About the same as one year ago	Somewhat worse now than one year ago	Much worse now than one year ago
☐	☐	☐	☐	☐

3. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes limited a lot	Yes limited a little	No, not limited at all
a) **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sport	☐	☐	☐
b) **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling or playing golf	☐	☐	☐
c) Lifting or carrying groceries	☐	☐	☐
d) Climbing **several** flights of stairs	☐	☐	☐
e) Climbing **one** flight of stairs	☐	☐	☐
f) Bending, kneeling or stooping	☐	☐	☐
g) Walking **more than one mile**	☐	☐	☐
h) Walking **several hundred yards**	☐	☐	☐
i) Walking **one hundred yards**	☐	☐	☐
j) Bathing or dressing yourself	☐	☐	☐

4. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a) Cut down on the **amount of time** you spent on work or other activities	☐	☐	☐	☐	☐
b) **Accomplished less** than you would like	☐	☐	☐	☐	☐
c) Were limited in the **kind** of work or other activities	☐	☐	☐	☐	☐
d) Had **difficulty** performing the work or other activities (for example, it took extra effort)	☐	☐	☐	☐	☐

5. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a) Cut down on the **amount of time** you spent on work or other activities	☐	☐	☐	☐	☐
b) **Accomplished less** than you would like	☐	☐	☐	☐	☐
c) Did work or other **activities less carefully than usual**	☐	☐	☐	☐	☐

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with the family, friends, neighbours, or groups?

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

7. How much bodily pain have you had during the past 4 weeks?

None	Very mild	Mild	Moderate	Severe	Very severe
☐	☐	☐	☐	☐	☐

8. During the past 4 weeks, how much did pain interfere with your normal work (including both outside the home and housework)?

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks…

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a) Did you feel full of life?	☐	☐	☐	☐	☐
b) Have you been very nervous?	☐	☐	☐	☐	☐
c) Have you felt so down in the dumps that nothing could cheer you up?	☐	☐	☐	☐	☐
d) Have you felt calm and peaceful?	☐	☐	☐	☐	☐
e) Did you have a lot of energy?	☐	☐	☐	☐	☐
f) Have you felt downhearted and depressed?	☐	☐	☐	☐	☐
g) Did you feel worn out?	☐	☐	☐	☐	☐
h) Have you been happy?	☐	☐	☐	☐	☐
i) Did you feel tired?	☐	☐	☐	☐	☐

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
☐	☐	☐	☐	☐

11. How TRUE or FALSE is each of the following statements for you?

		Definitely true	Mostly true	Don't know	Mostly false	Definitely false
a)	I seem to get sick a little easier than other people	☐	☐	☐	☐	☐
b)	I am as healthy as anybody I know	☐	☐	☐	☐	☐
c)	I expect my health to get worse	☐	☐	☐	☐	☐
d)	My health is excellent	☐	☐	☐	☐	☐

SECTION J - HEALTH CARE RELATED QUESTIONS

In the following questions, we are trying to find out about some of the costs you incurred over the last 12 MONTHS as a result of your health problems.

If you are not sure or cannot remember exact details, please give the best answer you can.

1. CURRENT EMPLOYMENT

Please tick the box, which best describes your current employment status.

- ☐ Full time employment
- ☐ Part time employment
- ☐ Student
- ☐ Retired
- ☐ Housework
- ☐ Seeking work
- ☐ Other

2. TIME AWAY FROM WORK, DUE TO ILLNESS

If you are in paid employment, how many days off work have you had in the past 12 MONTHS because of health problems?

☐☐☐ Days in total

☐☐☐ Days because of reflux symptoms

3. VISITS TO NHS HEALTH CARE FACILITIES

a) How many times in the past 12 MONTHS have you personally visited your GP? Do not include visits made on behalf of others, or if you are a woman attending routine visits because of your pregnancy.

☐ Total number of visits

☐ Visits because of your reflux symptoms

b) How many times in the past 12 MONTHS have you personally had a visit from your GP?

☐ Total number of visits

☐ Visits because of your reflux symptoms

Please give details of the visits that you have had TO or FROM your GP in the spaces below
(continue on a separate sheet if necessary).

Visit 1

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

Visit 2

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

Visit 3

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

Visit 4

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

c) How many times in the past 12 MONTHS have you personally had to attend the outpatients or casualty department of a hospital?

☐ Total number of visits

☐ Visits because of your **reflux** symptoms

d) How many times in the past 12 MONTHS have you personally been admitted to a hospital as a day case (do not stay overnight)?

☐ Total number of day case admissions

☐ Admissions because of your **reflux** symptoms

Please give details of the day case admissions you have had and approximate date, in the spaces below *(continue on a separate sheet if necessary)*.

Admission 1

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

Admission 2

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

Admission 3

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

Admission 4

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

e) How many times in the past 12 MONTHS have you personally been admitted to a hospital for treatment as an inpatient (overnight or longer)?

☐ Total number of **inpatient** admissions

☐ Admissions because of your **reflux** symptoms

Please give details of the inpatient stays you have had, in the spaces below. (continue on a separate sheet if necessary)

Admission 1

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

Admission 2

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

Admission 3

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

Admission 4

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

4. PRESCRIBED MEDICATION FOR REFLUX

Are you currently being PRESCRIBED medication for reflux symptoms?

YES ☐ **NO** ☐ → *If NO, please go to question 5 on page 29*

If YES, please put a cross in the box against the current dose you are being prescribed and write in the number of tablets you have taken in the last two weeks.

(Please note the dose can be found on the side of your tablet bottle or packet)

	Dose (mg)						Number of tablets taken in the last 2 weeks
Omeprazole (Losec)	10mg ☐		20mg ☐		40mg ☐		☐
Lansoprazole (Zoton)	15mg ☐		30mg ☐				☐
Pantoprazole (Protium)	20mg ☐		40mg ☐				☐
Rabeprazole (Pariet)	10mg ☐		20mg ☐				☐
Esomeprazole (Nexium)	20mg ☐		40mg ☐				☐
Ranitidine (Zantac)	150mg ☐		300mg ☐				☐
Famotidine (Pepcid)	20mg ☐		40mg ☐				☐
Nizatidine (Axid)	150mg ☐		300mg ☐				☐
Cimetidine (Tagamet)	400mg ☐		800mg ☐				☐
Domperidone (Motilium)	10mg ☐		20mg ☐				☐
Metoclopramide (Maxolon)	10mg ☐		20mg ☐				☐

If you are prescribed any other medication (tablets or liquid) for your reflux symptoms that are not listed above, please list below the name(s) of the medicine(s) and include the number of times you have taken it in the last two weeks.

Names of medication	Number of times taken in last 2 weeks
e.g. Gaviscon	

5. NON PRESCRIBED MEDICATION FOR REFLUX

Please list below the names of any NON PRESCRIBED (over the counter) medication (tablets/liquid) you take for your REFLUX symptoms and include the number of times you have taken it in the last two weeks.

Names of medication	Number of times taken in last 2 weeks
e.g. Rennies	

IF YOU HAVE ANY OTHER COMMENTS about your gastro-oesophageal reflux symptoms, your reflux treatment or this study, please write them below.

THANK YOU FOR YOUR HELP IN COMPLETING THIS QUESTIONNAIRE

Once you have completed the form, please return it in the pre-paid envelope provided or to the following address:

REFLUX Trial Office
Health Services Research Unit
Polwarth Building
Foresterhill
Aberdeen AB25 2ZD
Tel: 01224 XXXXXX
Fax: 01224 XXXXXX
E-mail: reflux@hsru.abdn.ac.uk

Appendix 2 Intra- and postoperative surgical outcomes

TABLE 40 Intra- and post-operative surgical outcomes

Surgical outcome	Surgical participants, *n* (%)	
	Randomised (*n* = 111)	Preference (*n* = 218)
Conversion	2 (1.8)	0 (0.0)
Liver injury	1 (0.9)	1 (0.5)
Splenic injury	0 (0.0)	1 (0.5)
Pleural injury	1 (0.9)	2 (0.9)
Oesophageal injury	0 (0.0)	0 (0.0)
Other visceral injury	0 (0.0)	0 (0.0)
Haemorrhage	1 (0.9)	1 (0.5)
Pneumothorax	0 (0.0)	2 (0.9)
Blood transfusion	0 (0.0)	1 (0.5)[a]
Other postoperative event	3 (2.7)	5 (2.3)
ICU admission	0 (0.0)	0 (0.0)
HDU admission	1 (0.9)	2 (0.9)
Reoperation within 12 months	0 (0.0)	3 (1.4)
Stricture dilatation or food disimpaction required within 12 months	1 (0.9)	2 (0.9)
Discharged status		
Home	107 (96.4)	213 (97.7)
Other	4 (3.6)	5 (2.3)
Length of stay (days), median (IQR)	2 (2–3)	2 (2–3)

HDU, high-dependency unit; ICU, intensive-care unit; IQR, interquartile range.

a Participant was transfused with three units.

Appendix 3 Tables showing medication use in preceding fortnight at each time point of follow-up

TABLE 41 Follow-up at the time point equivalent to 3 months after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	150	109	158	150	230	203	182	178
PPIs, *n* (%)[a]								
Any PPI	47 (31.3)	8 (7.3)	140 (88.6)	133 (88.7)	41 (17.8)	13 (6.4)	167 (91.8)	152 (84.9)
Omeprazole	16 (10.7)	5 (4.6)	45 (28.5)	45 (30.0)	15 (6.5)	3 (1.5)	57 (31.3)	57 (31.8)
Lansoprazole	19 (12.7)	3 (2.8)	55 (34.8)	54 (36.0)	13 (5.7)	7 (3.5)	67 (36.8)	64 (35.8)
Pantoprazole	1 (0.7)	0 (0.0)	9 (5.7)	8 (5.3)	3 (1.3)	2 (1.0)	14 (7.7)	14 (7.8)
Rabeprazole	4 (2.7)	1 (0.9)	9 (5.7)	9 (6.0)	3 (1.3)	0 (0.0)	13 (7.1)	13 (7.3)
Esomeprazole	7 (4.7)	1 (0.9)	22 (13.9)	21 (14.0)	7 (3.0)	3 (1.5)	21 (11.5)	21 (11.7)
H_2RAs, *n* (%)[a]								
Any H_2RA	1 (0.7)	0 (0.0)	12 (7.6)	10 (6.7)	4 (1.7)	2 (1.0)	14 (7.7)	13 (7.3)
Ranitidine	0 (0.0)	0 (0.0)	8 (5.1)	8 (5.3)	2 (0.9)	1 (0.5)	10 (5.5)	9 (5.0)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.6)
Cimetidine	1 (0.7)	0 (0.0)	1 (0.6)	1 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Nizatidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.6)
Over-the-counter H_2RA	0 (0.0)	0 (0.0)	4 (2.5)	4 (2.7)	2 (0.9)	1 (0.5)	3 (1.7)	3 (1.7)
Prokinetics, *n* (%)[a]								
Any prokinetic	7 (4.7)	3 (2.8)	6 (3.8)	5 (3.3)	7 (3.0)	6 (3.0)	5 (2.7)	4 (2.2)
Domperidone	3 (2.0)	1 (0.9)	6 (3.8)	5 (3.3)	3 (1.3)	2 (1.0)	4 (2.2)	3 (1.7)
Metoclopramide	4 (2.7)	2 (1.8)	0 (0.0)	0 (0.0)	4 (1.7)	4 (2.0)	1 (0.5)	1 (0.6)
Any reflux-related drug, *n* (%)[a]	50 (33.3)	10 (9.2)	146 (92.4)	139 (92.7)	45 (19.6)	17 (8.4)	176 (96.7)	161 (89.9)
Other prescribed drugs, *n*[b]								
Alginates	0	0	4	4	0	0	2	2
Antispasmodics (e.g. dicycloverine)	0	0	1	1	2	2	0	0
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0

continued

TABLE 41 Follow-up at the time point equivalent to 3 months after surgery: medications (*continued*)

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Other ulcer-healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	0	0	0	0	1	1
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	1	1	0	0	2	1	1	1
Anti-motility	0	0	0	0	1	1	0	0

a Percentage is for responders completing the relevant section of the questionnaire.
b More than one prescription per person possible.

TABLE 42 Follow-up at the time point equivalent to 12 months after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	154	104	164	155	230	202	177	174
PPIs, *n* (%)[a]								
Any PPI	56 (36.4)	13 (12.5)	142 (86.6)	139 (89.7)	42 (18.3)	19 (9.4)	156 (88.1)	154 (88.5)
Omeprazole	19 (12.3)	6 (5.8)	47 (28.7)	45 (29.0)	14 (6.1)	4 (2.0)	61 (34.5)	60 (34.5)
Lansoprazole	21 (13.6)	2 (1.9)	51 (31.1)	50 (32.3)	17 (7.4)	12 (5.9)	56 (31.6)	5 (31.6)
Pantoprazole	2 (1.3)	1 (1.0)	9 (5.5)	9 (5.8)	3 (1.3)	1 (0.5)	16 (9.0)	16 (9.2)
Rabeprazole	3 (1.9)	1 (1.0)	12 (7.3)	12 (7.7)	2 (0.9)	0 (0.0)	9 (5.1)	9 (5.2)
Esomeprazole	11 (7.1)	3 (2.9)	25 (15.2)	25 (16.1)	8 (3.5)	3 (1.5)	15 (8.5)	15 (8.6)
H_2RAs, *n* (%)[a]								
Any H_2RA	4 (2.6)	3 (2.9)	9 (5.5)	9 (5.8)	5 (2.2)	2 (1.0)	13 (7.3)	13 (7.5)
Ranitidine	3 (1.9)	2 (1.9)	7 (4.3)	7 (4.5)	2 (0.9)	0 (0.0)	8 (4.5)	8 (4.6)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.6)
Cimetidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Nizatidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	1 (0.6)	1 (1.0)	2 (1.2)	2 (1.3)	3 (1.3)	2 (1.0)	5 (2.8)	5 (2.9)
Prokinetics, *n* (%)[a]								
Any prokinetic	6 (3.9)	2 (1.9)	4 (2.4)	4 (2.6)	5 (2.2)	4 (2.0)	6 (3.4)	5 (2.9)
Domperidone	4 (2.6)	1 (1.0)	4 (2.4)	4 (2.6)	1 (0.4)	0 (0.0)	5 (2.8)	4 (2.3)
Metoclopramide	2 (1.3)	1 (1.0)	0 (0.0)	0 (0.0)	4 (1.7)	4 (2.0)	1 (0.6)	1 (0.6)
Any reflux-related drug, *n* (%)[a]	58 (37.7)	15 (14.4)	148 (90.2)	144 (92.9)	46 (20.0)	22 (10.8)	165 (93.2)	163 (93.7)
Other prescribed drugs, *n*[b]								
Alginates	3	0	4	4	1	0	5	5
Antispasmodics (e.g. dicycloverine)	1	1	4	4	1	1	1	1
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0
Other ulcer- healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	1	1	0	0	0	0
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	2	2	6	5	4	4	3	3
Anti-motility	0	0	0	0	0	0	0	0

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 43 Follow-up at the time point equivalent to 2 years after surgery: medications

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	128	86	142	136	203	182	156	153
PPIs, n (%)[a]								
Any PPI	47 (36.7)	13 (15.1)	121 (85.2)	119 (87.5)	43 (21.2)	29 (15.9)	129 (82.7)	128 (83.7)
Omeprazole	17 (13.3)	4 (4.7)	43 (30.3)	42 (30.9)	11 (5.4)	8 (4.4)	48 (30.8)	47 (30.7)
Lansoprazole	17 (13.3)	3 (3.5)	39 (27.5)	39 (28.7)	15 (7.4)	13 (7.1)	42 (26.9)	42 (27.5)
Pantoprazole	1 (0.8)	1 (1.2)	6 (4.2)	5 (3.7)	2 (1.0)	1 (0.5)	12 (7.7)	12 (7.8)
Rabeprazole	4 (3.1)	1 (1.2)	12 (8.5)	12 (8.8)	3 (1.5)	0 (0.0)	6 (3.8)	6 (3.9)
Esomeprazole	8 (6.3)	4 (4.7)	20 (14.1)	20 (14.7)	9 (4.4)	5 (2.7)	14 (9.0)	14 (9.2)
H_2RAs, *n* (%)[a]								
Any H_2RA	2 (1.6)	1 (1.2)	5 (3.5)	5 (3.7)	1 (0.5)	0 (0.0)	13 (8.3)	13 (8.5)
Ranitidine	2 (1.6)	1 (1.2)	4 (2.8)	4 (2.9)	1 (0.5)	0 (0.0)	12 (7.7)	12 (7.8)
Famotidine	0 (0.0)	0 (0.0)	1 (0.7)	1 (0.7)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.7)
Cimetidine	0 (0.0)	0 (0.0)	1 (0.7)	1 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	1 (0.8)	0 (0.0)	2 (1.4)	2 (1.5)	2 (1.0)	2 (1.1)	2 (1.3)	2 (1.3)
Prokinetics, *n* (%)[a]								
Any prokinetic	5 (3.9)	4 (4.7)	7 (4.9)	6 (4.4)	5 (2.5)	3 (1.6)	4 (2.6)	4 (2.6)
Domperidone	5 (3.9)	4 (4.7)	3 (2.1)	3 (2.2)	2 (1.0)	0 (0.0)	3 (1.9)	3 (2.0)
Metoclopramide	0 (0.0)	0 (0.0)	4 (2.8)	3 (2.2)	3 (1.5)	3 (1.6)	1 (0.6)	1 (0.7)
Any reflux-related drug, *n* (%)[a]	48 (37.5)	14 (16.3)	124 (87.3)	122 (89.7)	46 (22.7)	30 (16.5)	140 (89.7)	139 (90.8)
Other prescribed drugs, *n*[b]								
Alginates	5	2	4	4	5	3	6	6
Antispasmodics (e.g. dicycloverine)	1	1	0	0	0	0	1	1
Chelates (e.g. sucralfate)	0	0	0	0	1	1	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	1	1
Mucogel	1	1	0	0	0	0	0	0
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	1	0	0	0	3	3	5	5
Anti-nausea	0	0	1	1	1	1	0	0
Anti-motility	0	0	0	0	0	0	0	0
Over the counter – not prescribed	6	0	8	7	2	1	6	6

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 44 Follow-up at the time point equivalent to 3 years after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	132	92	134	133	196	175	159	156
PPIs, *n* (%)[a]								
Any PPI	50 (37.9)	18 (19.6)	112 (83.6)	112 (84.2)	47 (24.0)	32 (18.3)	129 (81.1)	128 (82.1)
Omeprazole	19 (14.4)	8 (8.7)	43 (32.1)	43 (32.3)	16 (8.2)	9 (5.1)	51 (32.1)	50 (32.1)
Lansoprazole	13 (9.8)	1 (1.1)	37 (27.6)	37 (27.8)	20 (10.2)	16 (9.1)	45 (28.3)	45 (28.8)
Pantoprazole	3 (2.3)	3 (3.3)	4 (3.0)	4 (3.0)	3 (1.5)	3 (1.7)	10 (6.3)	10 (6.4)
Rabeprazole	5 (3.8)	2 (2.2)	8 (6.0)	8 (6.0)	1 (0.5)	0 (0.0)	6 (3.8)	6 (3.8)
Esomeprazole	8 (6.1)	5 (5.4)	20 (14.9)	20 (15.0)	6 (3.1)	4 (2.3)	12 (7.5)	12 (7.7)
H_2RAs, *n* (%)[a]								
Any H_2RA	2 (1.5)	1 (1.1)	2 (1.5)	2 (1.5)	2 (1.0)	0 (0.0)	10 (6.3)	10 (6.4)
Ranitidine	2 (1.5)	1 (1.1)	2 (1.5)	2 (1.5)	2 (1.0)	0 (0.0)	9 (5.7)	9 (5.8)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.6)
Over-the-counter H_2RA	1 (0.8)	0 (0.0)	2 (1.5)	2 (1.5)	6 (3.1)	2 (1.1)	4 (2.5)	2 (1.3)
Prokinetics, *n* (%)[a]								
Any prokinetic	3 (2.3)	2 (2.2)	6 (4.5)	6 (4.5)	5 (2.6)	2 (1.1)	4 (2.5)	4 (2.6)
Domperidone	2 (1.5)	2 (2.2)	4 (3.0)	4 (3.0)	2 (1.0)	0 (0.0)	3 (1.9)	3 (1.9)
Metoclopramide	1 (0.8)	0 (0.0)	2 (1.5)	2 (1.5)	3 (1.5)	2 (1.1)	1 (0.6)	1 (0.6)
Any reflux-related drug, *n* (%)[a]	51 (38.6)	18 (19.6)	113 (84.3)	113 (85.0)	47 (24.0)	32 (18.3)	135 (84.9)	134 (85.9)
Other prescribed drugs, *n*[b]								
Alginates	9	4	12	12	10	3	12	6
Antispasmodics (e.g. dicycloverine)	1	1	1	1	0	0	1	1
Chelates (e.g. sucralfate)	0	0	0	0	1	1	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	1	1
Mucogel	0	0	0	0	1	0	0	0
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	2	2	6	6	1	3	1	5
Anti-nausea	0	0	0	0	0	0	0	0
Anti-motility	0	0	0	0	1	1	0	0
Over the counter – not prescribed	0	0	1	1	2	1	2	6

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 45 Follow-up at the time point equivalent to 4 years after surgery: medications

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	126	88	129	127	168	152	142	139
PPIs, *n* (%)[a]								
Any PPI	52 (41.3)	21 (23.9)	104 (80.6)	104 (81.9)	42 (25.0)	30 (19.7)	118 (83.1)	117 (84.2)
Omeprazole	21 (16.7)	6 (6.8)	40 (31.0)	40 (31.5)	17 (10.1)	13 (8.6)	44 (31.0)	43 (30.9)
Lansoprazole	15 (11.9)	6 (6.8)	34 (26.4)	34 (26.8)	16 (9.5)	13 (8.6)	48 (33.8)	48 (34.5)
Pantoprazole	1 (0.8)	1 (1.1)	4 (3.1)	4 (3.1)	1 (0.6)	0 (0.0)	6 (4.2)	6 (4.3)
Rabeprazole	5 (4.0)	2 (2.3)	6 (4.7)	6 (4.7)	2 (1.2)	0 (0.0)	4 (2.8)	4 (2.9)
Esomeprazole	8 (6.3)	6 (6.8)	16 (12.4)	16 (12.6)	3 (1.8)	2 (1.3)	12 (8.5)	12 (8.6)
H_2RAs, *n* (%)[a]								
Any H_2Ra	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.8)	4 (2.4)	1 (0.7)	6 (4.2)	6 (4.3)
Ranitidine	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.8)	4 (2.4)	1 (0.7)	5 (3.5)	5 (3.6)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.7)	1 (0.7)
Over-the-counter H_2RA	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.8)	2 (1.3)	2 (1.4)	2 (1.4)
Prokinetics, *n* (%)[a]								
Any prokinetic	6 (4.8)	4 (4.5)	8 (6.2)	7 (5.5)	5 (3.0)	3 (2.0)	7 (4.9)	7 (5.0)
Domperidone	5 (4.0)	3 (3.4)	5 (3.9)	4 (3.1)	3 (1.8)	1 (0.7)	6 (4.2)	6 (4.3)
Metoclopramide	1 (0.8)	1 (1.1)	3 (2.3)	2 (1.6)	2 (1.2)	2 (1.3)	1 (0.7)	1 (0.7)
Any reflux-related drug, *n* (%)[a]	51 (40.5)	21 (23.9)	106 (82.2)	105 (82.7)	43 (25.6)	31 (20.4)	125 (88.0)	124 (89.2)
Other prescribed drugs, *n*[b]								
Alginates	12	4	11	11	12	9	13	13
Antispasmodics (e.g. dicycloverine)	2	2	2	1	0	0	0	0
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	0	0	0	0	0	0
Asilone	1	1	1	1	0	0	1	1
Non-gastrointestinal	4	4	3	1	1	0	0	0
Anti-nausea	0	0	1	1	1	1	0	0
Anti-motility	0	0	0	1	1	1	0	0
Over the counter – not prescribed	0	0	0	0	3	2	0	0

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 46 Follow-up at the time point equivalent to 5 years after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	127	90	119	116	176	158	136	133
PPIs, *n* (%)[a]								
Any PPI	55 (43.3)	23 (25.6)	98 (82.4)	97 (83.6)	48 (27.3)	36 (22.8)	116 (85.3)	113 (85.0)
Omeprazole	24 (18.9)	10 (11.1)	44 (37.0)	43 (37.1)	22 (12.5)	15 (9.5)	48 (35.3)	46 (34.6)
Lansoprazole	16 (12.6)	5 (5.6)	31 (26.1)	31 (26.7)	18 (10.2)	15 (9.5)	45 (33.1)	45 (33.8)
Pantoprazole	1 (0.8)	1 (1.1)	2 (1.7)	2 (1.7)	0 (0.0)	0 (0.0)	7 (5.1)	7 (5.3)
Rabeprazole	3 (2.4)	2 (2.2)	4 (3.4)	4 (3.4)	0 (0.0)	0 (0.0)	2 (1.5)	2 (1.5)
Esomeprazole	9 (7.1)	6 (6.7)	17 (14.3)	17 (14.7)	7 (4.0)	6 (3.8)	12 (8.8)	11 (8.3)
H_2RAs, *n* (%)[a]								
Any H_2RA	0 (0.0)	0 (0.0)	2 (1.7)	2 (1.7)	3 (1.7)	2 (1.3)	6 (4.4)	6 (4.5)
Ranitidine	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.9)	3 (1.7)	2 (1.3)	5 (3.7)	5 (3.8)
Famotidine	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	1 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	4 (2.3)	1 (0.6)	4 (2.9)	4 (3.0)
Prokinetics, *n* (%)[a]								
Any prokinetic	6 (4.7)	5 (5.6)	5 (4.2)	4 (3.4)	8 (4.5)	6 (3.8)	5 (3.7)	4 (3.0)
Domperidone	5 (3.9)	4 (4.4)	4 (3.4)	3 (2.6)	4 (2.3)	2 (1.3)	4 (2.9)	3 (2.3)
Metoclopramide	1 (0.8)	1 (1.1)	2 (1.7)	1 (0.9)	3 (1.7)	3 (1.9)	1 (0.7)	1 (0.8)
Any reflux-related drug, *n* (%)[a]	56 (44.1)	24 (26.7)	98 (82.4)	97 (83.6)	49 (27.8)	37 (23.4)	121 (89.0)	118 (88.7)
Other prescribed drugs, *n*[b]								
Alginates	11	3	11	11	12	9	15	15
Antispasmodics (e.g. dicycloverine)	1	1	0	0	1	0	1	1
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	0	0	0	0	0	0
Asilone	1	1	0	0	0	0	0	0
Non-gastrointestinal	1	1	3	1	1	0	1	1
Anti-nausea	0	0	0	0	1	0	0	0
Anti-motility	0	0	0	0	1	1	0	0
Over the counter – not prescribed	0	0	0	0	2	0	0	0

a Percentage is for responders completing the relevant section of the questionnaire.
b More than one prescription per person possible.

Appendix 4 Tables showing health status measures at each time point of follow-up

TABLE 47 Follow-up at the time point equivalent to 3 months after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/ allocated	178	111	179	169	261	218	192	189
Number of responders	149	97	157	141	229	186	182	168
REFLUX QoL, mean (SD)[a]	83.9 (19.4)	85.9 (19.0)	70.6 (24.6)	70.8 (24.4)	80.4 (21.6)	82.5 (20.3)	80.2 (18.2)	80.6 (17.7)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	84.8 (17.3)	89.4 (14.0)	66.9 (26.2)	66.5 (26.0)	84.1 (19.6)	87.2 (16.6)	75.7 (19.6)	76.0 (19.5)
Wind and frequency symptom score	58.1 (19.7)	55.9 (19.7)	53.7 (22.6)	54.4 (22.5)	52.2 (21.1)	52.6 (20.7)	60.7 (22.2)	60.9 (22.3)
Nausea and vomiting symptom score	91.5 (15.7)	93.1 (15.7)	82.1 (20.7)	82.3 (20.2)	90.2 (15.2)	91.6 (13.7)	89.5 (12.9)	90.0 (11.9)
Activity limitation symptom score	88.2 (17.0)	89.9 (16.7)	81.6 (19.6)	81.9 (19.0)	88.4 (18.0)	89.7 (17.5)	87.9 (13.2)	88.0 (13.3)
Constipation and swallowing symptom score	79.2 (20.0)	78.7 (20.7)	75.8 (20.9)	77.0 (19.8)	77.1 (21.2)	76.9 (21.3)	84.2 (16.9)	84.6 (16.5)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	49.2 (10.0)	49.3 (10.4)	46.5 (11.5)	46.6 (11.6)	49.9 (9.7)	50.4 (9.4)	47.6 (10.3)	47.5 (10.4)
Norm-based role physical	47.7 (11.8)	47.4 (12.1)	44.8 (12.1)	45.0 (12.1)	48.1 (11.3)	48.7 (10.7)	47.1 (10.4)	47.1 (10.4)
Norm-based bodily pain	48.5 (10.3)	48.8 (10.8)	45.3 (11.4)	45.3 (11.3)	48.4 (11.3)	49.0 (11.2)	46.5 (10.2)	46.5 (10.3)
Norm-based general health	46.3 (11.0)	47.4 (11.0)	40.7 (11.2)	40.7 (11.2)	47.2 (11.3)	48.2 (11.1)	42.5 (10.5)	42.6 (10.4)
Norm-based vitality	47.1 (11.9)	48.0 (12.1)	43.9 (12.4)	44.3 (12.2)	48.0 (11.9)	48.4 (11.9)	44.7 (11.4)	44.8 (11.4)
Norm-based social functioning	47.2 (11.5)	47.5 (12.1)	43.6 (12.7)	43.8 (12.6)	46.8 (12.3)	47.6 (12.0)	46.9 (10.5)	46.9 (10.5)
Norm-based role emotional	48.3 (12.3)	48.4 (12.5)	43.9 (14.2)	44.1 (14.2)	47.0 (12.6)	48.9 (11.7)	47.0 (11.4)	46.9 (11.4)
Norm-based mental health	48.7 (12.0)	49.7 (11.9)	44.5 (12.2)	44.7 (11.9)	48.3 (12.2)	49.2 (11.8)	47.1 (10.6)	47.1 (10.7)
EQ-5D, mean (SD)[a]	0.788 (0.233)	0.806 (0.239)	0.689 (0.301)	0.696 (0.299)	0.806 (0.245)	0.817 (0.240)	0.763 (0.231)	0.765 (0.229)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 48 Follow-up at the time point equivalent to 12 months after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	154	98	165	149	232	192	181	169
REFLUX QoL, mean (SD)[a]	84.6 (17.9)	88.3 (15.6)	73.4 (23.3)	73.1 (23.7)	83.3 (20.7)	86.0 (17.9)	79.2 (19.2)	79.4 (19.0)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	84.7 (17.5)	90.2 (14.0)	67.4 (25.8)	66.7 (25.8)	85.0 (19.4)	87.7 (16.5)	73.9 (20.7)	74.0 (20.8)
Wind and frequency symptom score	56.7 (21.0)	56.9 (21.7)	52.6 (23.3)	52.7 (23.5)	56.9 (22.5)	57.5 (22.1)	61.4 (21.9)	61.5 (22.0)
Nausea and vomiting symptom score	91.9 (14.4)	94.7 (11.8)	84.0 (18.6)	83.3 (18.8)	91.1 (16.5)	93.3 (13.8)	88.6 (15.4)	88.9 (14.4)
Activity limitation symptom score	90.7 (12.8)	93.3 (11.5)	82.2 (19.2)	81.6 (19.4)	90.8 (16.8)	92.4 (14.8)	87.3 (14.7)	87.4 (14.8)
Constipation and swallowing symptom score	79.3 (19.1)	80.2 (19.6)	74.5 (22.8)	75.2 (22.3)	78.5 (20.2)	79.1 (19.7)	83.6 (17.6)	83.8 (17.4)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	48.9 (10.3)	49.6 (10.3)	47.2 (11.0)	47.2 (10.9)	49.7 (10.8)	50.3 (10.5)	47.4 (10.5)	47.4 (10.6)
Norm-based role physical	46.7 (11.4)	47.4 (11.3)	45.8 (11.8)	46.0 (11.7)	49.0 (11.2)	49.6 (10.5)	46.8 (10.7)	46.8 (10.7)
Norm-based bodily pain	47.7 (10.4)	48.5 (10.7)	44.5 (10.9)	44.5 (10.9)	49.1 (11.3)	49.9 (11.1)	47.4 (9.9)	47.4 (10.0)
Norm-based general health	45.2 (11.1)	46.2 (11.8)	40.7 (11.2)	40.5 (11.1)	46.4 (10.8)	47.2 (10.6)	42.3 (10.1)	42.3 (10.1)
Norm-based vitality	46.9 (11.5)	47.6 (11.6)	44.2 (11.9)	44.4 (11.7)	47.3 (12.0)	48.0 (11.7)	45.1 (10.3)	45.2 (10.3)
Norm-based social functioning	46.9 (11.6)	47.8 (11.7)	45.2 (12.2)	45.4 (12.1)	46.9 (12.5)	47.8 (12.1)	46.6 (10.6)	46.6 (10.6)
Norm-based role emotional	46.4 (13.5)	47.2 (12.9)	44.2 (14.4)	44.4 (14.2)	47.3 (13.3)	48.1 (12.7)	46.2 (12.0)	46.1 (12.0)
Norm-based mental health	47.2 (11.7)	48.5 (11.6)	46.4 (12.1)	46.5 (12.2)	46.9 (12.0)	47.4 (12.0)	46.5 (10.9)	46.6 (10.9)
EQ-5D, mean (SD)[a]	0.754 (0.247)	0.777 (0.232)	0.709 (0.272)	0.710 (0.270)	0.791 (0.263)	0.803 (0.252)	0.741 (0.240)	0.743 (0.238)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 49 Follow-up at the time point equivalent to 2 years after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	128	86	142	136	203	182	156	153
REFLUX QoL, mean (SD)[a]	85.5 (17.3)	89.2 (15.1)	76.9 (22.8)	77.0 (22.9)	85.3 (19.0)	87.1 (17.5)	80.4 (19.1)	80.5 (19.1)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	83.1 (18.5)	88.2 (15.0)	71.8 (25.4)	71.5 (25.4)	85.7 (19.9)	88.2 (17.1)	75.2 (19.5)	75.1 (19.6)
Wind and frequency symptom score	57.1 (20.0)	57.1 (20.0)	54.9 (24.1)	55.4 (24.4)	56.1 (23.2)	56.9 (22.5)	61.1 (21.6)	61.3 (21.5)
Nausea and vomiting symptom score	92.4 (13.1)	94.3 (11.6)	86.3 (18.1)	86.3 (18.4)	91.9 (14.6)	93.0 (13.1)	89.3 (14.0)	89.5 (13.4)
Activity limitation symptom score	91.2 (12.1)	93.7 (10.5)	83.3 (20.6)	83.6 (20.2)	92.4 (14.7)	93.3 (13.5)	86.8 (15.6)	86.9 (15.6)
Constipation and swallowing symptom score	80.5 (19.5)	81.3 (20.2)	77.6 (22.5)	77.5 (22.9)	80.1 (21.0)	80.7 (20.7)	81.5 (17.6)	81.6 (17.7)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	48.4 (9.8)	48.9 (9.7)	46.7 (11.3)	47.0 (11.0)	49.0 (11.0)	49.6 (10.2)	46.2 (11.7)	46.2 (11.7)
Norm-based role physical	48.6 (10.6)	49.2 (10.0)	45.8 (12.2)	46.0 (12.0)	49.3 (10.8)	49.9 (10.1)	46.1 (11.1)	45.9 (11.2)
Norm-based bodily pain	47.6 (9.3)	48.1 (8.8)	44.8 (10.7)	45.1 (10.6)	47.7 (10.0)	48.2 (9.6)	45.7 (9.1)	45.9 (9.1)
Norm-based general health	44.6 (11.1)	45.2 (11.9)	41.3 (11.4)	41.3 (11.3)	46.5 (10.7)	46.9 (10.4)	41.6 (10.4)	41.7 (10.3)
Norm-based vitality	46.6 (10.7)	46.9 (10.9)	43.4 (11.6)	43.4 (11.3)	47.1 (11.6)	47.5 (11.4)	44.4 (10.6)	44.4 (10.5)
Norm-based social functioning	47.4 (11.4)	48.1 (11.2)	45.3 (12.0)	45.5 (11.7)	47.7 (12.1)	48.1 (11.8)	46.2 (11.6)	46.2 (11.6)
Norm-based role emotional	48.4 (11.8)	49.0 (11.6)	45.5 (14.1)	45.8 (13.8)	48.9 (11.8)	49.8 (10.7)	45.5 (12.4)	45.3 (12.4)
Norm-based mental health	47.9 (11.5)	47.9 (12.0)	45.6 (11.7)	45.9 (11.6)	48.0 (12.0)	48.8 (11.5)	45.6 (10.8)	45.5 (10.8)
EQ-5D, mean (SD)[a]	0.762 (0.272)	0.790 (0.244)	0.717 (0.313)	0.721 (0.308)	0.796 (0.257)	0.816 (0.233)	0.736 (0.235)	0.735 (0.237)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 50 Follow-up at the time point equivalent to 3 years after surgery: health status

Health status measure	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	132	92	134	133	196	175	159	156
REFLUX QoL, mean (SD)[a]	87.0 (15.0)	88.0 (15.3)	79.9 (20.1)	79.7 (20.1)	85.6 (18.2)	87.3 (17.2)	81.9 (16.4)	81.9 (16.5)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	85.4 (17.4)	88.4 (17.1)	74.8 (23.1)	74.6 (23.1)	85.7 (19.6)	87.3 (18.6)	77.4 (18.7)	77.2 (18.7)
Wind and frequency symptom score	59.5 (22.9)	57.9 (23.6)	56.1 (25.5)	55.8 (25.4)	55.3 (22.5)	55.9 (22.4)	62.0 (22.8)	62.1 (22.7)
Nausea and vomiting symptom score	94.0 (10.2)	95.4 (9.0)	89.2 (16.2)	89.1 (16.2)	92.1 (14.8)	93.0 (14.3)	90.3 (13.4)	90.4 (13.3)
Activity limitation symptom score	91.6 (13.2)	93.0 (13.3)	87.6 (16.7)	87.5 (16.7)	92.7 (12.1)	93.6 (11.6)	88.9 (12.9)	88.9 (13.0)
Constipation and swallowing symptom score	82.1 (16.8)	81.1 (17.0)	79.6 (20.1)	79.4 (20.1)	78.4 (21.7)	78.7 (21.4)	83.1 (17.4)	83.3 (17.4)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	49.1 (10.2)	49.8 (9.9)	47.8 (11.3)	47.8 (11.3)	49.3 (10.6)	49.6 (10.4)	46.9 (11.2)	46.9 (11.3)
Norm-based role physical	48.1 (10.9)	48.1 (11.1)	47.0 (11.4)	46.9 (11.4)	48.5 (11.3)	48.9 (10.9)	46.7 (11.4)	46.5 (11.4)
Norm-based bodily pain	47.4 (9.7)	46.3 (9.8)	46.3 (10.3)	46.3 (10.3)	48.1 (10.2)	48.6 (10.0)	46.4 (9.1)	46.3 (9.1)
Norm-based general health	45.3 (10.0)	45.8 (10.1)	42.4 (11.8)	42.3 (11.8)	46.2 (11.2)	46.6 (11.1)	41.8 (10.2)	41.8 (10.2)
Norm-based vitality	46.0 (11.5)	46.8 (11.2)	44.7 (12.7)	44.6 (12.7)	47.1 (11.7)	47.7 (11.6)	44.5 (10.3)	44.5 (10.3)
Norm-based social functioning	48.5 (10.4)	48.7 (10.6)	46.2 (11.9)	46.1 (11.9)	47.6 (12.4)	47.9 (12.3)	47.0 (11.3)	46.9 (11.3)
Norm-based role emotional	49.6 (9.9)	49.2 (10.4)	45.9 (13.3)	45.8 (13.4)	48.0 (12.9)	48.2 (12.8)	47.0 (11.8)	46.8 (11.8)
Norm-based mental health	49.5 (10.8)	49.7 (11.0)	46.1 (12.0)	46.0 (12.0)	47.9 (12.0)	48.4 (11.8)	46.6 (10.6)	46.5 (10.6)
EQ-5D, mean (SD)[a]	0.803 (0.231)	0.790 (0.252)	0.747 (0.262)	0.745 (0.262)	0.803 (0.249)	0.805 (0.251)	0.763 (0.231)	0.761 (0.232)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 51 Follow-up at the time point equivalent to 4 years after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	126	88	129	127	168	152	142	139
REFLUX QoL, mean (SD)[a]	85.2 (18.2)	87.7 (17.7)	81.1 (20.7)	81.9 (19.4)	86.2 (16.8)	86.9 (16.3)	83.7 (17.2)	83.7 (17.2)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	83.1 (20.8)	87.0 (20.2)	77.4 (21.9)	77.5 (21.8)	85.7 (19.4)	87.5 (17.4)	79.2 (19.7)	79.0 (19.8)
Wind and frequency symptom score	58.1 (22.1)	57.0 (23.1)	53.6 (23.9)	53.2 (23.7)	55.3 (23.0)	55.7 (22.8)	62.4 (23.4)	62.5 (23.5)
Nausea and vomiting symptom score	91.0 (16.7)	93.6 (14.9)	89.2 (16.4)	89.4 (16.1)	92.9 (12.7)	94.0 (11.3)	91.4 (11.7)	91.6 (11.4)
Activity limitation symptom score	91.8 (13.1)	93.6 (12.4)	87.5 (17.2)	88.1 (15.3)	92.1 (13.7)	92.9 (12.6)	90.4 (12.9)	90.4 (12.9)
Constipation and swallowing symptom score	80.4 (19.4)	79.7 (20.0)	79.6 (20.9)	79.8 (20.6)	79.0 (21.9)	79.5 (20.9)	82.8 (18.1)	82.9 (18.1)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	47.9 (10.1)	48.5 (10.4)	47.5 (11.7)	47.7 (11.4)	49.5 (10.6)	50.0 (10.1)	47.2 (10.4)	47.1 (10.5)
Norm-based role physical	47.5 (11.9)	47.1 (12.4)	46.7 (12.2)	46.8 (12.0)	49.4 (10.6)	49.8 (10.2)	47.6 (10.3)	47.5 (10.4)
Norm-based bodily pain	46.1 (10.4)	46.4 (11.0)	46.3 (10.2)	46.4 (10.0)	47.8 (10.1)	48.1 (10.1)	47.9 (9.0)	47.9 (9.0)
Norm-based general health	44.6 (10.4)	45.5 (11.0)	42.2 (11.4)	42.4 (11.2)	46.5 (11.2)	47.0 (10.9)	42.2 (11.4)	42.1 (11.5)
Norm-based vitality	44.8 (10.7)	44.8 (10.9)	45.6 (11.7)	45.6 (11.4)	47.4 (11.6)	48.1 (11.5)	45.1 (11.0)	45.1 (11.0)
Norm-based social functioning	45.6 (12.7)	45.8 (12.9)	46.1 (11.9)	46.3 (11.6)	48.3 (11.1)	48.9 (10.6)	46.7 (11.3)	46.8 (11.4)
Norm-based role emotional	48.1 (12.4)	48.7 (12.4)	46.5 (14.1)	46.7 (13.8)	48.9 (11.7)	49.4 (11.3)	47.2 (11.5)	47.3 (11.5)
Norm-based mental health	47.5 (11.8)	49.1 (11.2)	47.0 (12.0)	47.2 (11.6)	48.7 (11.5)	49.1 (11.3)	47.3 (11.2)	47.4 (11.2)
EQ-5D, mean (SD)[a]	0.771 (0.244)	0.778 (0.264)	0.754 (0.272)	0.760 (0.258)	0.806 (0.254)	0.825 (0.229)	0.773 (0.213)	0.773 (0.215)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 52 Follow-up at the time point equivalent to 5 years after surgery: health status

Health status measure	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	127	90	119	116	176	158	136	133
REFLUX QoL, mean (SD)[a]	86.7 (13.8)	89.8 (11.7)	80.7 (20.3)	80.6 (20.4)	85.3 (17.3)	86.2 (17.1)	84.8 (15.2)	85.0 (15.3)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	85.0 (17.5)	89.6 (14.2)	75.3 (22.6)	75.0 (22.7)	85.7 (19.0)	86.7 (18.6)	78.8 (19.2)	78.8 (19.3)
Wind and frequency symptom score	58.8 (21.8)	58.7 (22.5)	56.4 (22.7)	55.8 (22.5)	54.0 (23.5)	53.9 (23.3)	65.3 (22.2)	65.7 (22.0)
Nausea and vomiting symptom score	92.5 (12.7)	94.9 (10.1)	89.6 (15.1)	89.9 (14.5)	93.2 (11.8)	93.8 (10.8)	92.1 (11.3)	92.4 (10.9)
Activity limitation symptom score	93.2 (11.4)	95.3 (9.5)	87.7 (18.5)	88.3 (16.6)	92.6 (13.6)	93.4 (13.2)	91.2 (11.9)	91.3 (11.9)
Constipation and swallowing symptom score	81.2 (18.7)	81.0 (18.9)	78.3 (20.4)	78.8 (19.8)	80.3 (19.9)	80.5 (19.8)	84.3 (17.9)	84.5 (18.0)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	48.4 (9.6)	48.3 (9.9)	48.2 (11.1)	48.4 (11.0)	49.3 (10.6)	49.8 (10.2)	47.0 (12.0)	46.9 (12.1)
Norm-based role physical	47.3 (11.9)	47.3 (12.4)	47.7 (11.8)	47.9 (11.6)	49.7 (10.1)	50.1 (9.9)	47.9 (10.0)	47.8 (10.0)
Norm-based bodily pain	46.3 (10.3)	47.3 (11.1)	46.2 (10.9)	46.6 (10.7)	47.9 (10.5)	48.0 (10.5)	48.0 (8.8)	47.9 (8.9)
Norm-based general health	44.1 (10.3)	44.9 (10.6)	43.2 (11.5)	43.4 (11.4)	47.0 (10.8)	47.2 (10.6)	43.3 (9.2)	43.2 (9.3)
Norm-based vitality	45.3 (11.2)	46.0 (11.5)	46.4 (12.0)	46.5 (12.0)	47.7 (11.9)	47.8 (11.9)	45.2 (10.9)	45.2 (11.0)
Norm-based social functioning	47.4 (11.7)	48.0 (12.1)	47.0 (12.0)	47.3 (11.7)	48.4 (11.4)	48.7 (11.3)	48.2 (10.4)	48.1 (10.4)
Norm-based role emotional	48.6 (11.8)	49.0 (11.7)	47.5 (12.6)	47.7 (12.2)	49.8 (10.0)	50.1 (9.7)	48.2 (10.6)	48.1 (10.6)
Norm-based mental health	47.7 (11.9)	48.8 (11.8)	48.9 (11.5)	49.2 (11.2)	49.4 (11.6)	49.4 (11.6)	47.5 (10.3)	47.4 (10.3)
EQ-5D, mean (SD)[a]	0.774 (0.259)	0.777 (0.281)	0.761 (0.282)	0.770 (0.269)	0.800 (0.253)	0.807 (0.249)	0.794 (0.206)	0.793 (0.208)

a Mean (SD) based on responders completing relevant section of the questionnaire.

Appendix 5 Characteristics of the four randomised controlled trials of laparoscopic fundoplication compared with medical management

Anvari *et al.* trial[44–46]

Methods	**Randomisation:** computerised sequence generation **Allocation concealment:** apparently yes, although blocking used to ensure 1 : 1 randomisation ('blocking factor determined by data centre') **Blinding:** not possible; outcome assessment: at office visit (questionnaires before medical assessment) at 6 and 12 months, by telephone at 3 and 9 months **Follow-up:** 3, 6, 9 and 12 months and 3 years **Setting:** single centre in Canada (four experienced surgeons) **Inclusion criteria:** chronic symptoms of GORD requiring long-term therapy; dependent on PPIs for at least 12 months; adults aged 18–70 years; GORD symptom score of < 18 and a score of > 70 on visual analogue scale (VAS) (0–100) of symptom control at screening; % acid reflux > 4% at baseline **Exclusion criteria:** pregnancy, malignancy, aperistaltic esophagus, severe comorbidity and previous GORD surgery
Participants	**Sample size:** 216 (a priori) **Randomised:** 104; medical: 52 [50 received medication (96%)], surgical: 52 [51 received surgery (98%)] **Age, mean:** medical 42.1 years; surgical 42.9 years **Sex (M/F):** medical 26/26; surgical 29/23
Interventions	**Medical:** optimised PPI as per detailed symptom management algorithm **Surgical:** laparoscopic Nissen fundoplication. Comprised construction of 2.5- to 3-cm 360° wrap. Short gastric vessels divided routinely to achieve floppy wrap
Outcomes	**Primary outcome:** GERSS – includes heartburn, regurgitation, bloating, dysphagia and epigastric/retrosternal pain. Total scale score 0–60. Well controlled defined as score < 18 **Secondary outcomes:** oesophageal function: endoscopy, manometry and 24-hour pH; QoL: SF-36 (0–100), EQ-5D (0–1) and VAS 0–100 for patient satisfaction with symptom control. A score of 70 was considered the threshold for symptom control on the VAS
Type of trial design	On explanatory end of explanatory–pragmatic continuum
Clinical leadership	Upper gastrointestinal surgeon
Risk of bias	
Allocation concealment?	Probably concealed – explanation of randomisation and concealment given in methods, although blocking could have jeopardised this
Free of selective reporting?	One concern: heartburn-free days promoted to primary outcome at 3 years
Sequence generation?	Computerised sequence generation but blocked and size of block not stated
Incomplete outcome data addressed?	Some evidence to suggest differential loss to follow-up at 3 years: 8/52 vs 3/52; no responder analysis
Notes	Trial funded by the Canadian Institute of Health Research and Ontario Ministry of Health

LOTUS trial[47–50]

Methods	**Randomisation:** randomisation in blocks of four **Allocation concealment:** unclear **Blinding:** not possible; outcome assessment: primary outcome (treatment failure) dependent on clinical decision-making, which was not blinded **Follow-up:** 6 months and 1, 3 and 5 years **Setting:** 39 centres across 11 European countries **Inclusion criteria:** oesophagitis grade no more than Los Angeles grade B; GORD symptoms no more than mild; response to PPI in run-in phase **Exclusion criteria:** previous oesophageal, gastric or duodenal surgery; primary oesophageal disorders; inflammatory bowel disorders; any gastrointestinal absorption abnormality; other significant concomitant disease
Participants	**Sample size:** 550 – not clear if stated a priori **Randomised:** 554; medical: 266, surgical: 288 [248 received surgery (86%)] – specialist surgery **Age, mean (SD):** medical 45.4 (11.5) years; surgical 44.8 (10.9) years **Sex (M/F):** medical 199/67; surgical 199/89
Interventions	**Medical:** esomeprazole 20 mg once daily, which could be increased stepwise **Surgical:** laparoscopic anti-reflux surgery. Used crural repair and short floppy total fundoplication in standardised approach
Outcomes	**Primary outcome:** time to treatment failure **Secondary outcomes:** symptoms related to GORD (heartburn, acid regurgitation and dysphagia severity); other gastrointestinal symptoms (flatulence, diarrhoea, epigastric pain, bloating) from GSRS; endoscopy; QoL using QOLRAD; perioperative and postoperative mortality (<30 days); dysphagia requiring further treatment; serious adverse events; rate of conversion to open surgery
Type of trial design	Principally explanatory with some pragmatic features (calls itself 'exploratory')
Clinical leadership	Upper gastrointestinal surgeon
Risk of bias	
Allocation concealment?	Unclear; randomisation in blocks of four, otherwise not reported
Free of selective reporting?	No evidence of selective reporting, although QOLRAD data only reported in supplementary table at 5 years
Sequence generation?	Unclear; randomisation in blocks of four
Incomplete outcome data addressed?	Not fully: follow-up at 3 years: 204/288 vs 208/266; at 5 years: 180/288 (62.5%) vs 192/266 (72.2%). No data on 14% allocated surgery who did not have an operation
Notes	Trial funded by AstraZeneca R&D, with three authors employed by AstraZeneca

Mahon *et al.* trial[51–53]

Methods	**Randomisation:** 'computerised randomisation' – no details **Allocation concealment:** unclear, not reported **Blinding:** not possible **Follow-up:** 3 months and 1 year; separate follow-up of participants from one centre at 7 years **Setting:** two UK centres (two experienced surgeons) **Inclusion criteria:** GORD for at least 6 months, dependent on PPIs for at least 3 months and aged >16 to <70 years **Exclusion criteria:** significant oesophageal dysmotility and morbid obesity (BMI > 35 kg/m^2)
Participants	**Sample size:** a priori apparently 215 although basis not clear **Randomised**: 217; medical: 108, surgical: 109 (apparently all received surgery) **Age, median (range):** medical 47 (35–57) years; surgical 48 (39–56) years **Sex (M : F ratio):** medical 1 : 2.6; surgical 1 : 1.9
Interventions	**Medical:** one of four different PPI regimens, aiming to abolish symptoms **Surgical:** laparoscopic Nissen fundoplication. Used crural repair and short floppy wrap of 3 cm; division of short gastric vessels as deemed necessary
Outcomes	PGWI, GSRS, dysphagia, DeMeester score, operation time, length of stay, conversion to open surgery, reoperation rate, mortality rate, lower oesophageal sphincter pressure, postoperative complications, % time pH <4, cost, patient satisfaction only at 7 years (scale 1–3)
Type of trial design	At explanatory end of explanatory–pragmatic continuum
Clinical leadership	Upper gastrointestinal surgeon
Risk of bias	
Allocation concealment?	Unclear, not reported
Free of selective reporting?	Unclear, primary outcome not clearly prespecified
Sequence generation?	'Computerised randomisation'
Incomplete outcome data addressed?	Among 108 in medical group, well-being scores were available for 108 at baseline and 96 at one year; equivalent figures among 109 in surgical group were 104 and 99, respectively
Notes	Trial partially funded by Jansen Pharmaceutics; economic evaluation funded by Ethicon Endo-Surgery. All participants in medical group offered surgery at 1 year: 54/92 (59%) underwent surgery

REFLUX trial[1–3]

Methods	**Randomisation:** computer-generated sequence **Allocation concealment:** yes **Blinding:** not possible; outcome assessment by patient-completed postal questionnaires **Follow-up:** 3 months and annually for 5 years **Setting:** 21 UK centres **Inclusion criteria:** GORD symptoms for >12 months requiring PPI; evidence of GORD (endoscopy and/or pH monitoring) **Exclusion criteria:** BMI >40 kg/m^2; Barrett's esophagus >3 cm; paraoesophageal hernia; oesophageal stricture
Participants	**Sample size:** 600 (sample size recalculated from 600 to 392 after advice from DMC) **Randomised:** 357; medical: 179, surgical: 178 [111 received surgery (62%)] – by, or supervised by, experienced surgeon **Age, mean (SD):** medical 45.9 (11.9) years; surgical 46.7 (10.3) years **Sex (M/F):** medical 120/59; surgical 116/62
Interventions	**Medical:** best medical management after review. Lansoprazole was predominant PPI at study entry; omeprazole and lansoprazole most commonly reported at follow-up **Surgical:** laparoscopic surgery. Type of fundoplication was left to discretion of surgeon and all surgical techniques considered as a single policy
Outcomes	**Primary outcome:** REFLUX questionnaire score (heartburn, acid reflux, wind, eating and swallowing, bowel movements, sleep, work, physical and social activity) **Secondary outcomes:** QoL: EQ-5D and SF-36; serious morbidity; mortality; patient costs; NHS costs
Type of trial design	Pragmatic on explanatory–pragmatic continuum. Also included parallel, non-randomised preference groups
Clinical leadership	Upper gastrointestinal surgeon and gastroenterologist partnerships
Risk of bias	
Allocation concealment?	Allocation conducted by trials unit independent of all clinical teams
Free of selective reporting?	ITT and PP analysis presented as prespecified
Sequence generation?	Computerised randomisation
Incomplete outcome data addressed?	Adjusted treatment received and PP analyses reported in addition to ITT. Follow-up at 12 months: 154/178 (87%) vs 164/179 (92%)
Notes	Trial funded by NIHR HTA programme

Appendix 6 Search strategies for economic evaluation review

Gastro-oesophageal reflux disease search terms used in a recent Cochrane Review and adapted for use in the systematic review described in *Chapter 5*.[57]

Economic evaluation search

The Cochrane Library (includes NHS Economic Evaluation Database)

http://onlinelibrary.wiley.com/o/cochrane/cochrane_search_fs.html
Searched: 19 April 2011.

#1 MeSH descriptor Gastroesophageal Reflux explode all trees (1356)
#2 (gastroesophageal near/3 reflux):ti,ab,kw (1764)
#3 (gastro near/3 oesophageal near/3 reflux):ti,ab,kw (657)
#4 (gastro near/3 esophageal near/3 reflux):ti,ab,kw (657)
#5 (gord):ti,ab,kw (103)
#6 (gerd):ti,ab,kw (413)
#7 MeSH descriptor Duodenogastric Reflux explode all trees (50)
#8 (duodenogastric near/3 reflux):ti,ab,kw (58)
#9 MeSH descriptor Bile Reflux explode all trees (22)
#10 (bile near/3 reflux):ti,ab,kw (78)
#11 (acid near/3 reflux):ti,ab,kw (281)
#12 MeSH descriptor Dyspepsia explode all trees (864)
#13 (dyspep*):ti,ab,kw (2165)
#14 (belch* or burp*):ti,ab,kw 100
#15 MeSH descriptor Eructation explode all trees (18)
#16 (eructation):ti,ab,kw (52)
#17 MeSH descriptor Heartburn explode all trees (255)
#18 (heartburn or indigestion):ti,ab,kw (985)
#19 MeSH descriptor Esophagitis explode all trees (583)
#20 (esophagitis or oesophagitis):ti,ab,kw (1273)
#21 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20), from 2005 to 2011 (1367)

1367 results made up of:

- Economic evaluations (NHS EED): 85
- Cochrane reviews (CDSR): 54
- Other systematic reviews (DARE): 59
- Technology assessments (HTA): 13
- Clinical trials (Cochrane Central Register of Controlled Trials, CENTRAL): 1147
- Methods studies (Cochrane Methodology Register): 9.

A total of 10 of the CDSR records were pre 2005 and so they were deleted. In addition to the 85 NHS EED records, all CDSR, DARE and HTA records were also saved to EndNote library reflux.enl (marked CDSR, DARE, HTA or NHS EED in the Custom 4 field) in case they are useful (Thomson Reuters, CA, USA).

Quality-of-life searches

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations

Ovid MEDLINE(R) and Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations – 1948 to present.
Searched: 20 April 2011 via OVID interface.

1. exp gastroesophageal reflux/ (18,908)
2. (gastroesophageal adj3 reflux).tw. (11,313)
3. (gastro adj3 oesophageal adj3 reflux).tw. (3146)
4. (gastro adj3 esophageal adj3 reflux).tw. (898)
5. gord.tw. (562)
6. gerd.tw. (4147)
7. exp duodenogastric reflux/ (1511)
8. (duodenogastric adj3 reflux).tw. (813)
9. exp bile reflux/ (649)
10. (bile adj3 reflux).tw. (895)
11. (acid adj3 reflux).tw. (2044)
12. exp dyspepsia/ (6549)
13. dyspep$.tw. (9171)
14. (belch$ or burp$).tw. (771)
15. exp eructation/ (257)
16. eructation.tw. (173)
17. exp heartburn/ (1395)
18. (heartburn or indigestion).tw. (3760)
19. exp esophagitis/ (8478)
20. (esophagitis or oesophagitis).tw. (10,081)
21. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 (44,775)
22. exp life tables/ (10,574)
23. "quality of life"/ (89,134)
24. health status/ (47,125)
25. exp health status indicators/ (150,921)
26. (utilit$ approach$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab. (1099)
27. (health measurement$ scale$ or health measurement$ questionnaire$).ti,ab. (31)
28. (standard gamble$ or categor$ scal$ or linear scal$ or linear analog$ or visual scal$ or magnitude estimat$).ti,ab. (3759)
29. (time trade off$ or rosser$ classif$ or rosser$ matrix or rosser$ distress$ or hrqol).ti,ab. (5160)
30. (index of wellbeing or quality of wellbeing or qwb).ti,ab. (150)
31. (rating scale$ or multiattribute$ health ind$ or multi attribute$ health ind$).ti,ab. (26,439)
32. (health utilit$ index or health utilit$ indices).ti,ab. (484)
33. (multiattribute$ theor$ or multi attribute$ theor$ or multiattribute$ analys$ or multi attribute$ analys$).ti,ab. (9)
34. (health utilit$ scale$ or classification of illness state$ or 15d or 15 d or 15 dimension).ti,ab. (2878)
35. (health state$ utilit$ or 12d or 12 d or 12 dimension).ti,ab. (2009)
36. well year$.ti,ab. (20)
37. (multiattribute$ utilit$ or multi attribute$ utilit$).ti,ab. (152)
38. health utilit$ scale$.ti,ab. (7)
39. (qol or 5d or 5-d or 5 dimension or quality of life or eq-5d or eq5d or eq 5d or euroqol).ti,ab. (114,629)
40. (qualy or qaly or qualys or qalys or quality adjusted life year$).ti,ab. (4631)
41. life year$ gain$.ti,ab. (1289)
42. willingness to pay.ti,ab. (1517)
43. (hye or hyes or health$ year$ equivalent$).ti,ab. (58)

44. (person trade off$ or person tradeoff$ or time tradeoff$ or time trade off$).ti,ab. (776)
45. theory utilit$.ti,ab. (7)
46. life table$.ti,ab. (6627)
47. health state$.ti,ab. (2838)
48. (sf36 or sf 36).ti,ab. (9654)
49. (short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).ti,ab. (4429)
50. (6d or 6-d or 6 dimension).ti,ab. (4590)
51. or/22-50 (366,981)
52. 21 and 51 (3337)
53. limit 52 to yr="2005 - 2011" (1726)

A total of 1726 results saved to EndNote library reflux.enl (marked MEDLINE in the Custom 4 field).

EMBASE

EMBASE – 1996 to week 15 2011.
Searched: 20 April 2011 via OVID interface.

1. exp gastroesophageal reflux/ (24,724)
2. (gastroesophageal adj3 reflux).tw. (10,538)
3. (gastro adj3 oesophageal adj3 reflux).tw. (2717)
4. (gastro adj3 esophageal adj3 reflux).tw. (867)
5. gord.tw. (670)
6. gerd.tw. (5466)
7. exp duodenogastric reflux/ (993)
8. (duodenogastric adj3 reflux).tw. (292)
9. exp bile reflux/ (509)
10. (bile adj3 reflux).tw. (505)
11. (acid adj3 reflux).tw. (1939)
12. exp dyspepsia/ (15,437)
13. dyspep$.tw. (7897)
14. (belch$ or burp$).tw. (706)
15. exp eructation/ (315)
16. eructation.tw. (100)
17. exp heartburn/ (5566)
18. (heartburn or indigestion).tw. (3660)
19. exp esophagitis/ (12,163)
20. esophagitis.tw. (6246)
21. oesophagitis.tw. (1696)
22. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 (51,561)
23. life tables/ (1840)
24. exp "quality of life"/ (158,965)
25. health status/ (53,662)
26. health survey/ (104,329)
27. (utilit$ approach$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab. (1100)
28. (health measurement$ scale$ or health measurement$ questionnaire$).ti,ab. (35)
29. (standard gamble$ or categor$ scal$ or linear scal$ or linear analog$ or visual scal$ or magnitude estimat$).ti,ab. (2771)
30. (time trade off$ or rosser$ classif$ or rosser$ matrix or rosser$ distress$ or hrqol).ti,ab. (6267)
31. (index of wellbeing or quality of wellbeing or qwb).ti,ab. (139)
32. (rating scale$ or multiattribute$ health ind$ or multi attribute$ health ind$).ti,ab. (25,950)
33. (health utilit$ index or health utilit$ indices).ti,ab. (550)

34. (multiattribute$ theor$ or multi attribute$ theor$ or multiattribute$ analys$ or multi attribute$ analys$).ti,ab. (9)
35. (health utilit$ scale$ or classification of illness state$ or 15d or 15 d or 15 dimension).ti,ab. (2773)
36. (health state$ utilit$ or 12d or 12 d or 12 dimension).ti,ab. (1630)
37. well year$.ti,ab. (7)
38. (multiattribute$ utilit$ or multi attribute$ utilit$).ti,ab. (132)
39. health utilit$ scale$.ti,ab. (5)
40. (qol or 5d or 5-d or 5 dimension or quality of life or eq-5d or eq5d or eq 5d or euroqol).ti,ab. (131,090)
41. (qualy or qaly or qualys or qalys or quality adjusted life year$).ti,ab. (5406)
42. life year$ gain$.ti,ab. (1495)
43. willingness to pay.ti,ab. (1724)
44. (hye or hyes or health$ year$ equivalent$).ti,ab. (40)
45. (person trade off$ or person tradeoff$ or time tradeoff$ or time trade off$).ti,ab. (785)
46. theory utilit$.ti,ab. (7)
47. life table$.ti,ab. (3428)
48. health state$.ti,ab. (2931)
49. (sf36 or sf 36).ti,ab. (12,136)
50. (short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).ti,ab. (5063)
51. (6d or 6-d or 6 dimension).ti,ab. (3389)
52. or/23-51 (364,882)
53. 22 and 52 (4601)
54. limit 53 to yr="2005 - 2011" (2906)

A total of 2906 results saved to EndNote library reflux.enl (marked EMBASE in the Custom 4 field).

Results of literature search

EndNote library records were deduplicated as far as possible.

Source	Results	Results after deduplication
NHS EED	85	85
CDSR	44	44
DARE	59	56
HTA	13	12
MEDLINE	1726	1640
EMBASE	2906	1825
Total	4833	3662

Appendix 7 Within-trial cost-effectiveness analysis: health-related quality-of-life and cost-effectiveness results

TABLE 53 Within-trial cost-effectiveness analysis: health-related quality-of-life and cost-effectiveness results

Study	Grant *et al*. 2008[1]	Goeree *et al*. 2011[46]
Trial	REFLUX (multicentre UK)	Anvari (single centre in Canada)
Follow-up	Within-trial cost-effectiveness analysis over 1 year	Within-trial cost-effectiveness analysis over 3 years
Number of patients	318[a]	104
Perspective	UK NHS	Societal perspective
Price year	2006 UK pounds	2009 Canadian dollars (2010 tested in sensitivity analysis)
HRQoL instrument	EQ-5D	HUI (primary instrument); SF-6D and EQ-5D (tested in sensitivity analysis) QoL improved over time across all utility instruments; however, the QALYs gained estimated with EQ-5D were less than half of those estimated with HUI3 and SF-6D
Difference in mean QALYs	0.066 (95% CI 0.026 to 0.107)	0.109 (SD 0.784)
Difference in mean costs	£1280 (£1054 to £1468)	C$3205 (SD C$16,828)
ICER	£19,000 per QALY gained	C$29,400 per QALY gained (utilities from HUI3); C$76,310 per QALY gained (utilities from EQ-5D)
Probability of surgery being cost-effective	When k = £20,000, probability = 46%; when k = £30,000, probability = 86%	Laparoscopic Nissen fundoplication has the highest probability of being the most cost-effective treatment when k is >C$30,000

a The REFLUX economic analysis included both ITT and PP analysis. Results presented in this table are based on the ITT analysis.

Appendix 8 Validation of the multiple imputation

TABLE 54 Predictors of missingness at the 95% confidence level

	Predictors of missingness ($p < 0.05$)		
Follow-up	Variable	Coefficient	Pseudo-R^2
Year 1	EQ-5D at baseline	2.2842	0.0673
	EQ-5D at 3 months	–3.7987	
Year 2	EQ-5D at baseline	1.4209	0.0230
Year 3	EQ-5D at baseline	–3.4594	0.1681
	EQ-5D at 3 months	2.7446	
	EQ-5D at year 2	2.0889	
Year 4	–[a]	–[a]	0.0288
Year 5	EQ-5D at baseline	–7.4267	0.1358
	EQ-5D at year 3	3.1675	

a For year 4, no coefficient was statistically significant at the 95% confidence level. Pseudo-R^2 obtained with constant only.

Note: Only coefficients for the variables significant at the 95% confidence level are shown, despite all models tested including a similar set of variables: demographics (age, sex, BMI), ITT allocation, PP status, costs for the previous years and EQ-5D scores for the previous follow-up points.

The existence of predictors for missingness at the 95% confidence level indicates that data may not be MCAR and therefore that the multiple imputed data set is more reliable than the complete case.

Figures 24 and *25* compare the distribution of total costs and total QALYs, respectively, across the first 10 imputed data sets and the original data (imputation number 0). The distribution is similar, providing some assurance that the multiple imputation strategy was successful.

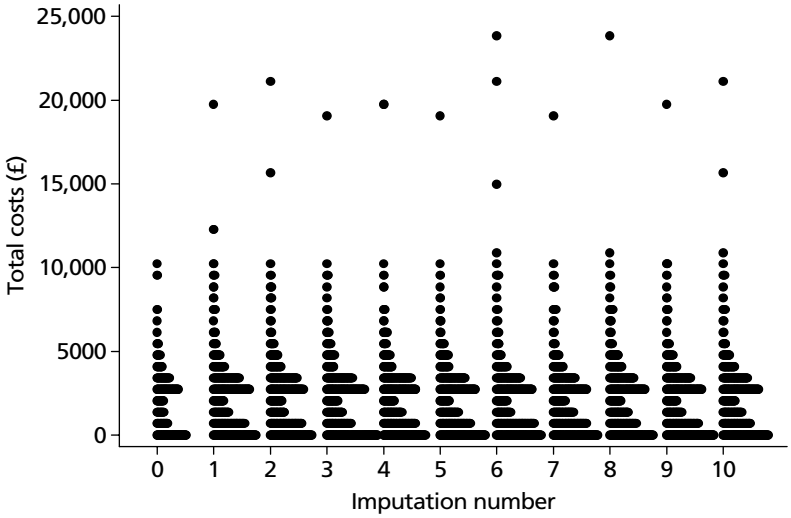


FIGURE 24 Distribution of total costs across the first 10 imputed data sets and for the original data set (imputation number 0).

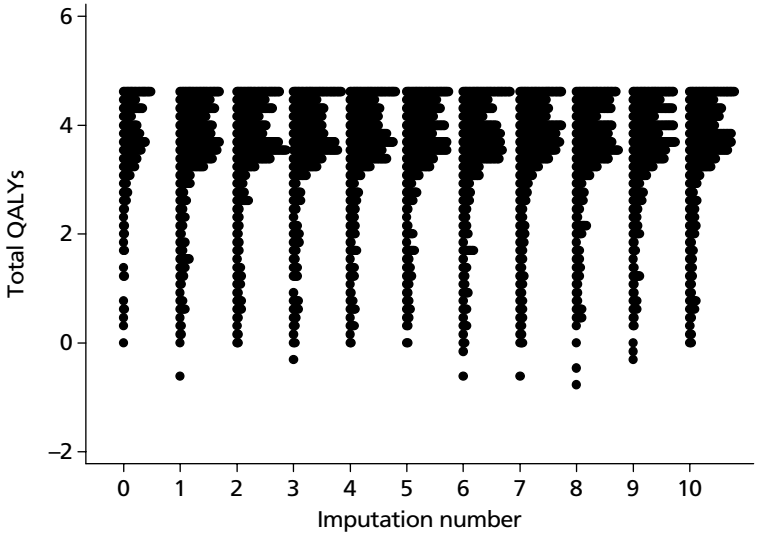


FIGURE 25 Distribution of total QALYs across the first 10 imputed data sets and for the original data set (imputation number 0).

Appendix 9 Costs and health-related quality of life for allocation according to per protocol at 1 year: structural sensitivity analysis

TABLE 55 Total mean costs and total mean QALYs for the medical management and surgery groups according to ITT and PP at 1 year for the complete case

		PP at 1 year		
	Treatment allocated (ITT)	Medical management	Surgery	Total
Total costs	Medical management	£1201.61	£3718.18	£1316.00
Total QALYs		3.5665	2.6076	3.5229
Number of patients		84	4	88
Total costs	Surgery	£989.06	£3525.38	£2981.89
Total QALYs		3.7016	3.7447	3.7354
Number of patients		18	66	84
Total costs	Total	£1164.10	£3536.40	£2129.57
Total QALYs		3.5904	3.6797	3.6268
Number of patients		102	70	172

Note: cells highlighted by shading refer to PP at 1 year groups considered for the incremental analysis.

TABLE 56 Health-related quality of life (EQ-5D) according to PP analysis

Completed questionnaires returned at each time point			Mean (SD) EQ-5D		
Surgery ($n = 178$[a])	Medical management ($n = 179$[a])	Follow-up	Surgery	Medical management	Difference in mean EQ-5D (surgery – medical management) (95% CI)[b,c]
108	162	Baseline	0.7184 (0.2394)	0.7266 (0.2553)	–0.0082 (–0.0691 to 0.0528)
108	143	3 months	0.8059 (0.2393)	0.6910 (0.3068)	0.1148 (0.0446 to 0.1851)
102	153	Year 1	0.7773 (0.2323)	0.7064 (0.2703)	0.0709 (0.0065 to 0.1353)
83	129	Year 2	0.7903 (0.2442)	0.7170 (0.3133)	0.0733 (–0.0068 to 0.1532)
89	127	Year 3	0.7897 (0.2521)	0.7563 (0.2492)	0.0336 (–0.0356 to 0.1018)
88	122	Year 4	0.7785 (0.2636)	0.7550 (0.2678)	0.0236 (–0.0498 to 0.0969)
87	110	Year 5	0.7771 (0.2812)	0.7654 (0.2782)	0.0117 (–0.0674 to 0.0908)

a *n* refers to number of patients originally randomised to each trial arm.
b CIs estimated using OLS regression.
c Unadjusted for baseline EQ-5D.

TABLE 57 Costs associated with resource use for PP analysis

Returned questionnaires in each year			Mean (SD) resource-use cost (£) according to PP at 1 year		Incremental mean cost (surgery–medical management) (95% CI[a]) (£)
Surgery	Medical management	Year	Surgery	Medical management	
104	154	1[b]	3241.78 (1263.80)	361.28 (668.12)	2880.50 (2634.08 to 3126.91)
86	133	2	82.57 (373.46)	159.80 (366.76)	–77.23 (–177.98 to 23.51)
92	128	3	79.80 (349.15)	289.27 (913.00)	–209.47 (–406.78 to –12.15)
88	124	4	96.49 (368.53)	314.70 (1362.96)	–218.21 (–512.14 to 75.72)
90	112	5	43.17 (133.98)	233.60 (645.33)	–190.43 (–326.93 to –53.93)
Cost category					
Surgery in year 1			2780.73 (1756.83)	0[c]	2780.73 (2704.69 to 2856.76)
Reflux-related hospital night admissions			403.07 (1305.30)	315.49 (833.69)	87.58 (–242.21 to 417.37)
Reflux-related hospital day admissions			159.78 (423.32)	231.64 (608.67)	–71.86 (–236.56 to 92.84)
Reflux-related GP visits			130.93 (940.68)	193.31 (465.70)	–62.37 (–176.00 to 51.26)
Medication			87.69 (217.69)	383.01 (525.63)	–295.32 (–424.17 to –166.47)

a CIs estimated using OLS regression.
b Refers to the patients who returned the final year 1 questionnaire.
c By definition, none of the medical management patients in PP underwent surgery.

Appendix 10 Protocol

THE PLACE OF MINIMAL ACCESS SURGERY AMONGST PEOPLE WITH GASTRO-OESOPHAGEAL REFLUX DISEASE (GORD)

A UK COLLABORATIVE STUDY FUNDED BY THE NHS R&D HTA PROGRAMME

PROTOCOL

VERSION 6a - November 2003

THE PLACE OF MINIMAL ACCESS SURGERY AMONGST PEOPLE WITH GASTRO-OESOPHAGEAL REFLUX DISEASE (GORD) A UK COLLABORATIVE STUDY

(Known as the REFLUX Trial)

PROTOCOL SUMMARY

AIM	To identify the optimal place within the NHS for minimal access surgery amongst people with GORD, whose symptoms would otherwise be managed with long-term medical therapy.
DESIGN	Multicentre, pragmatic randomised trial (with parallel non-randomised preference groups).
PATIENT ELIGIBILTY	• Documented evidence of GORD (endoscopy and/or manometry/24h pH monitoring) • Symptoms for more than 12 months and currently requiring maintenance proton pump inhibitor (PPI) therapy for reasonable symptom control • Received care from a participating clinician • Suitable for either policy (ASA grade I or II) • Recruiting doctor uncertain which management policy is better • Give informed consent to either random allocation of management or follow-up after preferred management
RECRUITMENT	Based on surgeon-physician 'partnership' in at least 15 centres.
INTERVENTIONS	A laparoscopic surgery based policy compared with a continued medical management policy.
OUTCOME MEASUREMENT	**Primary –** Disease specific outcome and NHS costs **Secondary –** Patient costs and Health-related quality of life (EQ5D, SF36)
ORGANISATION	• All whole-hearted contributors part of the GORD Trialist Group (with group authorship of main reports) • Conduct overseen by Steering Group • Trial Office in Aberdeen responsible for day-to-day non-clinical co-ordination • Sessional research nurses in each clinical centre • Health economic evaluation and outcome measure assessment jointly led from York and Aberdeen
FUNDING	NHS R&D Health Technology Assessment Programme

TABLE OF CONTENTS

See Grant *et al.*[1] for details of appendices.

1. OUTLINE OF THE TRIAL

Aim

The aim is to identify the optimal place within the NHS of minimal access surgery amongst people with gastro-oesophageal reflux disease (GORD). Its focus is people whose symptoms would otherwise be managed with long-term medical therapy. The background and justification are summarised in Appendix I.

Objectives

- To evaluate the clinical effectiveness, cost-effectiveness, and safety of a policy of relatively early laparoscopic surgery compared with continued medical management amongst people with GORD judged suitable for both policies.
- To explore factors which may influence the relative performance of the two policies, such as patient preference, surgeon experience, pre-enrolment symptoms and signs, underlying pathology, type of operative procedure used or choice of therapy, and time since surgery.
- To explore the impact that various policies for using laparoscopic surgery would have on the NHS and society in respect of the costs or savings that they would imply for (a) those providing surgical care (in secondary care settings), (b) those providing long-term medical management (usually in primary care settings), and (c) those with GORD.

Design

The study will have two complementary components:

A A randomised trial (with parallel non-randomised preference groups) comparing a laparoscopic surgery based policy with a continued medical management policy to assess their relative clinical effectiveness.

B An economic evaluation of laparoscopic surgery for GORD to compare the cost-effectiveness of the two management policies, to identify the most efficient provision of future care, and to describe the resource impact that various policies for fundoplication would have on the NHS.

The rationale for the study design is described in Appendix II.

2. THE RANDOMISED TRIAL (WITH PARALLEL PREFERENCE GROUPS)

Centre eligibility

Clinical centres will be based on local partnerships between surgeons with experience of laparoscopic fundoplication and the gastroenterologists, with whom they share the secondary care of patients with GORD. Centres will be eligible if they include:

1. a surgeon who has performed at least 50 laparoscopic fundoplication operations
2. one or more gastroenterologists who agree to collaborate with the surgeon in the trial.

Patient eligibility

Inclusion criteria

1. Documented evidence of GORD (based on endoscopy and/or manometry/24hr pH monitoring)
2. Symptoms for more than 12 months and currently requiring maintenance proton pump inhibitor (PPI) therapy for reasonable symptom control (Patients who are intolerant to PPIs and therefore require Histamine Receptor Antagonists (H_2RAs) therapy to control their symptoms will also be included)
3. Care provided by a participating clinician
4. Suitable for either policy (including ASA grade I or II)
5. Recruiting doctor uncertain which management policy is better
6. Informed consent either to random allocation of management or to follow-up after preferred management

Exclusion criteria

1. Morbidly obese (BMI >40 kg/m^2)
2. Barrett's oesophagus of more than 3 cm or have evidence of dysplasia
3. Paraoesophageal hernia
4. Oesophageal stricture

Although there is no formal age limit, it will be younger patients with GORD who will be eligible, who are expected to be aged between 18 and 65 years .

Health technology policies being compared

Laparoscopic surgery policy:

Most of those allocated to this policy will have surgery. Deferring or declining will remain an option, however, even after trial entry, particularly amongst those recruited by a gastroenterologist and referred to a surgeon for consideration of surgery within the trial. Participants who have not had manometry/pH studies will undergo these tests before surgery to exclude achalasia.

The surgery will be performed either by a surgeon who has undertaken more than 50 laparoscopic fundoplications or by a less experienced surgeon working under the supervision of an experienced surgeon. It is recommended that crural repair be routine and non-absorbable, synthetic sutures (not silk) be used for the repair. The type of fundoplication used will be left to the discretion of an experienced surgeon. For the purposes of the main comparisons, the different surgical techniques for laparoscopic fundoplication will be considered as parts of a single policy. The study design will, however, allow indirect comparisons between techniques.

It is expected that local policies for thromboembolism prophylaxis will include a suitable anticoagulant (such as heparin or tinzaparin) plus surgical stockings or pneumatic compression.

Medical therapy policy:

Most of those allocated to the medical therapy policy will continue 'best medical management' (appropriate PPI), as recommended by the clinician responsible for care. Management should conform to the principles of the Genval Workshop Report (see Appendix III). While all the recommendations of this workshop cannot be summarised here, they include stepping down antisecretory medication in most patients to the lowest dose that maintains acceptable symptom control. Patients who have had severe oesophagitis should not be managed on the basis of symptoms alone, however. While it is expected that most trial participants allocated medical management will continue to be managed in this way, surgery should be considered if a clear indication for it subsequently develops.

Outcome measurements

Primary:

a 'Disease-specific' outcome to include the need for changes in treatment, reflux and other gastro-intestinal symptoms, and the side effects and complications of both therapies.

b NHS costs including treatments, investigations, consultations and other contacts with the health service.

Secondary:

c Health-related quality of life – EQ5D and SF36.

d Patient costs including loss of earnings, reduction in activities, and the cost of prescriptions and travel to health care.

Other:

e Other serious morbidity, such as operative complications

f Mortality

The instrument for collecting this information are in Appendix IV. The ways in which these data will be displayed in the final report are illustrated in Appendix V.

Sample size and statistical analysis

A sample size of 600 will identify a difference between the two randomised groups of less than 0.25 of the standard deviation of the disease-specific instrument, EQ5D or SF36 with 80% power using a significance level of 5%. Based on the same arguments, about 300 people will be recruited to each arm of the preference study.

The cost savings of a surgical policy will largely depend on the number of patients managed surgically who no longer require PPI treatment. A trial with 300 surgically managed patients will estimate this proportion to within about 5% with 95% statistical confidence.

A single principal analysis is planned within the current time frame when all participants have been followed-up for at least 12 months after surgery (or an equivalent time if managed medically). Standard statistical techniques will be used with analysis by intention to treat and 95% confidence intervals. Secondary analyses will explore differential effects within pre-stated sub-groups, characterised by initial symptom severity and surgeon's preferred operative procedure; 99% confidence intervals will be

generated for such analyses to reflect their exploratory nature. The issue of continued surgeon 'learning' will also be investigated using curve fitting techniques.

3. THE ECONOMIC EVALUATION

The economic evaluation is described in detail in Appendix VI. It will have three components: a within-trial cost-effectiveness study; a detailed assessment of the preferences of patients with GORD; and an outside-trial cost-effectiveness analysis based on decision modelling.

4. PRACTICAL ARRANGEMENTS

Each clinical centre will be supported by a part-time research nurse.

Identification of potential participants

Potential participants will be identified in three ways:

- Retrospective case-note review
- Prospective identification of current case
- Referral from general practice

These are summarised in Figure 1. The actual approach used will vary between centres, but case note review is likely to be the principal method.

As a general rule, potentially eligible participants will be booked for an outpatient appointment. They will be sent a brief letter, together with a copy of the information leaflets in advance, letting them know that the trial is likely to be discussed with them (Appendix VII). At the appointment, the clinician will review the person's symptoms and current treatment regimen, and assess eligibility for the trial following the completion of a Patient Assessment Form (Appendix VIII). If eligibility is confirmed, the person will be invited to see the research nurse who will describe the study and discuss any issues that arise. This is summarised in Figure 2. The nurse will also give a supplementary information leaflet that describes the operation in more detail (Appendix IX). Information will also be sent to the general practitioner (GP) in case the participant consults them to discuss the trial (Appendix X); a specific clinic letter will follow from the consultant.

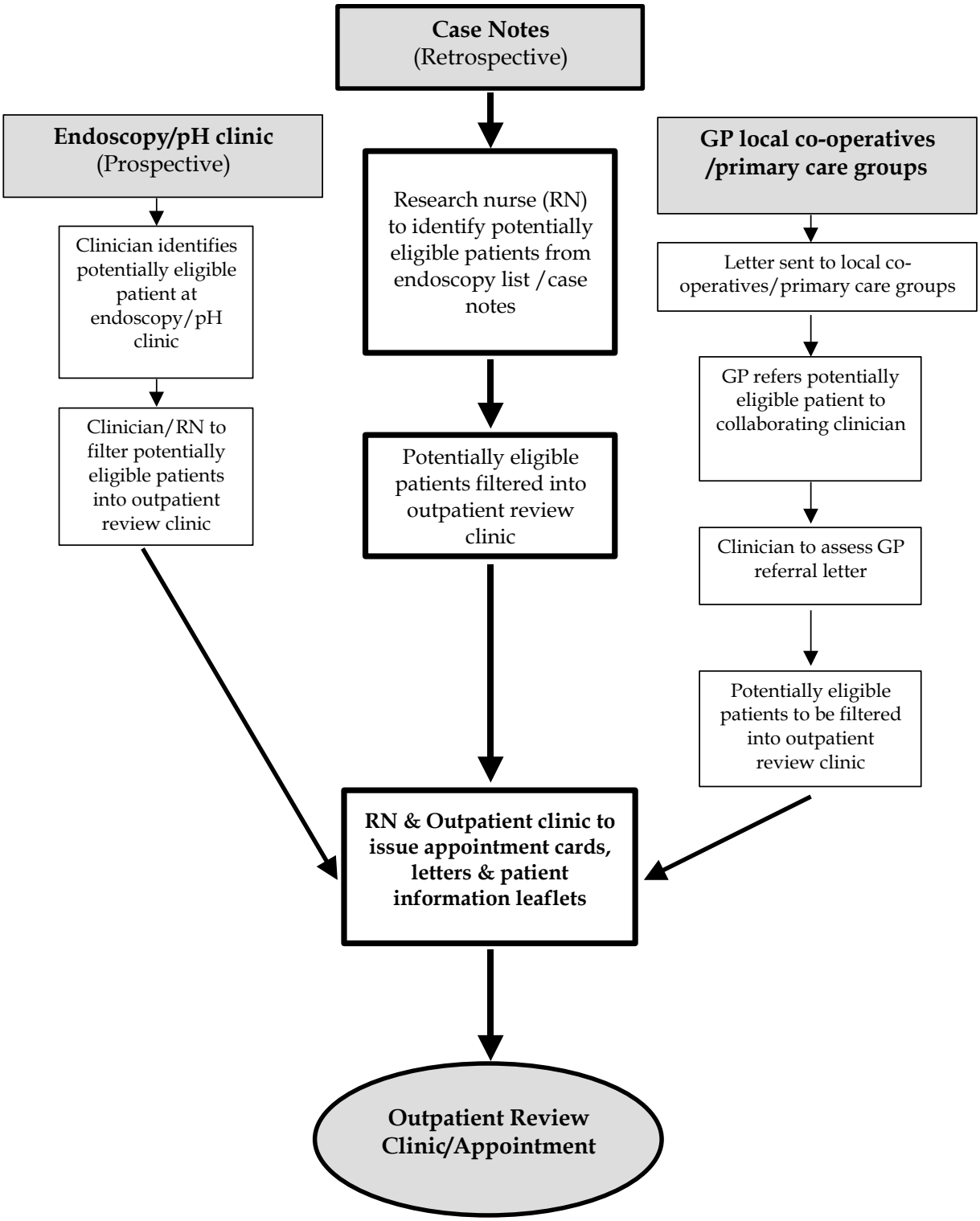


Figure 1. Flowchart describing sources for patient identification

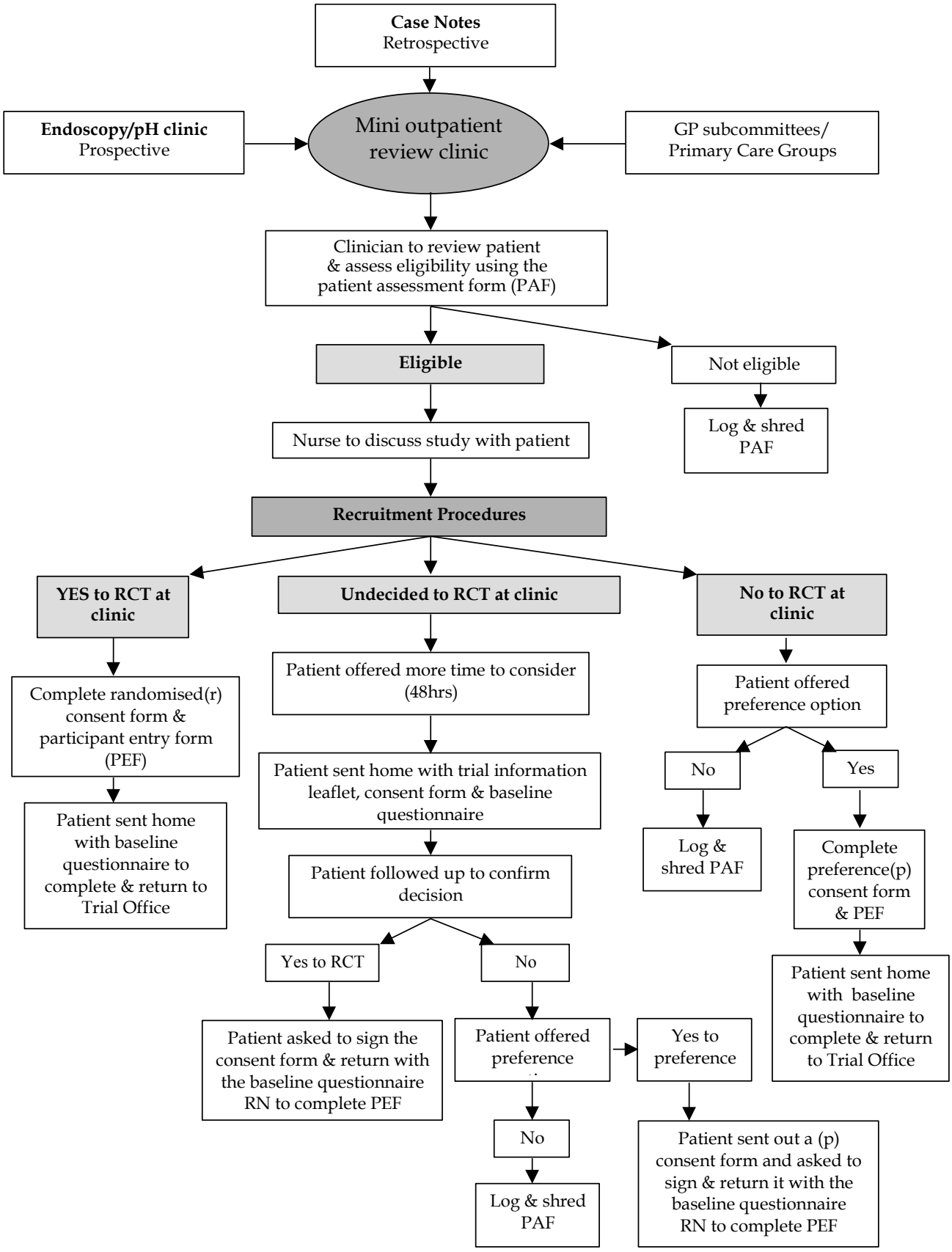


Figure 2. Flowchart describing patient recruitment

Consent to participate

The randomised trial:

Some potential participants will make a decision about participation at this appointment. Those who wish to participate in the randomised trial will be asked to sign a consent form (Appendix XI). On this, they will confirm that they have been given the information they require and that the study has been explained to them. They will also confirm that they understand that they will be sent questionnaires from the Trial Office at participant-specific time intervals after joining the study. (This will be at a time equivalent to around three months and 12 months after surgery.) They will also be told that it is anticipated that further follow-up will be performed periodically thereafter for some years.

The preference study:

A person who does not want to take part in the randomised trial because of a strong preference for one type of treatment management will be asked to take part in the preference arm of the study. Those who wish to participate in the preference study will be given a preference information leaflet and asked to sign a consent form (Appendix XII). In addition to the details collected on the randomised consent form, they will confirm their preferred treatment allocation.

Any person who is uncertain will be given at least 48 hours to consider participation. A research nurse will then phone them to find out their decision and make arrangements as appropriate for them to sign a randomised trial or a preference study consent form.

One copy of the consent form will be given to the participant, another will be filed in the patient's hospital case notes, and the third will be posted to the Trial Office.

Information to be collected at trial entry

Once a participant has agreed to join the trial, the research nurse will record basic identifying and descriptive information on a standard form (Appendix XIII). This information will be sent to the Trial Office.

The participant will take home a baseline questionnaire to complete, and will be asked to return it in a pre-paid envelope to the Trial Office.

Study registration (and treatment allocation when randomised)

The entry procedure will distinguish between those who have agreed to randomisation and those who have agreed to participate in the preference part of the study.

The treatment allocation for participants consenting to the randomised arm of the trial will be computer-generated in the Trial Office. The allocation will be stratified by centre, with balance in respect of other key prognostic variables – age (18-50 y or 51-65 y), sex (M or F), and BMI (≤28 or >29 kg/m^2) - by a process of minimisation.

A letter will be sent from the Trial Office to each participant (Appendix XIV), their GP (Appendix XV) and the local research nurse, confirming the treatment allocation and whether they are taking part in the randomised- or preference-arm of the trial. A letter will also be sent to the respective collaborating surgeon or gastroenterologist with respect to the treatment the participant is allocated.

Clinical management

Clinical management will be left to the discretion of the clinician responsible for care. A summary of the different clinical management pathways is illustrated in Figure 3.

Participants who are allocated to the surgical arm, will be invited to a consultation with the collaborating surgeon. (Participants who have not already had manometry/pH studies will be booked to undergo these tests prior to this consultation.) During this consultation, the surgeon will confirm that there is no contra-indication to surgery and discuss the operation in more detail with the participant, before arranging a date for the operation. The intra-operative details will be recorded by the surgeon on specially designed study forms (Appendix XVI).

All other in-hospital data collection will be the responsibility of the local study nurse.

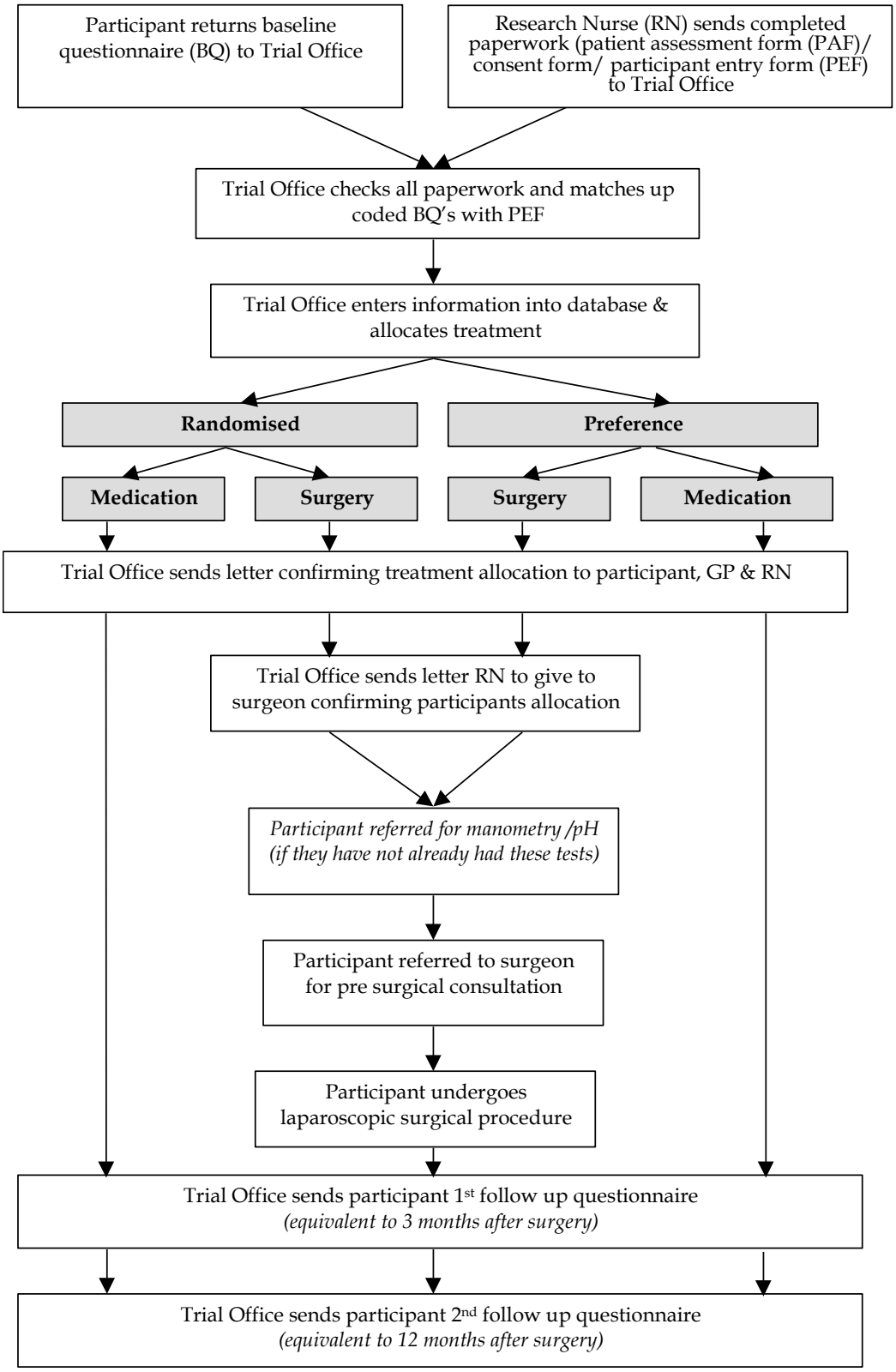


Figure 3. Flowchart showing clinical management post recruitment

Follow-up in the trial

Follow-up by postal questionnaire will be performed twice, at participant-specific time intervals after joining the study. (This will be at a time equivalent to around three and 12 months after surgery). When necessary, clarification of clinical management will be sought through the research nurses (while they are in post) and then subsequently through the recruiting doctor or general practitioner. While it is anticipated that further follow-up will be performed periodically thereafter for some years (dependent on funding being available at that time) these subsequent assessments are not part of this protocol.

Data collection after trial entry

All data will be sent to the Trial Office in Aberdeen for processing. Staff in Aberdeen will work closely with the research nurses to secure as complete and accurate data as possible. A random 10% sample of data will be double-entered to check accuracy. Extensive range and consistency checks will further enhance the quality of the data.

Organisation

Local organisation

The trial is designed to limit the extra work for collaborating clinicians to tasks that only they can do. Research nurses will facilitate the trial locally, and the central organisation will take responsibility for data management and participant follow-up.

Clinical collaborators (gastroenterologist and/or surgeon) will:

1. establish the trial locally (e.g. identifying a 'partnering' clinician or surgeon if not already agreed; facilitating local research ethics committee approval; identifying and appointing a local research nurse; and ensuring that all clinical staff involved in the care of patients with GORD are informed about the trial)
2. take responsibility for clinical aspects of the trial locally (e.g. if any particular concerns emerge)
3. notify the Trial Office of any unexpected clinical events that might be related to trial participation
4. provide support and supervision for all aspects of the work of the local research nurse
5. represent the centre at REFLUX trial collaborators' meetings

Research nurses will:

1. keep local staff informed about the trial and its progress
2. keep regular contact with the local gastroenterologist(s) and surgeon
3. maintain regular contact with the Trial Office
4. identify potential participants and log whether or not they are recruited to the trial (including the preference groups) - with reasons for non-participation
5. arrange for the initial letter of invitation and information leaflet to be sent to potential participants prior to an out-patient assessment.
6. assist the participating clinicians (e.g. at assessment clinics) to give additional information and seek consent to study entry
7. ensure that the baseline data describing participants are collected and sent back to the Trial Office
8. facilitate later follow-up by, for example, helping with local tracing
9. provide support for participants in other ways if there are difficulties
10. represent the centre at trial nurse meetings and collaborators' meetings

5. TRIAL CO-ORDINATION

Trial Offices

The main Trial Office is within the Health Services Research Unit in Aberdeen and gives day-to-day support to the clinical centres. This Office is responsible for all central co-ordination of the trial, including centre and research nurse support, study entry and randomisation, postal follow-up, data processing and statistical analysis.

The economic evaluation and the outcome development work is based in the Centre for Health Economics and the Department of Health Sciences and Clinical Evaluation, respectively, both within the University of York.

The Steering Group

The trial is co-ordinated by a Steering Group (listed in Appendix XVII). The Steering Group, in consultation with the Collaborative Group (see below), will take responsibility for any major decisions, such as the need to close recruitment early to one or more parts of the study or to change the protocol for any reason.

The Collaborative Group

The Collaborative Group is made up of the surgeons, gastroenterologists and research nurses contributing to the trial, members of the Steering Group, and representatives from the Trial Offices.

The Data Monitoring Committee

A data monitoring committee will be established. It will be independent of the trial organisers. During the period of recruitment to the trial, interim analyses will be supplied, in strict confidence, to the data monitoring committee, together with any other analyses that the committee may request. This may include analyses of data from other comparable trials. In the light of these interim analyses, the data monitoring committee will advise the Steering Group if, in its view, the trial has provided both (a) proof beyond reasonable doubt[1] that for all or some types of patients one intervention is clearly indicated in terms of clinical- and cost-effectiveness, and (b) evidence that might reasonably be expected to influence materially the care of people with GORD by clinicians who know the results of this and comparable trials. The Steering Group can then decide to consult the Collaborative Group about whether or not to modify intake into the trial or to report results early. Unless this happens, however, the Steering Group, the Collaborative Group and Trial Offices (except those who supply the confidential analyses) will remain ignorant of the interim results considered by the committee.

The frequency of interim analyses will depend on the judgement of the chairman of the committee, in consultation with the Steering Group.

6. FINANCE

The trial is supported by a grant from the Health Technology Assessment Programme of the NHS Executive Research and Development Programme.

Note:

[1] Appropriate criteria for proof beyond reasonable doubt cannot be specified precisely. A difference of at least three standard deviations in the interim analysis of a major endpoint may be needed to justify halting, or modifying, such a study prematurely. If this criteria were to be adopted, it would have the practical advantage that the exact number of interim analyses would be of little importance, and so no fixed schedule is proposed (Peto R et al *Br J Cancer* 1976; 34: 584-612).

7. A STUDY OF FACTORS IMPACTING ON PATIENTS DECISION TO PARTICIPATE IN THE REFLUX TRIAL (APPENDIX XVIII)

During the recruitment phase of the trial, it is anticipated that a CSO research fellow will undertake supplementary site visits to explore the patients' perspective in relation to trial recruitment. A small number of centres will be purposively selected using qualitative methods (non-participation observation and in-depth interviews). It is proposed that the selected centres will reflect varying recruitment rates.

It is expected that, subject to clinician and patient consent, the research fellow would sit-in and observe reflux clinics where patients are approached to join the study. The researcher would aim to supplement the observational work by interviewing some of the patients (again, subject to consent) about their experience of trial recruitment and factors impacting on their decision to join the trial or not.

It is hoped this small but very useful complementary study nested in the REFLUX trial, will help identify factors impacting on patient recruitment and enable us to look at ways of addressing these issues to facilitate improved future trial recruitment.

8. PUBLICATION

The success of the trial depends entirely on the whole-hearted collaboration of a large number of people. For this reason, chief credit for the trial will be given, not to the committees or central organisers, but to all those who have whole-heartedly collaborated in the trial. The trial's publication policy is described in detail in Appendix XIX. The results of the trial will be reported first to the trial collaborators. The main report will be drafted by the Steering Group, and circulated to all the clinical collaborators for comment. The final version will be agreed by the Steering Group before submission for publication, on behalf of the collaboration. To safeguard the integrity of the study, reports of sub-studies will not be submitted for publication without prior discussion with the Steering Group. Once the main report has been published, a lay summary will be sent to participants who have indicated that they would like to receive one.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library