

# 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,<sup>1\*</sup> C Boachie,<sup>1</sup> SC Cotton,<sup>1</sup> R Faria,<sup>2</sup>  
L Bojke,<sup>2</sup> DM Epstein,<sup>2</sup> CR Ramsay,<sup>1</sup> B Corbacho,<sup>2</sup>  
M Sculpher,<sup>2</sup> ZH Krukowski,<sup>3</sup> RC Heading<sup>4</sup> and  
MK Campbell<sup>1</sup> on behalf of the REFLUX trial group

<sup>1</sup>Health Services Research Unit, Health Sciences Building, University of Aberdeen, Aberdeen, UK

<sup>2</sup>Centre for Health Economics, University of York, York, UK

<sup>3</sup>Aberdeen Royal Infirmary, Aberdeen, UK

<sup>4</sup>School of Medicine and Health, Wolfson Research Institute, Durham University, Stockton-on-Tees, UK

\*Corresponding author

## Executive summary

### Five-year follow-up of the REFLUX trial

Health Technology Assessment 2013; Vol. 17: No. 22

DOI: 10.3310/hta17220

NIHR Journals Library [www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)

# 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 refute 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.

## Publication

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

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/](http://www.publicationethics.org/)).

Editorial contact: [nihredit@southampton.ac.uk](mailto:nihredit@southampton.ac.uk)

The full HTA archive is freely available to view online at [www.journalslibrary.nihr.ac.uk/hta](http://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](http://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/](http://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.

**© Queen's Printer and Controller of HMSO 2013. This work was produced by Grant et al. under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.**

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

## **Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library**

**Professor Tom Walley** Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

### **NIHR Journals Library Editors**

**Professor Ken Stein** Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

**Professor Andree Le May** Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

**Dr Martin Ashton-Key** Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

**Professor Matthias Beck** Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

**Professor Aileen Clarke** Professor of Health Sciences, Warwick Medical School, University of Warwick, UK

**Dr Tessa Crilly** Director, Crystal Blue Consulting Ltd, UK

**Dr Peter Davidson** Director of NETSCC, HTA, UK

**Dr Tom Marshall** Reader in Primary Care, School of Health and Population Sciences, University of Birmingham, UK

**Professor William McGuire** Professor of Child Health, Hull York Medical School, University of York, UK

**Professor Geoffrey Meads** Honorary Professor, Business School, Winchester University and Medical School, University of Warwick, UK

**Professor John Powell** Consultant Clinical Adviser, NICE, UK

**Professor James Raftery** Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

**Dr Rob Riemsma** Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

**Professor Helen Snooks** Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:  
[www.journalslibrary.nihr.ac.uk/about/editors](http://www.journalslibrary.nihr.ac.uk/about/editors)

**Editorial contact:** [nihredit@southampton.ac.uk](mailto:nihredit@southampton.ac.uk)