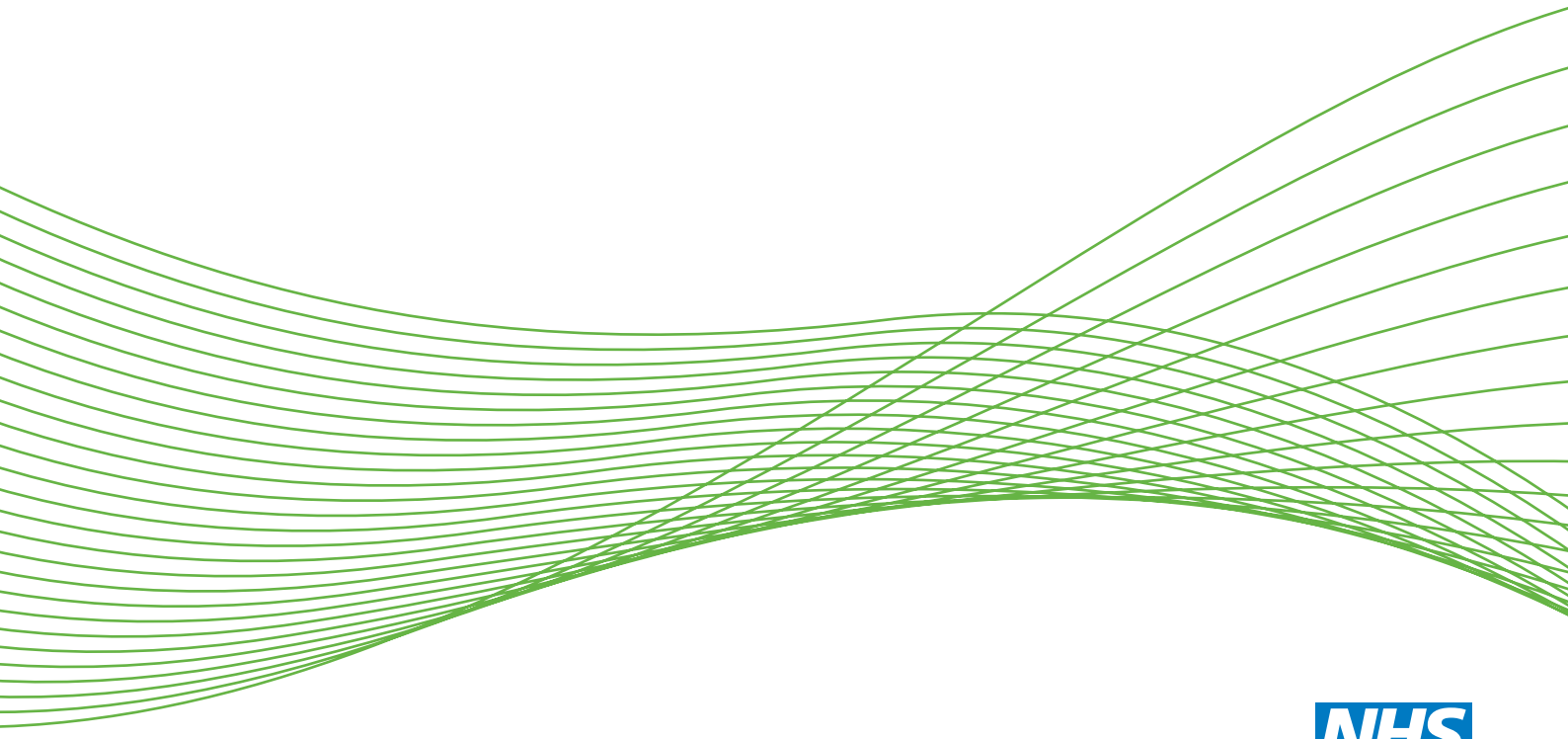
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 22 ISSUE 19 APRIL 2018
ISSN 1366-5278

Amaze: a double-blind, multicentre randomised controlled trial to investigate the clinical effectiveness and cost-effectiveness of adding an ablation device-based maze procedure as an adjunct to routine cardiac surgery for patients with pre-existing atrial fibrillation

Linda Sharples, Colin Everett, Jeshika Singh, Christine Mills, Tom Spyt, Yasir Abu-Omar, Simon Fynn, Benjamin Thorpe, Victoria Stoneman, Hester Goddard, Julia Fox-Rushby and Samer Nashef



NHS
National Institute for Health Research

DOI 10.3310/hta22190

Amaze: a double-blind, multicentre randomised controlled trial to investigate the clinical effectiveness and cost-effectiveness of adding an ablation device-based maze procedure as an adjunct to routine cardiac surgery for patients with pre-existing atrial fibrillation

Linda Sharples,[1]* Colin Everett,[2] Jeshika Singh,[3] Christine Mills,[4] Tom Spyt,[5] Yasir Abu-Omar,[6] Simon Fynn,[6] Benjamin Thorpe,[2] Victoria Stoneman,[4] Hester Goddard,[4] Julia Fox-Rushby[7] and Samer Nashef[6]

[1]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK
[2]Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
[3]Health Economics Research Group (HERG), Brunel University London, London, UK
[4]Papworth Trials Unit Collaboration, Papworth Hospital, Cambridge, UK
[5]Department of Cardiac Surgery, Glenfield Hospital, Leicester, UK
[6]Department of Cardiology and Department of Cardiothoracic Surgery, Papworth Hospital, Cambridge, UK
[7]Department of Population Science, King's College London, London, UK

*Corresponding author

Declared competing interests of authors: Samer Nashef received personal expenses from AtriCure, Inc. (Amsterdam, the Netherlands) for contributing to educational courses for surgeons interested in learning the maze procedure, independently of this study. AtriCure, Inc. is one of several manufacturers of atrial fibrillation ablation devices, the technology of which was used in atrial fibrillation surgery in this study.

Published April 2018
DOI: 10.3310/hta22190

This report should be referenced as follows:

Sharples L, Everett C, Singh J, Mills C, Spyt T, Abu-Omar Y, *et al*. Amaze: a double-blind, multicentre randomised controlled trial to investigate the clinical effectiveness and cost-effectiveness of adding an ablation device-based maze procedure as an adjunct to routine cardiac surgery for patients with pre-existing atrial fibrillation. *Health Technol Assess* 2018;**22**(19).

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

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

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

Editorial contact: journals.library@nihr.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: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 07/01/34. The contractual start date was in September 2008. The draft report began editorial review in January 2017 and was accepted for publication in August 2017. 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 and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

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

Abstract

Amaze: a double-blind, multicentre randomised controlled trial to investigate the clinical effectiveness and cost-effectiveness of adding an ablation device-based maze procedure as an adjunct to routine cardiac surgery for patients with pre-existing atrial fibrillation

Linda Sharples,[1]* Colin Everett,[2] Jeshika Singh,[3] Christine Mills,[4] Tom Spyt,[5] Yasir Abu-Omar,[6] Simon Fynn,[6] Benjamin Thorpe,[2] Victoria Stoneman,[4] Hester Goddard,[4] Julia Fox-Rushby[7] and Samer Nashef[6]

[1]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK
[2]Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
[3]Health Economics Research Group (HERG), Brunel University London, London, UK
[4]Papworth Trials Unit Collaboration, Papworth Hospital, Cambridge, UK
[5]Department of Cardiac Surgery, Glenfield Hospital, Leicester, UK
[6]Department of Cardiology and Department of Cardiothoracic Surgery, Papworth Hospital, Cambridge, UK
[7]Department of Population Science, King's College London, London, UK

*Corresponding author Linda.Sharples@lshtm.ac.uk

Background: Atrial fibrillation (AF) can be treated using a maze procedure during planned cardiac surgery, but the effect on clinical patient outcomes, and the cost-effectiveness compared with surgery alone, are uncertain.

Objectives: To determine whether or not the maze procedure is safe, improves clinical and patient outcomes and is cost-effective for the NHS in patients with AF.

Design: Multicentre, Phase III, pragmatic, double-blind, parallel-arm randomised controlled trial. Patients were randomised on a 1 : 1 basis using random permuted blocks, stratified for surgeon and planned procedure.

Setting: Eleven acute NHS specialist cardiac surgical centres.

Participants: Patients aged ≥ 18 years, scheduled for elective or in-house urgent cardiac surgery, with a documented history (> 3 months) of AF.

Interventions: Routine cardiac surgery with or without an adjunct maze procedure administered by an AF ablation device.

Main outcome measures: The primary outcomes were return to sinus rhythm (SR) at 12 months and quality-adjusted life-years (QALYs) over 2 years after randomisation. Secondary outcomes included return to SR at 2 years, overall and stroke-free survival, drug use, quality of life (QoL), cost-effectiveness and safety.

Results: Between 25 February 2009 and 6 March 2014, 352 patients were randomised to the control ($n = 176$) or experimental ($n = 176$) arms. The odds ratio (OR) for return to SR at 12 months was 2.06 [95% confidence interval (CI) 1.20 to 3.54; $p = 0.0091$]. The mean difference (95% CI) in QALYs at 2 years between the two trial arms (maze/control) was –0.025 (95% CI 0.129 to 0.078; $p = 0.6319$). The OR for SR at 2 years was 3.24 (95% CI 1.76 to 5.96). The number of patients requiring anticoagulant drug use was significantly lower in the maze arm from 6 months after the procedure. There were no significant differences between the two arms in operative or overall survival, stroke-free survival, need for cardioversion or permanent pacemaker implants, New York Heart Association Functional Classification (for heart failure), EuroQol-5 Dimensions, three-level version score and Short Form questionnaire-36 items score at any time point. Sixty per cent of patients in each trial arm had a serious adverse event ($p = 1.000$); most events were mild, but 71 patients (42.5%) in the maze arm and 84 patients (45.5%) in the control arm had moderately severe events; 31 patients (18.6%) in the maze arm and 38 patients (20.5%) in the control arm had severe events. The mean additional cost of the maze procedure was £3533 (95% CI £1321 to £5746); the mean difference in QALYs was –0.022 (95% CI –0.1231 to 0.0791). The maze procedure was not cost-effective at £30,000 per QALY over 2 years in any analysis. In a small substudy, the active left atrial ejection fraction was smaller than that of the control patients (mean difference of –8.03, 95% CI –12.43 to –3.62), but within the predefined clinically equivalent range.

Limitations: Low recruitment, early release of trial summaries and intermittent resource-use collection may have introduced bias and imprecise estimates.

Conclusions: Ablation can be practised safely in routine NHS cardiac surgical settings and increases return to SR rates, but not survival or QoL up to 2 years after surgery. Lower anticoagulant drug use and recovery of left atrial function support anticoagulant drug withdrawal provided that good atrial function is confirmed.

Further work: Continued follow-up and long-term clinical effectiveness and cost-effectiveness analysis. Comparison of ablation methods.

Trial registration: Current Controlled Trials ISRCTN82731440.

Funding: This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 22, No. 19. See the NIHR Journals Library website for further project information.

Contents

List of tables

List of figures

List of abbreviations

A&E	accident and emergency
AE	adverse event
AF	atrial fibrillation
ALAEF	active left atrial ejection fraction
AVR	aortic valve replacement or repair
CABG	coronary artery bypass graft operation
CI	confidence interval
CONSORT	Consolidated Standards of Reporting Trials
COPD	chronic obstructive pulmonary disease
CRF	clinical report form
DMEC	Data Monitoring and Ethics Committee
ECG	electrocardiography
EQ-5D-3L	EuroQol-5 Dimensions, three-level version
EuroSCORE	European System for Cardiac Operative Risk Evaluation
GP	general practitioner
HESTER	Has Electrical Sinus Translated into Effective Remodelling?
HR	hazard ratio
HRQoL	health-related quality of life
HTA	Health Technology Assessment
ICC	intraclass correlation coefficient
ICER	incremental cost-effectiveness ratio
ICU	intensive care unit
INMB	incremental net monetary benefit
ITT	intention to treat
LAEF	left atrial ejection fraction
LAV_{max}	maximum left atrial volume
LAV_{min}	minimum left atrial volume
LAV_{preA}	pre A-wave left atrial volume
LVEF	left ventricular ejection fraction
MCS	mental component score
MI	myocardial infarction
MRI	magnetic resonance imaging
MVR	mitral valve replacement or repair
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NYHA	New York Heart Association
OR	odds ratio
PCS	physical component score
QALY	quality-adjusted life-year
QoL	quality of life
RCT	randomised controlled trial
RR	relative risk
SAE	serious adverse event
SD	standard deviation
SF-6D	Short Form questionnaire-6 Dimensions
SF-36	Short Form questionnaire-36 items
SR	sinus rhythm
TSC	Trial Steering Committee
WTP	willingness to pay

Plain English summary

Irregular heartbeat is common, and the most common type is atrial fibrillation (AF). The heart has four chambers: two ventricles that propel blood from the heart and two atria that receive blood. In AF, the chambers of the heart lose their pumping action and the heart becomes less efficient. Blood can settle in the atria and form clots. These can detach and cause strokes, so AF patients are given blood-thinning drugs to reduce this risk. These drugs can increase the risk of bleeding. The maze procedure is an operation designed to stop AF and make the heart beat regularly again.

Many patients who need major heart surgery also have AF. The Amaze trial was designed to find out whether or not adding a maze procedure to heart operations is useful in making the heartbeat regular again, if it improves long-term survival and quality of life (QoL) and whether or not any benefits are worth the extra costs.

Between 25 February 2009 and 6 March 2014, 352 patients in 11 hospitals were recruited and randomly put into one of two treatment groups: (1) maze and planned heart surgery or (2) planned surgery alone. The results showed that patients in the maze group were more likely to have a normal, regular heartbeat afterwards; however, there were no differences in survival or QoL at 2 years. In addition, many maze patients who recovered a regular heartbeat did well in terms of the heart's pumping action, suggesting that longer follow-up may show better QoL and survival in these patients.

Scientific summary

Background

Atrial fibrillation (AF) is characterised by an irregular heartbeat resulting from abnormal electrical signals in the atria. Prevalence is 1–2% of the population in high-income countries, and this increases with age and comorbidities such as obesity, diabetes and hypertension. The UK prevalence is 7.2% in patients aged ≥ 65 years and 10.3% in patients aged ≥ 75 years. With the advancing age of the population and the increasing prevalence of obesity, this is likely to increase.

Atrial fibrillation causes palpitations, chest pain, dizziness and breathlessness, and imposes a heavy burden on both patient health-related quality of life (HRQoL) and NHS resources. Inefficient heart pumping as a result of AF increases the risk of blood clot formation, which can lead to stroke; anticoagulant medication reduces the risk of stroke, but confers an increased risk of bleeding. AF may also exacerbate existing heart failure or cause heart failure; treatment of AF and its consequences is expensive for the NHS.

The maze procedure, developed in the 1980s, involves multiple cutting and sewing of the atria and pulmonary veins to prevent AF. Despite success in restoring sinus rhythm (SR), the technical challenges required for this procedure mean that it is reserved for severely symptomatic patients. Less demanding methods of achieving the electrical block, using a range of energy sources (heat, cold, radiofrequency or microwave) to ablate atrial tissue, have been developed. Although technically easier, quicker and safer, these methods are a new and costly technology.

There is evidence that AF ablation increases rates of freedom from AF, atrial flutter and atrial tachycardia and decreases antiarrhythmic medication use 3 months after surgery. However, effects on cardiovascular mortality, adverse events (AEs), HRQoL and long-term outcomes are uncertain. Results of cost-effectiveness analyses are mixed and limited by the lack of evidence on HRQoL and other key outcomes in the medium term (1–5 years), which means that long-term economic models are not robust.

The Amaze trial aimed to evaluate the clinical and HRQoL benefits, as well as the cost-effectiveness for the NHS, of this technology. The HESTER (Has Electrical Sinus Translated into Effective Remodelling?) observational substudy explored atrial contractile function in maze patients who were in SR at least 1 year after the procedure, compared with cardiac surgery patients who were in SR both before and at 1 year after the procedure.

Objectives

The primary objective was to compare the maze procedure as an adjunct to routine cardiac surgery with routine cardiac surgery alone in terms of:

- return to stable SR at 12 months
- quality-adjusted survival over 2 years.

The key secondary objective was to assess cost-effectiveness over 2 years from a NHS perspective.

The other secondary objective was to compare the two trial arms for return to stable SR at 2 years, overall survival, thromboembolic neurological complications (e.g. stroke), stroke-free survival, anticoagulant and antiarrhythmic drug use and HRQoL.

Prespecified subgroup analysis explored differences in treatment effects between patients with paroxysmal AF and non-paroxysmal AF, surgical centres (as a random effect), cardiac surgical procedures and surgeons. Within the maze arm, the analysis explored differences between ablation devices and lesion sets treated.

The HESTER substudy objective was to assess whether or not patients in SR at least 1 year after an adjunct maze procedure had equivalent active left atrial ejection fraction (ALAEF) to control patients who had undergone cardiac surgery and were in SR both before and after surgery.

Methods

Amaze was a Phase III, pragmatic, multicentre, double-blind, parallel-arm randomised controlled trial to compare clinical, patient and cost outcomes for patients with pre-existing AF who underwent routine cardiac surgery either with or without an adjunct device-based ablation procedure.

Setting

Eleven acute NHS specialist cardiac surgical centres, co-ordinated by the Papworth Trials Unit Collaboration. Participating surgeons had at least 2 years' experience in the use of ablation devices.

Patient recruitment

Consecutive cardiac surgery patients with a history of AF were screened for eligibility. Trial inclusion criteria were as follows: patients aged $\geq$ 18 years, scheduled for elective or in-house urgent cardiac surgery (coronary, valve, combined coronary and valve or any other cardiac surgery requiring cardiopulmonary bypass), with a documented history (> 3 months) of AF (chronic, persistent or paroxysmal). Exclusion criteria included patients who had had previous cardiac operations, emergency or salvage operations surgery without cardiopulmonary bypass and patients who were unlikely to be available for the 2-year follow-up or who were unable to consent.

Randomisation

On the day of surgery, in the anaesthetic room, eligible patients were randomised (1 : 1) to either planned cardiac surgery (control arm) or planned cardiac surgery with additional device-based AF ablation (experimental arm). The allocation sequence was computer generated using permuted blocks (sizes 6 and 8), stratified by surgeon and planned procedure.

Blinding

Although theatre staff could not be blinded to treatment allocation, patients, researchers collecting HRQoL outcomes and cardiologists assessing the 4-day electrocardiography (ECG) results were unaware of treatment allocation.

Treatment arms

In this pragmatic trial, cardiac surgery and postoperative management in the control arm was completed in accordance with standardised hospital protocols.

For patients randomised to maze, the surgeon also administered ablation. The lesion set was at the discretion of the treating surgeon. Any AF ablation device routinely used within the NHS was permitted, including bipolar and unipolar radiofrequency, 'cut-and-sew', cautery, cryotherapy, ultrasound, laser and microwave energy. Postoperative management, subsequent follow-up and data collection were identical to the control arm.

Outcomes

Return to SR at 12 months after surgery and quality-adjusted survival over 2 years were joint primary outcomes. Return to SR was defined as absence of any AF on 4-day continuous ECG recordings,

analysed centrally at Papworth Hospital, by cardiologists unaware of patient identity or treatment arm. Quality-adjusted survival over 2 years was estimated using serial utility measurements from the UK population valuation of the EuroQol-5 Dimensions, three-level version (EQ-5D-3L), administered at randomisation and discharge, and at 6 weeks and 6, 12 and 24 months after the procedure. Quality-adjusted life-years (QALYs) over 2 years were estimated using the area under the curve method.

Secondary outcomes were return to SR at 2 years after surgery, overall survival, stroke-free survival, incidence of hospital admission for haemorrhage, anticoagulant and antiarrhythmic drug usage, HRQoL [measured by the EQ-5D-3L, the Short Form questionnaire-36 items (SF-36) and the New York Heart Association (NYHA)], resource use and trial-based cost-effectiveness of the adjunct maze procedure up to 2 years after randomisation.

Sample size

The maze procedure was considered effective if there was a significant effect for either return to SR at 12 months or QALYs over 2 years. The planned recruitment target of 200 patients per arm was based on detecting a target difference of 15% in the return to SR rate (45% for maze and 30% for control) or 1 additional month of quality-adjusted life (0.083 QALYs, standard deviation 0.3), with approximately 80% power, a two-sided significance of 5% and up to 15% death/loss to follow-up.

Owing to slower than expected accrual, recruitment was terminated in September 2014, when 352 patients had been randomised.

Statistical analysis

The primary analysis used intention to treat, with multiple imputation for missing primary outcomes. For AEs, patients were included in the arm corresponding to the intervention received (maze procedure completed vs. no maze procedure).

Return to SR rates were analysed using binary logistic regression, including surgeon (random effect), baseline heart rhythm and planned surgical procedure (fixed effects). For QALYs > 2 years, linear regression was fitted to utilities post treatment, including surgeon (random effect), baseline utility and treatment arm (fixed effects). For surviving patients with missing EuroQoL measurements, multiple imputation was used, and a confidence interval (CI) for the QALY difference was estimated using non-parametric bootstrapping. No discounting of QALY estimates was applied for the primary outcome. For both primary outcomes, subgroup effects were investigated by including interaction terms.

Overall survival and stroke-free survival were analysed using Kaplan–Meier and Cox regression methods. SF-36 scores were analysed using linear regression, including time point, treatment arm, time-by-treatment-arm interaction and baseline SF-36 scores (all fixed effects), with random intercepts for patients. Drug use was tabulated and analysed using logistic regression, including drug category, time period using drug, baseline drug usage and treatment arm.

Economic analysis

NHS resource use covered the primary admission (operation, time in intensive care, cardiac and acute care wards, transfers to rehabilitation centres or other hospitals), follow-up (including readmissions, diagnostic tests and health-care visits) and drugs (antiarrhythmic, anticoagulant, antiplatelet and cardiac drugs). Resource use was costed using national estimates of unit prices [Department of Health and Social Care (DHSC). *NHS Reference Costs 2014–15*. London: DHSC; 2015], literature (e.g. 24-hour blood pressure monitoring and chest radiography) or information from Papworth Hospital (e.g. theatre cost and cost of device). The ablation device was costed at £3000 per patient for high-intensity focused ultrasound, and at £1250 per patient for all other methods. Both costs and QALYs were discounted at 3.5% in year 2 for the cost-effectiveness analysis [National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013: Process and Methods*. URL: www.nice.org.uk/process/pmg9/chapter/the-reference-case#discounting (accessed 10 January 2018)].

Costs and QALYs were analysed using seemingly unrelated regression, including age, sex, baseline EQ-5D-3L score, baseline AF and, for QALYs only, specific procedure; regression coefficients were used to estimate incremental cost-effectiveness ratios (ICERs). Probabilistic sensitivity analysis used bootstrapping. Cost-effectiveness planes, the cost-effectiveness acceptability curve and incremental net monetary benefit were estimated. Deterministic sensitivity analysis explored the impact of using Short Form questionnaire-6 Dimensions QALYs, complete-case analysis, truncating costs and QALYs at discharge, excluding outliers and alternative imputation strategies.

The 'Has Electrical Sinus Translated into Effective Remodelling?' substudy

To assess whether or not contractile function after maze procedure was equivalent to that for non-AF patients, 22 maze procedure patients who were in SR at least 1 year postoperatively, were matched (1 : 1) to non-trial control patients who were in SR before, and at least 1 year after, routine cardiac surgery. Matching criteria were time since procedure, age, sex, procedure, preoperative left ventricular ejection fraction and logistic European System for Cardiac Operative Risk Evaluation (EuroSCORE).

Eligible patients underwent ECG to confirm SR, transthoracic two- and three-dimensional echocardiography and cardiac magnetic resonance imaging (MRI). The primary outcome was ALAEF; left atrial volumes and an ECG marker of left ventricular function (E/A ratio) were secondary end points.

Sample size

The minimum clinically important difference in ALAEF was set at 18.2%. Equivalence was concluded if the two-sided 95% CI of the estimated treatment effect (maze–control) was entirely in the interval (95% CI –18.2% to 18.2%). Twenty-two matched pairs provided 80% power to demonstrate equivalence.

Statistical analyses

For the primary end point, the linear regression model, including treatment, matching variables (fixed effects) and matched pairs (random effect), was fitted.

Results

Between 25 February 2009 and 6 March 2014, 1013 patients were screened in 11 UK specialist cardiac surgery centres and 352 patients were randomised to the control ($n = 176$) or experimental ($n = 176$) arms. Thirty surgeons participated in the trial. The SR status of patients at 12 months was available for 141 maze procedure and 145 control patients (80% and 82%, respectively); QALYs up to 2 years were available for 160 patients in each arm (91%).

Primary outcome results

Among complete cases in the maze procedure arm, 87 out of 141 patients (61.7%) were in SR compared with 68 out of 145 (46.9%) control patients. The odds ratio (OR) (95% CI) for return to SR was 2.06 (1.20 to 3.54; $p = 0.0091$). Surgical results varied by surgeon in both groups, but the treatment effects did not. Results were broadly consistent across subgroups.

In both trial arms, QALYs could be estimated for 160 patients. Unadjusted, undiscounted mean QALYs (95% CI) over 2 years were 1.489 (1.416 to 1.558) for maze procedure patients and 1.485 (1.403 to 1.559) for control patients. The mean difference (95% CI) in QALYs at 2 years (maze–control) was –0.025 (–0.129 to 0.078; $p = 0.6319$). Results did not vary by surgeon or subgroup.

Secondary outcomes

In the maze procedure arm, 69 out of 118 (58.5%) completers were in SR compared with 47 out of 129 (36.4%) completers in the control arm. The adjusted OR for patients in SR at 2 years was 3.24 (95% CI 1.76 to 5.96). The number of patients requiring anticoagulant drug therapy was significantly lower in the

maze arm from 6 months to 2 years post procedure. Slightly more maze procedure patients required antiarrhythmic drugs throughout follow-up, but the difference was not statistically significant.

There were no significant differences between the arms for any of the following secondary outcomes at any time point: operative or overall survival, stroke-free survival, need for cardioversion or permanent pacemaker implants, NYHA score, EQ-5D-3L utility and SF-36 dimensions.

Safety

Sixty per cent of patients in each arm had a serious adverse event ($p = 1.000$); most events were mild, but 71 (42.5%) maze procedure patients and 84 (45.5%) control patients had at least one moderately severe event, and 31 (18.6%) maze procedure patients and 38 (20.5%) control patients had a severe event. Twenty-three events in 17 (10.2%) patients were possibly related to treatment in the maze procedure arm compared with 28 events in 19 (10.3%) patients in the control arm; one patient (0.5%) in the control group was admitted to hospital for investigation of atrial flutter, classed as 'definitely related' to treatment.

Cost-effectiveness

The mean additional cost of the maze procedure was £3533 (95% CI £1321 to £5746), which was statistically significant, but the mean difference in QALYs was not statistically significant (–0.022, 95% CI –0.1231 to 0.0791). None of the analyses suggested that the maze procedure was cost-effective at £30,000 per QALY over 2 years. The smallest ICER was £83,625 per QALY for the complete-case analysis.

The 'Has Electrical Sinus Translated into Effective Remodelling?' substudy

Between 24 July 2013 and 8 July 2015, 22 eligible patients were recruited for each cohort and underwent echocardiography and MRI. The mean difference (95% CI) in ALAEF between maze procedure and control patients was –8.03 (–12.43 to –3.62). The 95% CI was contained entirely in the interval (–18.2 to 18.2), so that the predefined criterion for equivalence was met. However, the mean ALAEF was significantly lower in maze procedure patients than in control patients ($p = 0.0015$).

Mean E/A ratio was significantly higher and mean left atrial ejection fraction (four-chamber view and MRI) was significantly lower for maze procedure patients than for control patients. There were no significant differences in the other end points.

Conclusions

Implications for future health care

The Amaze trial demonstrated that ablation can be practised safely in a routine NHS cardiac surgical setting and that it increases the proportion of patients who return to SR up to 2 years after surgery. Clinical effects did not translate into improved survival or QALYs, and the addition of the maze procedure was not cost-effective over 2 years.

The reduction in anticoagulant drug use and results of the substudy provide support for anticoagulant drug withdrawal, but varying rates of left atrial functional recovery after the maze procedure mean that atrial function should be measured before considering withdrawal of anticoagulant drugs.

Implications for further research

The clinical results are promising, and continued follow-up of clinical events, HRQoL and long-term clinical effectiveness and cost-effectiveness analysis is warranted.

Subgroup analyses had low power to provide robust recommendations on specific methods. Further comparison of ablation methods would inform best practice.

Trial registration

This trial is registered as ISRCTN82731440.

Funding

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

Chapter 1 Background

The health problem

Atrial fibrillation (AF) is the most common disturbance of heart rhythm. It is characterised by an irregular heartbeat caused by low-amplitude, supraventricular oscillations.[1]

The normal rhythm of the heart is sinus rhythm (SR). The stimulus to beat is triggered by the sinoatrial node. This results in atrial contraction while the stimulus is conducted to the atrioventricular node, which relays it to the ventricles, initiating ventricular contraction. AF is a condition whereby the atria beat unevenly because of abnormal electrical signals. As a result, there is no SR, the ventricles respond to the disordered atrial electrical activity in a haphazard fashion, the pulse becomes irregular and the entire heartbeat is abnormal.

Atrial fibrillation has a prevalence of 1–2% of the population in high-income countries.[2,3] In the UK, the prevalence of AF is 7.2% in patients aged ≥ 65 years and 10.3% in patients aged ≥ 75 years.[4] Prevalence is associated with age and comorbidities, such as obesity, diabetes and hypertension. With the advancing age of the population and the increasing prevalence of obesity, this proportion is likely to increase.[5]

Consequences of atrial fibrillation

Atrial fibrillation can cause palpitations, chest pain, dizziness and breathlessness, and imposes a heavy burden on both patients and clinicians, as it has a considerable impact on quality of life (QoL) and NHS resources.[6]

When the atria fibrillate, they lose their pumping action, and this has two very important sequelae.

There is blood stagnation in the atria, which can lead to clot formation. Blood clots can then exit the heart (thromboembolism), leading to stroke and other complications. There is an associated four- to fivefold increased risk of thromboembolic stroke in AF, and if AF is left untreated, around 1 in 25 patients will have a stroke.[7] The NHS devotes 5% of its budget to preventing and treating strokes, and 15% of strokes can be attributed to AF.[4] Drugs and other treatments can control AF, but not without complications. Routine anticoagulant drug treatment reduces the risk of stroke in AF patients by two-thirds, but this incurs an increased risk of bleeding and needs careful monitoring. The substantial burden of monitoring anticoagulant therapy usually falls on general practice, anticoagulant clinics and haematology laboratories.

When the atria do not pump, the heart is less efficient. The presence of AF exacerbates heart failure that arises from other heart conditions, and can itself cause heart failure, especially if the fibrillating atria dilate and this results in leakage of the mitral and tricuspid valves.

Treatment of AF and its consequences (antiarrhythmic and anticoagulant drugs, hospital monitoring and stroke treatment) is expensive for the NHS. Implementation of the 2006 National Institute for Health and Care Excellence (NICE) guidelines,[6] on management of AF, was estimated to cost £21.86M per year.[3]

Types of atrial fibrillation

Atrial fibrillation is classified into three distinct subgroups. In the Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation,[8] paroxysmal AF is defined as recurrent AF (more than two episodes) that terminates spontaneously within 4 days, persistent AF is defined as AF that continues beyond 4 days and chronic or longstanding AF is persistent AF beyond 1 year.[8] These patterns of AF have slightly different electrophysiological mechanisms.

Diagnosis of atrial fibrillation

Atrial fibrillation is confirmed by electrocardiography (ECG) when the patient presents with symptoms of palpitations, as is the case with paroxysmal AF; however, both paroxysmal and more longstanding AF can be asymptomatic and, when that is the case, can be missed on the electrocardiogram, unless the AF coincides with the time at which ECG is carried out. This is especially important in assessing treatments aimed at correcting AF, as many historical series reported only the recurrence of symptomatic AF and have used only intermittent ECG for follow-up.[9] In the majority of published studies, success in the 'resumption of SR' was based on a single ECG recording at 12 months, which is not necessarily representative. More recently, there has been recognition of the importance of more robust AF documentation and Holter monitoring records have been reported in some AF trials.[10–12] Better evaluation of any residual AF after treatment, by continuous ECG monitoring over several days, documents the percentage of time for which a patient is in AF, and this is called the 'AF burden'.

A number of studies (cited in Calkins *et al.*[8]) suggest not only that AF increases the risk of a poor outcome from prospective cardiac surgery, but is also that AF an independent risk factor for early and late morbidity and mortality. This leads to the (unproven) hypothesis that efforts to eliminate pre-existing AF during cardiac surgery may improve survival and reduce adverse cardiac events after surgery.

Current treatment

Until the 1980s, AF was treated using antiarrhythmic drugs and direct current cardioversion. When that failed, AF was managed with rate control medication and anticoagulant drugs to reduce the risk of stroke. There has been a substantial development in our understanding of the pathophysiology of AF, and two very important findings have been the roles of the pulmonary veins and macro-re-entry circuits.

We now know that the majority of electrical trigger points that initiate AF lie within the pulmonary veins and not in the atria themselves.[13] Moreover, the maintenance of AF depends on the presence of macro-re-entry circuits, in which delayed conduction of the electrical signal means that the signal arrives at the originating point when that point is no longer refractory. These circuits are quite large (several centimetres). As a result of this knowledge, the maze procedure was developed in the 1980s by Cox and Boineau.[14] The procedure prescribed a number of surgical cuts aimed at achieving two objectives: electrical isolation of the pulmonary veins from the atria, thus dealing with the site of most AF trigger points, and further cuts in the atria to disrupt macro-re-entry circuits, thereby preventing the maintenance of the AF rhythm. The maze procedure therefore involves multiple cutting and sewing of the atria and pulmonary veins. Several studies have reported freedom from AF ranging from 75% to 90% after the Cox maze procedure (see Huffman *et al.*'s[9] systematic review and associated references), and one reported a 15-year success rate in restoring SR as high as 94%.[15] This traditional cut-and-sew technique, despite being available since 1987, has failed to achieve widespread use, as it is technically demanding and adds substantially to the operative burden of a heart operation. It is currently in very limited use by a few surgeons in a few centres and tends to be reserved for otherwise fit patients with severely symptomatic AF, who are prepared to take the risk of such a major intervention to relieve their symptoms.

Alternatives to the cut-and-sew maze procedure have been produced. A number of devices have been developed to achieve the electrical block needed, using energy sources to ablate atrial tissue. These have made a technically difficult and time-consuming operation easier, quicker and safer for cardiac surgeons to perform. Ablation devices use an energy source (heat, cold, radiofrequency or microwave) to replicate the lesion set produced by the cut-and-sew maze procedure.[16] As a rule, the procedure is safe and well tolerated and adds little to the length and burden of the operation, but these devices are a new and costly technology, which is currently being heavily marketed to treat AF.[17]

There has been no direct comparison between the traditional Cox maze procedure and the ablation maze procedure,[9] presumably because of the problems of incorporating such technically demanding surgery into an adequately powered randomised controlled design. However, a propensity analysis that matched patients who underwent the ablation maze procedure with those undergoing the Cox maze procedure showed no differences in freedom from AF at 3, 6 and 12 months afterwards.[18]

Patients who have AF before surgery are generally older and have an increased procedural risk and other comorbidities, so that treating AF at the time of cardiac surgery may be advantageous to the patient. When the Amaze trial was planned, the only evidence supporting this came from five small randomised controlled trials (RCTs) of the ablation maze procedure as an adjunct to surgery.[11,19–22] These trials found that SR was restored in 44–94% of treated patients compared with 5–33% of control patients. The trials were small and follow-up was short. Success was mostly defined on the basis of single ECG recordings. No trial looked at patient-centred outcomes or cost-effectiveness. Despite this lack of robust evidence, the number of patients with AF undergoing open heart surgery and being offered concomitant ablation maze procedures, or 'adjunct maze procedures', was increasing. Although there were instances of its use as a standalone procedure, the widest use of the ablation maze procedure in the NHS was in patients already having cardiac surgery for other problems.[23] This trial was designed in response to the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme call to evaluate ablation devices that were being rapidly assimilated into NHS practice without formal assessment.

Patient benefit

It is essential to recognise that the primary justification of the ablation procedure is to treat symptomatic AF.[24] When the Amaze trial was in the planning stage, no effectiveness studies had investigated the QoL benefit and cost-effectiveness of maze procedures. Since the original grant application, four small randomised trials have reported on health-related quality of life (HRQoL) after the maze procedure.[12,25–27] No difference was reported in the overall QoL scores. However, patients in the studies by Gillinov *et al.*[12] and Van Breugel *et al.*[27] were not blinded to the treatment they received, which could have influenced the reporting of QoL outcomes. Cherniavsky *et al.*[25] reported improvement in the Short Form questionnaire-36 items (SF-36) score; however, the trial overall reported that the adjunct maze procedure arm did no better than coronary artery bypass graft operation (CABG) alone.

Recent evidence

A Cochrane collaboration review[9] assessed the effects of adjunct AF surgery.[9] Using a comprehensive systematic review methodology, the authors identified 22 published trials (1899 participants) comparing cardiac surgery with and without adjunct AF surgery, with five additional ongoing studies and three studies not classified at the time of reporting.[10–12,19,20,22,25–40] All included studies were rated as being at a high risk of bias in at least one domain assessed. The Cochrane review[9] found that AF surgery, regardless of technique, doubles the rate of freedom from AF, atrial flutter and atrial tachycardia [51.0% vs. 24.1%; relative risk (RR) 2.04, 95% confidence interval (CI) 1.63 to 2.55], with more patients not taking antiarrhythmic medication 3 months after cardiac surgery. There was little evidence of a difference between patients with AF who were treated and those who were not treated in either 30-day mortality (2.3% vs. 3.1%; RR 1.25, 95% CI 0.71 to 2.20) or all-cause mortality (7.0% vs. 6.6%; RR 1.14, 95% CI 0.81 to 1.59). However, patients who were treated for AF were more likely to be fitted with a permanent pacemaker (6.0% vs. 4.1%; RR 1.69, 95% CI 1.12 to 2.54). The review authors concluded that there remained uncertainty about the effects on cardiovascular mortality, adverse events (AEs), HRQoL and long-term outcomes.[9]

In summary, there is little rigorous evidence that attempting to restore SR by treating AF with an ablation device during cardiac surgery is of benefit to the patient. Nevertheless, these devices are being incorporated into routine practice nationally and internationally. The Amaze trial provided a timely evaluation of this technology with the objective of assessing the clinical and HRQoL benefits for patients, as well as cost-effectiveness for the NHS.

Chapter 2 Amaze trial methods

Objectives

Primary objectives

The primary objectives were to compare patients undergoing the maze procedure as an adjunct to routine cardiac surgery with patients undergoing routine cardiac surgery alone, in terms of:

- return to stable SR at 12 months
- quality-adjusted survival over 24 months after randomisation.

Secondary objectives

The main secondary objective was to assess the cost-effectiveness of the adjunct maze procedure, relative to cardiac surgery alone, from a NHS perspective.

Other secondary objectives were to compare the following outcomes between the two arms:

- return to stable SR at 24 months after surgery
- overall survival
- thromboembolic neurological complications (e.g. stroke)
- stroke-free survival
- anticoagulant and antiarrhythmic drug use up to 24 months post randomisation
- HRQoL up to 24 months post randomisation, as measured by the EuroQol-5 Dimensions, three-level version (EQ-5D-3L), the SF-36 and the New York Heart Association (NYHA)
- resource use and costs.

Exploratory analyses

Prespecified subgroup analysis was planned to explore differences in treatment effects between:

- patients with paroxysmal AF and non-paroxysmal AF (i.e. persistent, chronic or longstanding AF)
- individual centres (as a random effect)
- cardiac surgical procedures
- surgeons.

Within the maze treatment arm, analysis was planned to explore differences between:

- different ablation devices
- different lesion sets treated.

Design

Overview

The Amaze trial was a Phase III, pragmatic, multicentre, double-blind, parallel-arm RCT to compare clinical, patient-based and cost outcomes for patients with pre-existing AF who undergo routine cardiac surgery either with or without an adjunct device-based ablation procedure.

Eligible patients were randomised (in a 1 : 1 ratio) to receive either:

- planned routine cardiac surgery with no additional procedure
- planned routine cardiac surgery with an additional device-based AF ablation procedure.

The study was reviewed and approved by the Essex 1 Research Ethics Committee (reference number 08/H0301/98) and was registered as International Standard Randomised Controlled Trial Number 82731440 (ISRCTN82731440). The trial protocol can be accessed at www.papworthhospital.nhs.uk/research/data/uploads/ptuc/protocol-v4-may-20151.pdf (accessed 2 March 2018)[41] and the HESTER (Has Electrical Sinus Translated into Effective Remodelling?) substudy protocol can be accessed at www.papworthhospital.nhs.uk/research/data/uploads/ptuc/hester-study-protocol-3.pdf (accessed 2 March 2018).[42]

As much as possible, the design and reporting of this trial adhered to the guidelines of the Consolidated Standards of Reporting Trials (CONSORT) statement[43] and incorporated recommendations of the Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation.[24]

Patient and public involvement

Mr Brian Elliott was an independent member and lay representative on the Trial Steering Committee (TSC), providing input on all aspects of the trial design, conduct and recruitment strategies. Professor Paul Kinnersley has a background in primary care and public health and provided further independent advice on the conduct and progress of the trial through membership of the TSC.

Setting and investigators

Eleven acute NHS specialist cardiac surgical centres (Papworth Hospital, Cambridgeshire; Royal Brompton Hospital, London; Brighton and Sussex University Hospitals, Brighton; University Hospitals Coventry and Warwickshire NHS Trust, Coventry; Glenfield Hospital, Leicester; Wythenshawe Hospital, Manchester; Northern General Hospital, Sheffield; Guy's and St Thomas' Hospital, London; Derriford Hospital, Plymouth; Freeman Hospital, Newcastle upon Tyne; and Blackpool Victoria Hospital, Blackpool) participated in the study, which was co-ordinated by the Papworth Trials Unit Collaboration. For surgeons to participate, they were required to be experienced in the use of ablation devices for at least 2 years. During the trial design period, participating surgeons met to agree the permissible procedures, ablation methods and lesion sets to be treated.

Participants

Consecutive cardiac surgical patients undergoing major cardiac surgery (e.g. coronary, valve or combined operations), with a history of paroxysmal, persistent or chronic AF beginning > 3 months before the date of the operation, were screened for eligibility.

- Paroxysmal AF was defined as recurrent AF (two or more episodes) that terminated spontaneously within 4 days.[24]
- Non-paroxysmal but persistent AF was defined as AF that continued for > 4 days.
- Chronic or longstanding AF was defined as AF that was persistent for > 1 year.

The Amaze trial inclusion criteria included patients who:

- were aged > 18 years
- were scheduled to undergo elective or in-house urgent cardiac surgery (coronary surgery, valve surgery, combined coronary and valve surgery or any other cardiac surgery requiring cardiopulmonary bypass)
- had a history of documented AF (chronic, persistent or paroxysmal) beginning > 3 months before entry into the study
- were willing to provide written informed consent to participate.

Exclusion criteria included patients who:

- had had previous cardiac operations
- had had emergency or salvage cardiac operations
- had had surgery without cardiopulmonary bypass
- were unlikely to be available for follow-up over a 2-year period
- were unable to provide consent.

Recruitment

The procedure for informing and obtaining consent from patients was devised to accommodate local variations in the patient pathway, but was otherwise identical for all centres. At most centres, potential participants were initially given a simple summary of the study by the local investigator at the initial surgical clinic when treatment options were discussed. Before the next attendance, the trial co-ordinator or a research nurse contacted the patients by telephone to assess interest in the trial, and posted the full patient information sheet to the homes of those who expressed an interest. Written consent was taken at the pre-admission clinic, at approximately 2 weeks prior to surgery, by the trial co-ordinator or research nurse. HRQoL questionnaires were administered by the research nurse, after consent, at this pre-admission clinic. Thereafter, patients were registered for the trial and provided with a 4-day ECG recording device and instructions on its use, to take home to monitor their heart rate for 4 days. On admission for surgery, the patient returned the 4-day ECG recorder.

Randomisation

Eligible patients who satisfied the inclusion criteria and provided written consent were randomised (in a 1 : 1 ratio) to receive either their planned cardiac surgery with no additional procedure or their routine cardiac surgery with an adjunct maze procedure.

The allocation sequence was generated by permuted block randomisation (using block sizes of 6 and 8), and randomisation was stratified by surgeon and planned cardiac procedure (CABG, aortic valve, mitral valve or combined procedure). On the day of surgery, when the patient was in the anaesthetic room, the local centre contacted the Papworth Trials Unit Collaboration by telephone. Patient details, surgeon and planned cardiac procedure were registered with the Papworth Trials Unit Collaboration, whose staff were not otherwise involved with the trial. Once registration was complete, the allocation was released to the surgical team, which was also responsible for completing the surgical clinical report form (CRF). The treatment allocation was not made available to any other staff who were directly or indirectly involved in the trial.

Blinding

Although theatre staff could not be blinded to the treatment allocation, the trial was double-blind to the extent that neither the patients themselves, nor any researchers collecting HRQoL outcomes nor the cardiologists assessing the 4-day ECG results were aware of the trial arm to which the patient had been allocated.

Patients' medical notes were labelled to indicate that they were Amaze trial participants. Routine reports provided details of their elective surgery and the fact that they were randomised within the Amaze trial. The surgical CRF describing the research intervention (maze procedure or no maze procedure) was placed in a sealed envelope labelled 'The Amaze Trial' and kept in the patient's notes. Clinicians were able to access this information in the event of a serious adverse event (SAE) considered to be related to surgery. At discharge, the data management staff retrieved the sealed envelope and uploaded the procedure details onto a secure database, before resealing the envelope and returning it to the notes. Cardiologists who analysed electrocardiograms for the primary outcome and researchers recording HRQoL outcomes did not have notes, including the procedural information, available at the time of outcome assessment.

Interventions

Control arm

The Amaze trial was planned as a pragmatic trial, following standard treatment and care as closely as possible in order to assess outcomes in a real-world context. Patients randomised to the control arm received preoperative management, elective or in-house urgent cardiac surgery and postoperative management in accordance with standardised hospital protocols.

Experimental arm

Patients randomised to the experimental arm received preoperative management and elective or in-house urgent cardiac surgery, as described in standardised hospital protocols. During the operation, the conduct of the adjunct maze procedure and the choice of lesion set was at the surgeon's discretion. At the time of trial design, there was no evidence for the superiority of one ablation device or one energy source over another. Therefore, any AF ablation device that was routinely used within the NHS by the investigators was permitted. This allowed surgeons to use the devices with which they were most familiar and comfortable, and which were in routine use at their institution. These included bipolar and unipolar radiofrequency, 'cut-and-sew', cautery, cryotherapy, ultrasound, laser and microwave energy. Postoperative management, subsequent follow-up and data collection were identical to the control arm.

Standardisation between centres

In order to minimise potential confounding by other components of a patient's care, the following aspects of the trial were standardised across participating hospitals.

Management of patients before, during and after surgery

Management was undertaken in accordance with the local site's normal practice, irrespective of randomisation. The only exceptions were processes required to maintain the blinding of the patient, cardiologist and QoL interviewer (see *Blinding*).

Conducting and reporting on the adjunct maze procedure and defining the prescribed lesion set

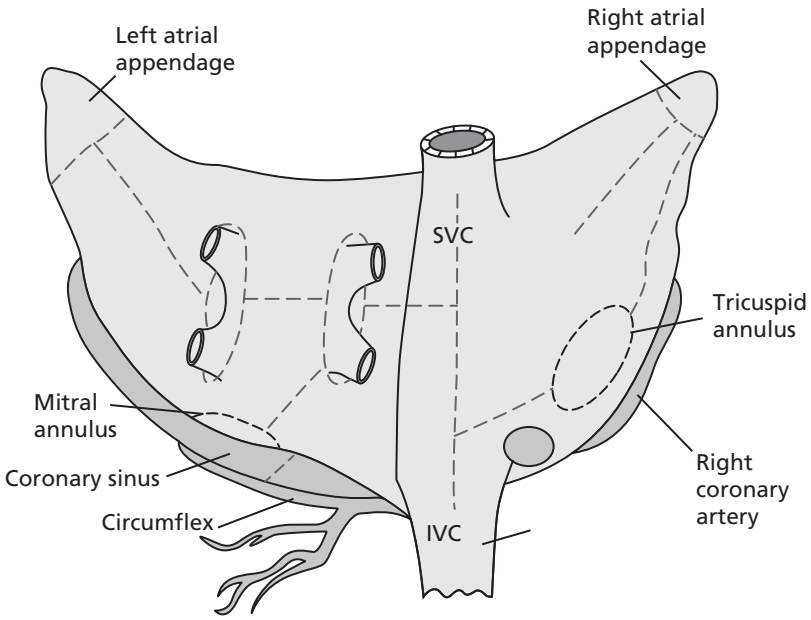
The full lesion set considered is illustrated in *Figure 1*, although the specific lesion set treated was left to the participating surgeon's discretion. Details of the ablation procedure and all lesions treated were documented. Published guidelines for reporting data and outcomes for surgical treatment of AF were followed.[8,15]

Postoperative drug use

- Amiodarone: unless contraindicated, 200 mg three times per day was prescribed, reducing over a period of 3 weeks to 200 mg per day for 6 weeks. The drug was stopped if stable SR was established at 6 weeks. Further prescription after this period was based on individual clinical judgement.
- Warfarin: prescribed until the patient was in stable SR. Thereafter, centres adopted normal practice.
- Beta-blockers: prescribed at the individual clinician's discretion.
- Other drugs: prescribed at the individual clinician's discretion; cardiac drugs with antiarrhythmic, antihypertensive and anticoagulant actions, including aspirin and warfarin, were documented.

Indications for cardioversion, timing and number of attempts

The protocol did not require cardioversion to be carried out at discharge, but if it was performed for clinical reasons, the details were recorded. For patients in AF at the first follow-up appointment, cardioversion was attempted within 3 months of surgery. If cardioversion was unsuccessful, then it was attempted again at 6 months after surgery.



Left side	Right side
1. Around RPV 2. Around LPV 3. Connecting RPV to LPV 4. Connecting RPV to mitral annulus 5. Left atrial appendage 6. Left atrial appendage to LPV Note that some lesions were not treated in some patients	1. SVC to IVC 2. SVC–IVC to tricuspid annulus 3. Trans-septal SVC–IVC to RPV 4. Right atrial appendage 5. Right atrial appendage to RA body 6. Right atrial appendage to tricuspid annulus 7. Coronary sinus ostium

FIGURE 1 Complete modified Cox maze procedure III lesion set. IVC, inferior vena cava; LPV, left pulmonary vein; RA, right atrium; RPV, right pulmonary vein; SVC, superior vena cava.

Outcome measures

Primary outcomes

Return to sinus rhythm at 12 months

Sinus rhythm at 12 months after surgery was determined by the absence of any AF on outputs from continuous monitoring by 4-day ECG recorders. All 4-day continuous ECG recordings were analysed centrally at Papworth Hospital. Participating hospitals forwarded the anonymised secure digital cards from the ECG recorders to Papworth Hospital. Analyses using the proprietary automated software package, together with manual checking of the recording in its entirety, were completed by cardiologists who were not aware of the patient's identity or allocated treatment arm. Total time spent in SR and in AF (AF burden) during the 4-day recording was calculated. Episodes of atrial flutter were noted and included in the AF burden.

Quality-adjusted survival over 2 years

The EQ-5D-3L was administered at randomisation, on discharge and at 6 weeks and 6, 12 and 24 months after the procedure (*Table 1*).[44] Although the current version of the EuroQol-5 Dimensions has five levels for each item, when the Amaze trial was designed, the three-level version was recommended for the cost-effectiveness analysis.[45] EQ-5D-3L responses were converted into utility scores reflecting values from a representative sample of the UK population.[46] Clinical effectiveness was measured by quality-adjusted life-years (QALYs) over 2 years using the area under the curve method (see *Statistical analyses*).[47]

TABLE 1 Health-related quality-of-life questionnaires administered

Type	Questionnaire	Description
Generic	EQ-5D-3L	• Five dimensions (morbidity, self-care, usual activities, pain/discomfort and anxiety/depression), each having three levels (no problems, moderate problems, severe problems) • One derived utility measure (EQ-5D-3L) measured on a scale from –0.591, through 0 (representing death), to 1 (representing full health) • One overall health scale from 0 to 100 (not analysed)
Generic	SF-36	• Eight dimensions (physical functioning, role limited because of physical problems, pain, energy/vitality, social functioning, mental health, role limited because of emotional problems and general health). Dimension scores range from 0 (minimum function) to 100 (maximum function) • One derived utility measure (SF-6D) ranging from 0 (death) to 1 (full health)
Specific	NYHA	Breathlessness was classified on a four-point scale: • Class I – no limitation during ordinary physical activity • Class II – ordinary physical activity slightly limited. Comfortable at rest, but ordinary physical activity results in fatigue, palpitation or dyspnoea • Class III – marked limitation of physical activity. Comfortable at rest, but less than ordinary activity causes fatigue, palpitation or dyspnoea • Class IV – unable to carry on any physical activity without discomfort; symptoms of cardiac insufficiency at rest

SF-6D, Short Form questionnaire-6 Dimensions.

Secondary outcomes

Sinus rhythm was determined by the absence of any AF on 4-day ECG recorders at 24 months after surgery. Other secondary outcomes were overall survival from the date of randomisation to date of death (all patients were registered with the Office for National Statistics' tracking system to allow long-term follow-up of survival); stroke-free survival, defined as the time between randomisation and the date of stroke or death (whichever occurred first); incidence of hospital admission for (anticoagulant-related) haemorrhage, anticoagulant and antiarrhythmic drug usage up to 24 months after randomisation; HRQoL measured by the EQ-5D-3L, the SF-36 and the NYHA for breathlessness (completed at baseline, on discharge and at 6 weeks and 6, 12 and 24 months post surgery after randomisation); and resource use and trial-based cost-effectiveness of the adjunct maze procedure up to 24 months after randomisation.

Health-related quality of life

Health-related quality-of-life interviews were conducted by clinical research co-ordinators/research nurses in face-to-face interviews at hospital research clinics. Six-month questionnaires were administered by telephone, as patients did not have clinical appointments. The questionnaires administered are summarised in *Table 1*. The SF-36 consists of eight dimensions, which are the weighted sums of the questionnaire item responses. Each dimension ranges from 0 to 100, with a higher score representing better health or fewer limitations for that domain. Standardised physical and mental health scores were calculated, which, for a general UK population, are expected to be approximately normally distributed with a mean of 50 and a standard deviation (SD) of 10.[48] For missing items, we used the methods recommended in the manual.[49] Briefly, if at least half the items were available for any scale, the mean of the recorded items was imputed for the missing items. If more than half the items for a scale were missing, then the scale was recorded as missing.

Sample size

The dual primary outcomes were a clinical end point (return to SR at 12 months) and an outcome of importance to patients and service providers (quality-adjusted survival over 2 years). The maze procedure was considered effective if there was a significant effect for return to SR or if the mean difference in QALYs between the groups did not include zero. No adjustments were made to the sample size to

accommodate the multiple testing of these two outcomes, which is inherent in this approach, as the focus for the QALY end point was on estimation of the treatment effect, rather than hypothesis testing. However, to guard against overinterpretation of hypothesis tests, we recommend that *p*-values between 0.025 and 0.05 are considered to have borderline significance.

Return to sinus rhythm at 12 months

Prior to the trial, published RCTs of ablation as an addition to cardiac surgery reported rates of return to SR at 12 months ranging from 44% to 87% in the maze procedure arms, and from 5% to 33% in the control arms.[19,22] We took a conservative estimate of the difference between the arms (45% vs. 30%) as the target effect. In order that a realistic recruitment target was achieved, 80% power was used in the calculation. Combining this with a two-sided significance of 5%, an estimated sample size of 176 in each arm (total of 352) would be sufficient to detect this effect. With planned recruitment of 400 patients, this allowed for approximately 15% death/loss to follow-up at 12 months.

Quality-adjusted survival over 2 years

The emphasis in cost-effectiveness studies is on estimation, rather than hypothesis testing, so that formal sample size calculations were considered less important. However, we provided a power calculation based on the effectiveness measure 'QALYs at 2 years post randomisation'. We could find no studies reporting comparative QALYs in similar patients undergoing ablation and cardiac surgery. From previous studies of patients undergoing angiography for suspected ischaemic heart disease and patients with refractory angina, the SD of QALYs over 12 and 18 months was at most 0.3.[50,51] Over 2 years, the minimum clinically important improvement was considered to be 1 extra month of quality-adjusted life or 0.083 QALYs. With a sample of 200 patients per arm (total of 400), we would have exactly 79% power to detect a difference of 0.083 QALYs (at a two-sided significance of 5%).

If the accepted threshold for cost-effectiveness was in the range £20,000–30,000 per QALY and we could demonstrate a significant increase in QALYs of 0.083, then the procedure would be cost-effective for an incremental cost of, at most, £2500.

Failure to reach target recruitment

Based on audit data, our target recruitment of 400 patients was expected to be achieved in 18 months at six centres. Owing to the slower than expected accrual, recruitment terminated in September 2014, when 352 patients had been randomised, with approximately 70% power to identify the target treatment effects.

Analysis populations

Intention-to-treat population

The primary analysis used the intention-to-treat (ITT) population, defined as all randomised patients, regardless of eligibility, withdrawal, compliance with the protocol, loss to follow-up or actual treatment received. No patients withdrew consent for their data to be used, despite withdrawing from trial follow-up. Multiple imputation was used for missing primary outcomes.

Quality-of-life population

For each instrument (the SF-36 and the EQ-5D-3L), all patients who returned a completed baseline questionnaire, regardless of subsequent questionnaire return, were included in the analysis. In addition, imputation (based on planned procedure and centre) of missing baseline EQ-5D-3L scores was completed for two patients.

Safety population

All patients were included in the safety population if they underwent a surgical procedure. Patients were included in the arm corresponding to the intervention received (maze procedure completed vs. no maze procedure).

Statistical analyses

All statistical analyses and reporting complied with CONSORT guidelines where possible.[52]

Formal analyses were conducted using a two-sided 5% level of significance, with no adjustment for multiple testing. All analyses were performed using SAS, version 9.4 (SAS Institute Inc., Cary, NC, USA). The statistical analysis plan is provided in *Appendix 1*.

In descriptive summaries, the number of non-missing items and the mean (SD) or median (upper and lower quartiles) were summarised for continuous variables, and the number and proportion by treatment arm were summarised for each level of categorical variables.

Return to sinus rhythm

The odds of being in SR for maze procedure patients was compared with the odds for control patients, and estimated using a binary logistic regression model, including surgeon (normal random effects on the logistic scale), baseline heart rhythm and planned surgical procedure (fixed effects). The odds ratio (OR) for return to SR was reported with the 95% CI and *p*-value from this model. Validity of logistic regression models was assessed by examining the following statistics and graphical summaries:

- Pearson residuals/deviance (half-normal plots)
- leverage values
- Cook's distance
- cross-validation probabilities (the probability of a particular observation, conditional on the remaining observations)
- *L*-statistics (the influence of an observation on the difference in deviance as a result of fitting the treatment effect).

The percentage of time in AF across the 4 days of monitoring at baseline and at 12 months was summarised by treatment arm.

Quality-adjusted survival

For the primary outcome, QALYs over 2 years were estimated from serial measurements of the EQ-5D-3L for each patient. The UK social tariff for the EQ-5D-3L, completed at baseline and on discharge, and at 6 weeks and 6, 12 and 24 months post surgery, as estimated by Dolan *et al.*,[46] was applied to calculate utility values. Using actual rather than nominal times of assessment, and assuming a linear change in values between time points, patient-specific utility curves up to 24 months post randomisation were calculated. A value of zero was assigned at the date of death for patients who died. QALYs were calculated as the area under the utility curve to 24 months or date of death, whichever occurred first. In order to adjust for differences in baseline utilities, a linear regression was fitted to the utilities post treatment, with baseline utility and treatment arm as explanatory variables. For patients who did not complete all EQ-5D-3L questionnaires or those who were censored, multiple imputation was used to estimate mean QALYs (see *Missing data*). Model fit was assessed by examining standardised residuals and association with the predicted values, as well as identifying influential observations by referring to leverage statistics. A CI for the true difference in QALYs was estimated using a non-parametric bootstrap resampling approach.[53]

Note that, for the primary outcome analysis, no discounting of the QALYs estimates was applied. However, when costs and benefits were estimated for the health economics analysis, both were discounted at 3.5% for the second year.

Missing data

Proportions of, and reasons for, missing data were investigated.

For the clinical primary end point (return to SR at 12 months), if a patient withdrew consent or was lost to follow-up within 12 months, the missing outcome (AF or SR at 12 months) was multiply imputed as a function of the baseline heart rhythm, surgeon, surgical procedure and treatment arm.[54] Rubin's rules were used to combine imputed data sets.[55]

For the patient-based primary end point (QALYs), whereby a patient died before the end of follow-up, the utility value of 0 was imputed for all subsequent assessments. If the response was missing, and the patient was alive, the missing value was imputed using the method of multiple imputation.[54] A number of sensitivity analyses related to missing data were completed; further details are provided in *Appendix 1*. In response to a reviewer's request, we also provided a supportive analysis using complete cases, but with predictors of missingness included in the model.

Subgroup analysis

Prespecified subgroups are listed above (see *Exploratory analysis*).

For the SR end point, a logistic regression model was fitted to heart rhythm at 12 months, including baseline heart rhythm, surgeon, surgical procedure, treatment arm and subgroup variable of interest and its interaction term with the treatment arm. Within-subgroup treatment effects and the interaction effect between subgroup and treatment arm were estimated with 95% CIs and *p*-values.

For the QALYs end point, a linear regression model was fitted to the area under the utility curve, with baseline EQ-5D-3L score, surgeon, surgical procedure, treatment arm, subgroup variable and the subgroup-by-treatment interaction variable.

Because analysis revealed an increasing OR for the maze procedure arm relative to the control arm as the trial progressed, we explored changes in baseline characteristics, surgery and cointerventions throughout the trial in an attempt to explain this finding.

Secondary end point analysis

Return to stable SR at 24 months was analysed in a similar way to return to SR at 12 months, using a binary logistic regression model, including baseline heart rhythm, surgeon, surgical procedure and treatment arm.

Overall survival was summarised using Kaplan–Meier methods for the time between randomisation and death. Patients who were alive at the end of the study, or who withdrew before the end of follow-up, were censored at the date they were last seen. Similarly, stroke-free survival, defined as the time between randomisation and the date of stroke or death, whichever occurred first, was summarised using Kaplan–Meier methods. Patients who were alive and stroke free at the end of the study, or who withdrew without having suffered a stroke before the end of follow-up, were censored at the date they were last seen. Cox regression models were used to estimate hazard ratios (HRs) for maze procedure patients relative to control patients.

Patients who had a stroke within 12 months of surgery, and the overall proportion of stroke events, were calculated by treatment arm, using the total number of patients participating in the trial as the denominator. The relationship between stroke and treatment arm was tested by Fisher's exact test, and the difference in stroke rates was reported along with 95% CIs for differences in proportions.

Short Form questionnaire-36 items dimension scores and summarised mental component score (MCS) and physical component score (PCS) were analysed using a linear regression model, including time point, treatment arm, time-by-treatment-arm interaction and baseline SF-36 scores (all modelled as fixed effects), and allowing random intercepts for patients.

Drug use for each arm was tabulated by time point (at baseline, discharge, 6 weeks and 6, 12 and 24 months) and drug category. Logistic regression for the outcome of each patient (1 = have one or more drugs during time period *t*, 0 = have no drugs during time period *t*) was fitted, including drug category, time period of drug usage, baseline drug usage and treatment arm as independent variables. ORs were estimated with 95% CIs and *p*-values.

The occurrence of atrial flutter and atrial tachycardia (organised atrial arrhythmia) and junctional rhythm were summarised by arm. The relationship between the completeness of the lesion set and the occurrences of organised atrial arrhythmia and junctional rhythm was tabulated.

Safety analysis

The number of AEs in each category and deaths from any cause were summarised by treatment arm, corresponding to the treatment received. Events were summarised according to whether or not they met the criteria of SAEs, severity and relationship to the procedure.

Economic analysis

Data collection and sources

NHS resource use was collected during primary admission and at the 6-week and 6-, 12- and 24-month follow-ups. Research nurses/clinical trial co-ordinators extracted data about inpatient stay from individual patient records and administered bespoke questionnaires about follow-up health service use either face to face or by telephone/post (for those who missed an appointment). Hospital records were checked to validate patient-reported hospital readmission.

Resources related to the primary admission (from randomisation to discharge) included theatre use (initial operation and returns to theatre), intensive care (days) and cardiac and acute care wards (days). The total length of stay was compared with the sum of recorded days in an intensive care unit (ICU) and ward, and any double-counting that was identified was subtracted. For patients who were not discharged home, subsequent admissions to rehabilitation centres or acute hospitals were added. Surgery-specific resource use, including equipment and energy sources for the maze procedure, were retrieved from patient notes.

The resource use recorded during follow-up was divided into three categories: hospital readmissions (length of stay in hospitals or rehabilitation centres), 16 types of test [e.g. cardiac related, magnetic resonance imaging (MRI) and radiographic] and 12 types of health-care visit [e.g. accident and emergency (A&E), outpatient, primary and community health services]. In all resource-use calculations, a value of zero was assigned to any unused resource item, including all resource items after death.

Medication use was limited to antiarrhythmic, anticoagulant and antiplatelet drugs and seven classes of cardiac drugs (beta-blockers, diuretics, calcium channel blockers, nitrates, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers and statins). The daily dose of each drug type was based on patient record review for inpatient stay and patient reports after discharge. As specific drug names were not recorded for cardiac drugs, the chief investigator (SN) identified the most likely drugs to be prescribed, and the daily dose was taken to be the most common reported daily dose in the CRFs for that category of drugs. For example, bisoprolol and atenolol were the assumed drugs when beta-blocker was indicated, with the dose equalling that reported by 90% of respondents.

The total amount of each drug used per patient was estimated by taking a mid-value of daily dose at consecutive follow-up time points, multiplied by duration between follow-up time points, and costed using the *NHS Prescription Services Electronic Drug Tariff*[56] and the *British National Formulary* (2016).[57] Missing drug use at each time point was replaced, depending on the nature of missingness, as follows: patients who died were assigned zero medication use from date of death; if a patient indicated use of a drug at only one follow-up, the duration was taken to be the mid-point between this and the next follow-up

(except for drugs recorded only at baseline, which were excluded, as the trial focused on drugs post randomisation). For patients who either had completely missing drug data at a follow-up time point or were lost to follow-up, costs were multiply imputed using chained equations with predictive mean matching, stratified by treatment arm (see *Missing data*).

Unit costs were multiplied by the frequency of resource use to provide total resource cost for each item. National estimates of unit prices were sourced[58,59] to increase generalisability. For resources for which national prices were not available, estimates were sourced either from the literature (e.g. 24-hour blood pressure monitoring and chest radiography) or from Papworth Hospital (e.g. theatre cost and cost of device). The hospital and community health services pay and price index[58] was applied to adjust for inflation when necessary (see *Appendix 2*, *Table 23*). The ablation device was costed at £3000 per patient for high-intensity focused ultrasound, and £1250 per patient for all other methods. All resource costs, from the date of operation (randomisation) up to 2 years post randomisation, were summed, with year 2 costs discounted at 3.5%.[45]

Health-related quality of life, assessed using the EQ-5D-3L and SF-36 questionnaires, was an important outcome, and is described in *Secondary end point analysis*. SF-36 health state responses were converted to the Short Form questionnaire-6 Dimensions (SF-6D) utility scale using values from the UK population.[47] QALYs, as described in *Statistical analyses* on page 12, were discounted at 3.5% in year 2 for the cost-effectiveness analysis.[45]

Missing data

Missing baseline variables that were required for the imputation model, for example missing baseline EQ-5D-3L assessments, were replaced by the mean value for each trial arm.[60] Logistic regression identified variables that were related to missingness.

Missing resource use and utility data were imputed jointly using chained equations with predictive mean matching. The imputation models included age, sex, paroxysmal AF and baseline EQ-5D-3L score, and were stratified by trial arm. A total of 60 imputed data sets were created to attain a stable imputation. The distribution of imputed values was checked for comparability with observed data [e.g. counts of general practitioner (GP) visits, matched observations].

For 28 resource-use variables (tests and health-care visits), multiple imputation at each data collection point was not possible, as a result of the small numbers of events, and, therefore, the annual average for each resource-use variable was imputed for each arm. To assess the sensitivity of results to this assumption, an alternative imputation model was fitted, in which each test and health-care visit was multiplied by the corresponding unit cost. All tests and (separately) all health-care visits were grouped for each trial data collection point using total cost, and multiple imputation was applied to these categories of resource-use cost.

Incremental cost-effectiveness analysis and sensitivity analyses

Differences in estimated costs and QALYs between trial arms were explored using two-sample *t*-tests with equal variances. Linear regression analysis was used to adjust for differences in age, sex, baseline EQ-5D-3L score, AF at baseline and, for QALYs only, the primary surgery [isolated mitral valve replacement or repair (MVR), isolated CABG, isolated aortic valve replacement or repair (AVR), CABG and MVR, CABG and AVR and all others]. The incremental cost-effectiveness ratio (ICER) was calculated using adjusted mean estimates of costs and QALYs from 'seemingly unrelated regression', to allow for correlation between costs and effects at the patient level, and for skewness of data.

One thousand bootstraps were generated for each sample for the probabilistic sensitivity analysis. Costs for the resource-use components were sampled from gamma distributions and applied to the bootstrapped samples, and the total costs and QALYs for each sample were estimated using seemingly unrelated regression. The probability that the maze procedure was cost-effective was considered at varying willingness-to-pay (WTP)

threshold values, using cost-effectiveness planes, the cost-effectiveness acceptability curve and incremental net monetary benefit (INMB).

Deterministic and probabilistic sensitivity analyses were used to explore the robustness of cost-effectiveness results that adopted different methodological approaches or assumptions. These analyses included the use of SF-6D QALYs, clinical effectiveness as measured by conversion of AF to SR, complete case analysis, examining costs and QALYs only up to discharge, examining the impact of outliers, excluding maze device cost, limiting the patient group to those randomised from April 2001 (to match the time-based post hoc statistical analysis) and an alternative imputation technique. The probability distribution of unit costs could not be resampled when the alternative imputation technique was used, as the imputation was for cost at each follow-up point, rather than resource use.

Chapter 3 Trial results: clinical effectiveness and health-related quality of life

Recruitment and compliance

Between 25 February 2009 and 6 March 2014, 1013 patients were screened for the Amaze trial in 11 UK specialist cardiac surgery centres: (1) Papworth Hospital, Cambridgeshire ($n = 546$); (2) Glenfield Hospital, Leicester ($n = 186$); (3) Derriford Hospital, Plymouth ($n = 95$); (4) Freeman Hospital, Newcastle ($n = 72$); (5) Northern General Hospital, Sheffield ($n = 49$); (6) Blackpool Victoria Hospital, Blackpool ($n = 27$); (7) Royal Brompton Hospital, London ($n = 16$); (8) Guy's and St Thomas' NHS Foundation Trust, London ($n = 13$); (9) Wythenshawe Hospital, Manchester ($n = 10$); (10) University Hospitals Coventry and Warwickshire NHS Trust, Coventry ($n = 4$); and (11) Brighton and Sussex University Hospitals, Brighton ($n = 3$). The flow of Amaze trial patients from the initial screening to final follow-up is illustrated in *Figure 2*.

A total of 661 patients were excluded at screening, but screening logs were only completed for all patients in the co-ordinating centre (Papworth Hospital). At Papworth Hospital, 366 out of 546 patients (67%) were excluded between registration and randomisation (see *Appendix 3*, *Table 32*); 107 of these patients declined to participate, mostly because of concerns about either the trial requirements or the planned surgery they were about to undergo. Only one patient cited concerns about not knowing the treatment arm until 2 years after the procedure (as a result of patient blinding), although 38 patients declined without giving a reason. A further 115 patients were excluded by the consultant surgeon, with 49 (43%) of these patients undergoing the maze procedure outside the trial (as a result of severe or symptomatic AF or patient preference) and 11 patients opting for minimally invasive access or another procedure. The reason for exclusion by the surgeon was not recorded for 23 cases. Other reasons were related to trial exclusion criteria, such as patient participation in other clinical trials ($n = 19$), lack of time to recruit some in-house urgent cases ($n = 6$), not having a well-documented history of AF ($n = 5$) and having previous cardiac surgery ($n = 3$). Between screening and randomisation, eight patients died, and for a further four, their conditions deteriorated to such an extent that trial participation was not considered appropriate. A further 42 patients were excluded for various administrative reasons (see *Appendix 3*, *Table 32*).

After exclusions, 352 patients were randomised to either the planned cardiac procedure alone ($n = 176$) or maze procedure in addition to the planned procedure ($n = 176$). Thirteen patients (3.7%) did not receive their allocated treatment: 11 (6.3%) maze and two (1.1%) control patients. The maze procedure was not completed for a number of patients, as a result of (1) operation complexity and concern about prolonged cross-clamp time ($n = 4$); (2) an enlarged atrium or other technical difficulty ($n = 3$); (3) patient withdrawal from surgery after randomisation was revealed ($n = 1$); and (4) unrecorded surgeon decision ($n = 3$). Two control patients had the maze procedure as a result of perceived patient benefit by the consultant post randomisation.

Complete blinding was maintained for 339 (96%) patients. Treatment allocation was revealed in the notes of 13 patients (nine at Papworth Hospital, three at Derriford Hospital and one at Wythenshawe Hospital); of these, 10 underwent the maze procedure and three were control patients. The unblinding was attributable to initial protocol misunderstanding; after re-education of trial personnel, complete blinding was achieved for subsequent patients. The cardiologist reviewing the ECG recording did not have access to the patient's medical notes. All patients and HRQoL assessors remained unaware of treatment allocation.

At 12 months, 150 maze procedure patients and 151 control patients (85% and 86%, respectively) remained in the trial. The reasons recorded for loss to follow-up at 12 months were death in 33 cases, patient withdrawal in 14 cases and loss to follow-up in four cases. Note that one maze procedure patient

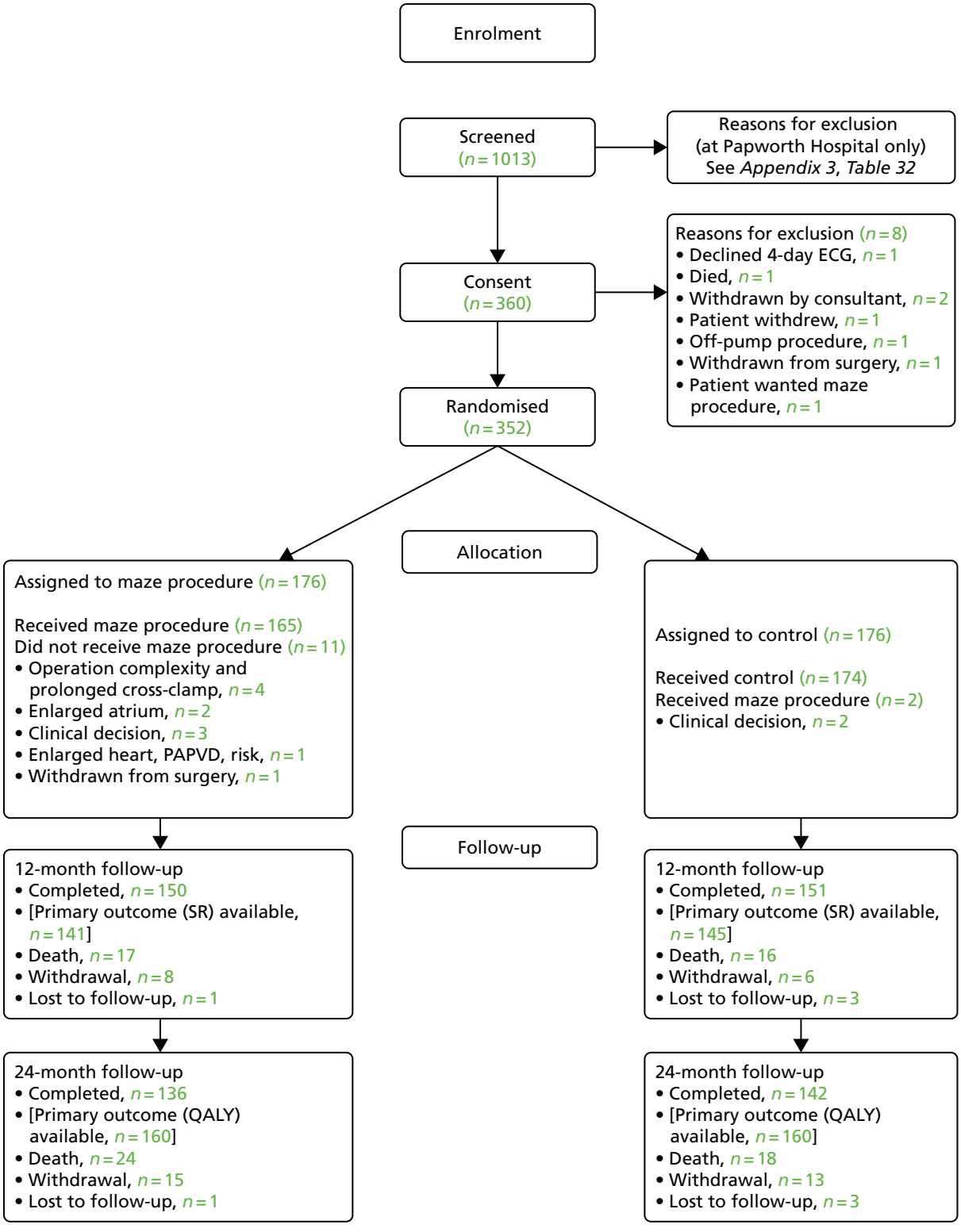


FIGURE 2 Patient flow through the Amaze trial. PAPVD, partial anomalous pulmonary venous drainage.

(out of 33) died just after 12 months, but the patient was too sick to complete follow-up. The clinical primary end point (SR at 12 months post randomisation) was completed for 141 (80%) maze procedure patients and 145 (82%) control patients, as 11 patients declined the 4-day ECG (eight maze procedure patients and three control patients) and, for four patients, the recordings were not usable (one maze procedure patient and three control patients). The frequency of missing outcomes and associated reasons was similar for the two trial arms.

The patient-based primary end point (QALYs up to 24 months post randomisation) was completed for 160 patients in each arm (91%). We note that patients who died during follow-up were included in the calculation of QALYs, contributing zero to the estimate from the date of death. Thirty-two patients were excluded from this analysis, as a result of either patient withdrawal from the study ($n = 28$) or loss to follow-up ($n = 4$).

Baseline characteristics

Patient characteristics at baseline are shown in *Table 2*. Almost 50% of cases were recruited in the co-ordinating centre (Pathworth Hospital) by 13 surgeons, and over one-quarter were recruited in the second highest recruiting centre (Glenfield Hospital) by four surgeons. The Amaze trial population had a mean age of 71.9 years (SD 7.67 years), almost two-thirds (65.9%) were men and the mean risk of

TABLE 2 Baseline characteristics for patients randomised in the Amaze trial

Characteristic	Treatment arm		
	Maze procedure ($n = 176$)	Control ($n = 176$)	Total ($n = 352$)
Number of patients at each randomising centre, *n* (%)			
Papworth Hospital	89 (50.6)	85 (48.3)	174 (49.4)
Glenfield Hospital	49 (27.8)	44 (25.0)	93 (26.4)
Derriford Hospital	16 (9.1)	16 (9.1)	32 (9.1)
Northern General Hospital, Sheffield	12 (6.8)	14 (8.0)	26 (7.4)
Freeman Hospital, Newcastle upon Tyne	3 (1.7)	5 (2.8)	8 (2.3)
Guy's and St Thomas' Hospital, London	1 (0.6)	5 (2.8)	6 (1.7)
Wythenshawe Hospital	4 (2.3)	1 (0.6)	5 (1.4)
Brighton and Sussex University Hospitals	1 (0.6)	1 (0.6)	2 (0.6)
University Hospitals Coventry and Warwickshire NHS Trust	1 (0.6)	1 (0.6)	2 (0.6)
Royal Brompton Hospital	–	2 (1.1)	2 (0.6)
Blackpool Victoria Hospital	–	2 (1.1)	2 (0.6)
Patient age (years)			
Mean (SD)	72.3 (7.53)	71.4 (7.81)	71.9 (7.67)
Range	50.0–86.0	48.0–89.0	48.0–89.0
Patient sex, *n* (%)			
Male	112 (63.6)	120 (68.2)	232 (65.9)
Female	64 (36.4)	56 (31.8)	120 (34.1)
Body mass index (kg/m^2)			
Mean (SD)	28.1 (5.27)	27.6 (4.62)	27.9 (4.96)
Range	17.4–46.0	17.9–42.8	17.4–46.0
Logistic EuroSCORE[61] (%)[a]			
Mean score (SD)	6.94 (5.489)	6.64 (4.869)	6.79 (5.184)
Range	0.88–30.41	1.40–23.85	0.88–30.41

EuroSCORE, European System for Cardiac Operative Risk Evaluation.
a EuroSCORE was not recorded for one control patient.

in-hospital death as a result of the procedure [the 2003 logistic European System for Cardiac Operative Risk Evaluation (EuroSCORE)[61]] was 6.79% (SD 5.18%). The characteristics of Amaze patients were broadly similar to those of UK NHS cardiac surgery patients, but Amaze patients were slightly older and more likely to be female, and had a slightly lower average EuroSCORE.[62] On average, the two treatment arms had similar characteristics.

Table 3 summarises symptoms at baseline. Heart failure symptoms, defined by the NYHA classification, were common, with 40.3% of patients reporting mild symptoms or slight limitations during ordinary activity, and 41.4% of patients reporting either marked or severe limitations, even during mild activity or at rest. Symptoms of angina, as defined by the Canadian Cardiovascular Society's grading scale for angina pectoris, were less common, with 73.3% of patients being angina free at baseline and only a small proportion (5.7%) reporting moderate or severe limitations as a result of angina.

Other markers of cardiac function were also similar between the two arms (*Table 4*). For example, approximately two-thirds of patients (66.5%) had a left ventricular ejection fraction (LVEF) of > 50% at baseline, and 2.6% of patients had suffered a recent myocardial infarction (MI). The frequency of other risk factors for heart disease was similar between the two arms: 3.4% of patients were insulin-dependent diabetics, 12.5% were non-insulin-dependent diabetics, 37.8% were treated for high cholesterol and 57.4% had hypertension. The frequency of comorbidities was similar in both treatment arms, with chronic obstructive pulmonary disease (COPD) present in 9.7% of patients and pulmonary hypertension present in 15.1% of patients. In both treatment arms, 6.3% of patients had a history of cerebrovascular accidents and 8.8% had previous transient ischaemic attacks, with 2.8% of these patients having neurological dysfunction at baseline (see *Appendix 3*, *Table 33*).

TABLE 3 Symptoms of heart failure at baseline

Classification	Treatment arm, *n* (%)		Total number of patients (*n* = 352), *n* (%)
	Maze procedure (*n* = 176)	Control (*n* = 176)	
CCS classification			
Class 0	125 (71.0)	133 (75.6)	258 (73.3)
Class 1	13 (7.4)	17 (9.7)	30 (8.5)
Class 2	21 (11.9)	16 (9.1)	37 (10.5)
Class 3	10 (5.7)	8 (4.5)	18 (5.1)
Class 4	1 (0.6)	1 (0.6)	2 (0.6)
Missing/not known	6 (3.4)	1 (0.6)	7 (2.0)
NYHA classification at baseline			
I	31 (17.6)	30 (17.0)	61 (17.3)
II	74 (42.0)	68 (38.6)	142 (40.3)
III	59 (33.5)	71 (40.3)	130 (36.9)
IV	10 (5.7)	6 (3.4)	16 (4.5)
Missing/not known	2 (1.1)	1 (0.6)	3 (0.9)

CCS, Canadian Cardiovascular Society's grading scale for angina pectoris.

TABLE 4 Markers of cardiac function and cardiovascular risk factors at baseline

Marker	Treatment arm, *n* (%)		Total number of patients (*n* = 352), *n* (%)
	Maze procedure (*n* = 176)	Control (*n* = 176)	
Left ventricular function			
Poor (LVEF of < 30%)	4 (2.3)	8 (4.5)	12 (3.4)
Moderate (LVEF of 30–50%)	50 (28.4)	56 (31.8)	106 (30.1)
Good (LVEF of > 50%)	122 (69.3)	112 (63.6)	234 (66.5)
Recent MI	4 (2.3)	5 (2.8)	9 (2.6)
Previous PCI	16 (9.1)	14 (8.0)	30 (8.5)
Congestive cardiac failure	5 (2.8)	1 (0.6)	6 (1.7)
Diabetes			
Insulin dependent	5 (2.8)	7 (4.0)	12 (3.4)
Non-insulin dependent	27 (15.3)	17 (9.7)	44 (12.5)
Hyperlipidaemia/hypercholesterolaemia	70 (39.8)	63 (35.8)	133 (37.8)
Systemic hypertension	103 (58.5)	99 (56.3)	202 (57.4)

LVEF, left ventricular ejection fraction; MI, myocardial infarction; PCI, percutaneous coronary intervention.

Table 5 documents patients' medical history associated with AF. For 26.1% of patients, AF was paroxysmal; the 73.9% of patients who had non-paroxysmal AF included almost 60% of patients classed as having chronic/longstanding AF and 13.9% classed as having persistent intermittent symptoms. Over two-thirds (68.5%) of patients had AF for > 12 months. Only 4.3% of patients had been fitted with a permanent pacemaker and 13.4% had previously undergone cardioversions; previous ablation had been attempted in slightly more maze procedure patients (1.7%) than control patients (0.6%). Anticoagulant and antiarrhythmic drugs were prescribed for 77.6% and 83.2% of patients, respectively.

Table 6 summarises the HRQoL for the EQ-5D-3L utility score and the SF-36 dimensions at baseline. The mean EQ-5D-3L utility score was 0.75 (SD 0.22) at baseline, which compares well with the UK norms of 0.78 (SD 0.26) for people aged 65–74 years and 0.73 (SD 0.27) for people aged ≥ 75 years.[63] Thus, patients selected for cardiac surgery, who entered the Amaze trial, have comparable limitations to the general population of the same age, as measured by this generic HRQoL scale. In contrast, the mean scores for the SF-36 dimensions were very much lower than the published norms at baseline, particularly for the physical dimensions (see *Table 6*).[64] The mean standardised MCS at baseline was 50.19 (SD 10.32) for this population, almost exactly the same as the mean score for the UK population, whereas the mean for the standardised PCS was 30.59 (SD 13.36), which is significantly lower than the mean (SD) score for the UK population.

Surgical results

Table 7 summarises the surgical procedures completed; the most common were isolated MVR (24.7%), CABG (19.6%) and AVR (15.6%), followed by combined CABG with either AVR (10.5%) or MVR (7.7%). All other procedures were combinations of CABG and/or multiple valve procedures, with the exception of two patients for whom no procedure could be completed.

TABLE 5 Atrial fibrillation-related clinical history at baseline

Marker of AF	Treatment arm, *n* (%)		
	Maze procedure (*n* = 176)	Control (*n* = 176)	Total (*n* = 352), *n* (%)
AF classification			
Paroxysmal (intermittent)	44 (25.0)	48 (27.3)	92 (26.1)
Persistent (intermittent)	30 (17.0)	19 (10.8)	49 (13.9)
Chronic/longstanding (continuous)	102 (58.0)	109 (61.9)	211 (59.9)
AF-related medical history			
0–3 months ago	4 (2.3)	2 (1.1)	6 (1.7)
3–6 months ago	25 (14.2)	25 (14.2)	50 (14.2)
6–12 months ago	31 (17.6)	23 (13.1)	54 (15.3)
> 12 months ago	115 (65.3)	126 (71.6)	241 (68.5)
Not known	1 (0.6)	–	1 (0.3)
Permanent pacemaker	7 (4.0)	8 (4.5)	15 (4.3)
Time of pacemaker implant			
0–3 months ago	–	2 (1.1)	2 (0.6)
3–6 months ago	2 (1.1)	–	2 (0.6)
6–12 months ago	2 (1.1)	2 (1.1)	4 (1.1)
> 12 months ago	3 (1.7)	4 (2.3)	7 (2.0)
Previous cardioversions			
0–3 months ago	1 (0.6)	1 (0.6)	2 (0.6)
3–6 months ago	1 (0.6)	–	1 (0.3)
6–12 months ago	5 (2.8)	3 (1.7)	8 (2.3)
> 12 months ago	17 (9.7)	19 (10.8)	36 (10.2)
Previous ablation	3 (1.7)	1 (0.6)	4 (1.1)
Arrhythmias other than AF/flutter	2 (1.1)	2 (1.1)	4 (1.1)
Any anticoagulant use at baseline	137 (77.8)	137 (77.3)	274 (77.6)
Any antiarrhythmic use at baseline	145 (82.4)	148 (84.1)	293 (83.2)

Descriptions of surgical indices are given in *Table 8*. As expected, the time spent in theatre was longer for the maze procedure arm; the difference (maze procedure vs. control) in the mean length of time spent in theatre was 13.8 minutes (95% CI –4.4 to 32.0 minutes; $p = 0.1375$). Similarly, there was a mean difference in the time taken for cross-clamp of 5.1 minutes (95% CI –4.0 to 14.2 minutes; $p = 0.2725$) and in the time taken for cardiopulmonary bypass of 18.9 minutes (95% CI 9.9 to 27.8 minutes; $p < 0.0001$). Note that three patients' surgical procedures were completed with a beating heart (with one patient randomised to the maze procedure arm and two patients randomised to the control arm), so that the time taken for both cross-clamp and cardiopulmonary bypass was zero minutes; on these occasions, the maze procedure was not performed. One more maze procedure patient had zero minutes recorded for cross-clamp, but 205 minutes recorded for the time taken for cardiopulmonary bypass.

TABLE 6 The mean scores for the EQ-5D-3L utility and SF-36 dimensions at baseline[a]

HRQoL measurement	Treatment arm, mean score (SD)		Total, mean score (SD) ($n = 352$)	UK norm, mean score (SD)
	Maze procedure ($n = 176$)	Control ($n = 176$)		
EQ-5D-3L utility score	0.74 (0.22)	0.75 (0.21)	0.75 (0.22)	–
SF-36 dimensions				
Bodily pain	72.00 (28.47)	72.26 (26.18)	72.13 (27.30)	81.49 (21.69)
General health	57.22 (19.11)	55.61 (20.76)	56.41 (19.94)	73.52 (19.90)
Physical function	47.18 (26.16)	48.40 (27.33)	47.79 (26.72)	88.40 (17.98)
Role emotional	71.10 (41.76)	65.90 (45.61)	68.49 (43.75)	82.93 (31.76)
Role physical	27.75 (37.20)	30.57 (40.84)	29.17 (39.04)	85.82 (29.93)
Social functioning	64.73 (29.19)	64.44 (31.81)	64.58 (30.49)	88.01 (19.58)
Vitality	43.67 (21.73)	44.71 (23.76)	44.19 (22.74)	61.13 (19.67)
Mental health	75.24 (15.44)	73.51 (18.21)	74.37 (16.88)	73.77 (17.24)
PCS	30.18 (13.17)	31.00 (13.56)	30.59 (13.36)	50 (10)
MCS	50.81 (9.92)	49.58 (10.69)	50.19 (10.32)	50 (10)

a Three maze procedure patients and two control patients had missing baseline QoL data.

TABLE 7 Cardiac procedure completed

Procedure	Treatment arm, *n* (%)		Total ($n = 352$), *n* (%)
	Maze procedure ($n = 176$)	Control ($n = 176$)	
Actual procedure category			
MVR	39 (22.2)	48 (27.3)	87 (24.7)
CABG	35 (19.9)	34 (19.3)	69 (19.6)
AVR	32 (18.2)	23 (13.1)	55 (15.6)
CABG and AVR	16 (9.1)	21 (11.9)	37 (10.5)
CABG and MVR	14 (8.0)	13 (7.4)	27 (7.7)
All other procedures, including none	40 (22.7)	37 (21.0)	77 (21.9)

The left atrial appendage was also significantly more likely to be excised in the maze procedure arm (55.1%) than in the control arm (30.1%).

Table 9 provides details of the lesion sets completed. Eleven patients in the maze procedure arm did not have the adjunct procedure at all. The most common ablation procedure was applied to the left and right atria and the mitral annulus (43.8%), with 22.2% applied to the left atrium and mitral annulus and 18.2% applied to the left atrium only. The mean number of lesions was 6.5 (SD 3.55) in the maze procedure arm, with 47.7% of patients having 5–9 lesions and 23.3% of patients having ≥ 10 lesions. The most common mode of delivery was bipolar radiofrequency ablation (81.8%), with unipolar radiofrequency ablation, cryotherapy and ultrasound applied to smaller numbers of maze procedures; no procedures applied laser or microwave energy.

TABLE 8 Summary of theatre times and excision of left atrial appendage

Procedure characteristic	Treatment arm		Total ($n = 352$)
	Maze procedure ($n = 176$)	Control ($n = 176$)	
Total length of time (minutes) taken for cross-clamp			
Mean (SD)	82.2 (37.25)	77.2 (48.60)	79.7 (43.31)
Median (quartiles)	74.0 (57.5–102.0)	67.5 (51.0–92.0)	72.0 (53.0–99.0)
Range	0.0–245.0	0.0–530.0	0.0–530.0
Total length of time (minutes) taken for cardiopulmonary bypass			
Mean (SD)	118.1 (43.39)	99.3 (41.81)	108.7 (43.59)
Median (quartiles)	110.5 (84.0–145.0)	93.0 (72.5–120.0)	100.5 (80.0–132.0)
Range	0.0–342.0	0.0–300.0	0.0–342.0
Total length of time (minutes) spent in theatre			
Mean (SD)	261.2 (79.68)	247.5 (93.27)	254.4 (86.89)
Median (quartiles)	260.0 (210.0–300.0)	218.0 (195.0–277.5)	240.0 (198.0–291.5)
Range	75.0–582.0	100.0–775.0	75.0–775.0
Excised left atrial appendage, *n* (%)			
Yes	97 (55.1)	53 (30.1)	150 (42.6)
No	79 (44.9)	123 (69.9)	202 (57.4)

TABLE 9 Details of number and location of lesion sets that were operated on

Lesion sets	Treatment arm		Total ($n = 352$)
	Maze procedure ($n = 176$)	Control ($n = 176$)	
Number of lesions treated			
Mean (SD)	6.5 (3.55)	0.1 (0.91)	3.3 (4.11)
Median (quartiles)	7.0 (4.0–9.0)	0.0 (0.0–0.0)	0.0 (0.0–7.0)
Range	0.0–14.0	0.0–11.0	0.0–14.0
Lesion number category, *n* (%)			
0	11 (6.3)	174 (98.9)	225 (63.9)
1–4	40 (22.7)	–	40 (22.7)
5–9	84 (47.7)	1 (0.6)	85 (24.1)
≥ 10	41 (23.3)	1 (0.6)	42 (11.9)
Lesion set treated, *n* (%)			
I: minimal left atrial lesion set: pulmonary vein isolation either with or without left atrial appendage line	32 (18.2)	–	32 (9.1)
II: more extensive left atrial lesion set, excluding mitral annulus	4 (2.3)	–	4 (1.1)
III: more extensive left atrial only lesion set, including mitral annulus	39 (22.2)	1 (0.6)	40 (11.4)
IV: minimal left atrial lesion set and right atrial lesion set	2 (1.1)	–	2 (0.6)
V: more extensive left atrial lesion set excluding mitral annulus and right atrial lesion set	11 (6.3)	–	11 (3.1)
VI: more extensive left atrial lesion set including mitral annulus and right atrial lesion set	77 (43.8)	1 (0.6)	78 (22.2)
No lesions	11 (6.3)	174 (98.9)	185 (52.6)

At least one perioperative complication was recorded for 34 (19.3%) maze procedure patients and 38 (21.6%) control patients (see *Appendix 3*, *Table 34*). As expected, the most common complication was bleeding (for 10.8% of maze procedure patients and 9.7% of control patients) and pleural effusion (for 8% of maze procedure patients and 13.6% of control patients). There were no important differences in the number of patients requiring transfusion of red blood cells, platelets, fresh-frozen plasma, cryoprecipitate or human albumin (see *Appendix 3*, *Table 35*).

Intensive care unit and hospital stay did not vary by treatment arm. The median (quartiles) duration of stay in an ICU was 1.1 days (0.9–2.9 days) in the maze procedure arm and 1.0 days (0.9–2.0 days) in the control arm, whereas the median (quartiles) total length of hospital stay was 9 days (7–13 days) and 8 days (6–12 days) for the maze procedure and control arms, respectively. Eleven maze procedure patients and 12 control patients returned to an ICU on one or more occasions, with those returning having a median (quartiles) total length of stay of 4.6 days (1.3–6.5 days) and 2.5 days (1.5–10.4 days) in the maze procedure and control arms, respectively.

Primary outcome results

Sinus rhythm at 12 months

Despite a history of AF, 30 patients (17.0%) in the maze arm and 32 patients (18.2%) in the control arm did not have any arrhythmias recorded by the 4-day ECG at baseline. At 12 months, 286 (81.3%) patients completed the 4-day ECG recording; of these patients, 266 (93.0%) were either in SR 100% of the time or in AF 100% of the time. Patients were classified as being in AF if any AF was observed during the 4-day ECG recording; this was decided before linking trial outcomes to either treatment arm. Among complete cases in the maze procedure arm, 87 out of 141 patients (61.7%) were in SR compared with 68 out of 145 (46.9%) control patients. In the ITT analysis, using multiple imputation of missing data, the OR for return to SR was 2.06 (95% CI 1.20 to 3.54; $p = 0.0091$); see *Appendix 3*, *Table 36* for the full results. Overall, results varied substantially by surgeon, with an associated intraclass correlation coefficient (ICC) of 0.089, suggesting that 8.9% of the total variation in return to SR rates over both treatment arms resulted from surgeon effects. However, there were no differences in the treatment effect (maze procedure vs. control) among surgeons (the ICC on the treatment coefficient was zero). *Table 10* shows that the difference between the treatment arms arises almost solely from an additional 19 patients in the maze procedure arm changing from having AF to SR (61 vs. 42) and 21 fewer patients remaining in AF at 12 months (50 vs. 71).

Sensitivity analysis

An exploratory analysis of the ITT population highlighted four outlying patients in accordance with cross-validation probabilities of having undue influence; a secondary analysis excluding these patients was conducted, with little change in the results (OR 2.20, 95% CI 1.27 to 3.82). In the sensitivity analyses, the OR changed to 2.00 (95% CI 1.21 to 3.32) if only complete cases were included, 1.70 (95% CI 1.07 to 2.69) if patients who died or withdrew were assumed to have been in AF, 1.92 (95% CI 1.17 to 3.15) if patients who died were assumed to have been in AF and 1.75 (95% CI 1.10 to 2.79) using the last observation carried forward; all remained statistically significant. An additional analysis, including the variables that were most associated with a missing status for 12-month SR (baseline SR, sex, diabetes, left ventricular function, history of rheumatic fever, non-AF/atrial flutter arrhythmias and COPD), resulted in a treatment effect estimate of 2.13 (95% CI 1.27 to 3.59). As the treatment effect was significant in all sensitivity analyses, we were confident that it was robust. Finally, in order to assess the treatment effect in patients for whom the surgery was completed as planned, the complier average causal effect for the difference in 12-month SR rates between the groups (maze procedure vs. control) was calculated; this was 15.8% (95% CI 3.9% to 27.6%) compared with 14.8% (95% CI 3.2% to 26.3%) for completers.

TABLE 10 Summary of ECG rhythm changes between baseline and follow-up

Rhythm changes	Treatment arm, *n* (%)		
	Maze procedure (*n* = 176)	Control (*n* = 176)	Total (*n* = 352), *n* (%)
Change from baseline to 12 months in AF/SR			
Either baseline or follow-up missing	39 (22.2)	33 (18.8)	72 (20.5)
SR at baseline, SR at 12 months	23 (13.1)	25 (14.2)	48 (13.6)
SR at baseline, AF at 12 months	3 (1.7)	5 (2.8)	8 (2.3)
AF at baseline, SR at 12 months	61 (34.7)	42 (23.9)	103 (29.3)
AF at baseline, AF at 12 months	50 (28.4)	71 (40.3)	121 (34.4)
Change from baseline to 24 months in AF/SR			
Either baseline or follow-up missing	62 (35.2)	50 (28.4)	112 (31.8)
SR at baseline, SR at 24 months	18 (10.2)	23 (13.1)	41 (11.6)
SR at baseline, AF at 24 months	3 (1.7)	4 (2.3)	7 (2.0)
AF at baseline, SR at 24 months	47 (26.7)	23 (13.1)	70 (19.9)
AF at baseline, AF at 24 months	46 (26.1)	76 (43.2)	122 (34.7)

Note
SR is defined as an absence of any AF recorded during the 4-day monitoring period.

Subgroup analysis

Figure 3 shows the results of the subgroup analysis (see also *Appendix 3*, *Table 37*); no interactions were statistically significant. The odds on returning to SR were increased by the adjunct maze procedure for both paroxysmal and non-paroxysmal AF groups. The maze patients had increased ORs on return to SR irrespective of the planned procedure, although the small numbers of patients in each subgroup meant that the ORs varied widely and were not significant in most cases.

Within the maze procedure arm, there was little evidence that return to SR was associated with the number of lesions treated, although this may simply reflect the skill of the surgeon in identifying the areas to be treated (*Table 11*). Moreover, there was no evidence of variation in return to SR between different ablation techniques, although bipolar ablation was clearly the preferred technique for many surgeons.

In October 2013, the independent Data Monitoring and Ethics Committee (DMEC) concluded that the results did not look promising and the TSC should consider stopping the trial on the basis of futility. After full and serious consideration of the recommendations made by the DMEC, the TSC decided to recommend continuation of recruitment and follow-up for the following reasons: there were no safety concerns at that stage of the trial; > 80% of patients had already been recruited; review of accruing data suggested that the results may be more promising; the trial results had the potential to have a high impact, particularly if no treatment effect was found. However, the investigators decided that the final arbiters should be the HTA programme board, which concluded that, as there were no safety concerns, it would support the continuation of recruitment to the trial until the end of May 2014. This discussion was relayed to the study teams at all sites via the steering group and circulation of the minutes, and they remained supportive of the decision to continue.

As a result of these concerns, a post hoc exploratory analysis was undertaken to assess whether or not there were changes in effects over time. The analysis found that the adjusted OR for return to SR at 12 months increased from 1.6 (95% CI 0.6 to 4.0) for the first 120 patients (considered by the DMEC) to 2.9 (95% CI 0.9 to 9.6) for the final 71 patients randomised in the final year of recruitment (2013).

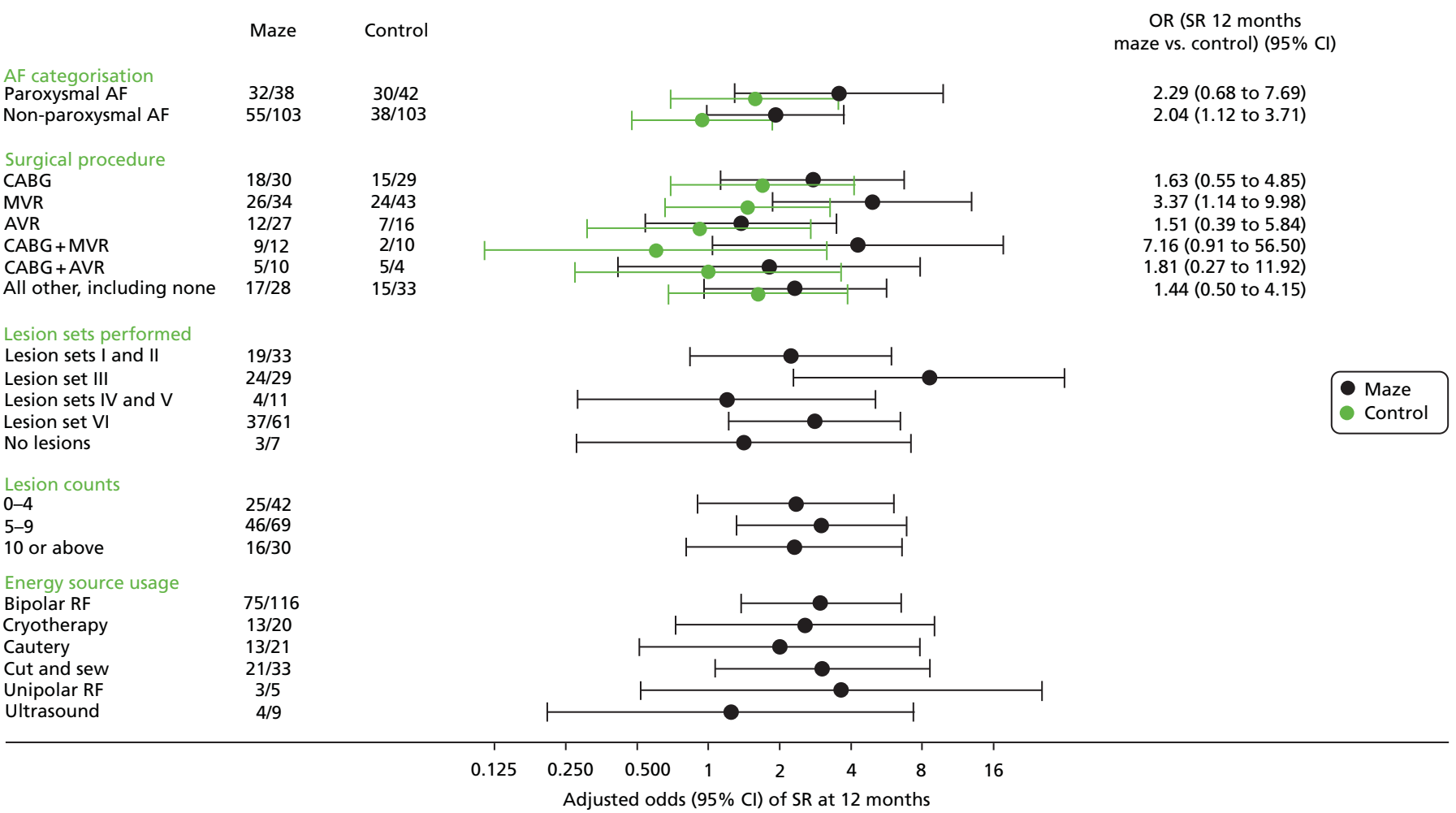


FIGURE 3 Forest plot showing (adjusted) odds on return to SR at 12 months after randomisation (details of lesion sets are given in *Table 9*). RF, radiofrequency.

TABLE 11 Odds and 95% CI of return to SR at 12 months in patients randomised to the maze procedure arm by subgroup (details of lesion sets are given in *Table 9*)

Subgroup	Number in SR/number in subgroup	Adjusted OR	95% CI
Lesion sets performed			
Lesion sets I and II	19/33	2.236	0.838 to 5.962
Lesion set III	24/29	8.585	2.295 to 32.109
Lesion sets IV and V	4/11	1.198	0.282 to 5.095
Lesion set VI	37/61	2.803	1.214 to 6.472
Lesion counts			
No lesions	3/7	1.412	0.279 to 7.142
0–4 lesions	25/42	2.337	0.903 to 6.050
5–9 lesions	46/69	3.009	1.320 to 6.861
≥ 10 lesions	16/30	2.312	0.813 to 6.577
Energy source usage			
Bipolar RF ablation	75/116	2.991	1.372 to 6.521
Cryotherapy	13/20	2.561	0.731 to 8.971
Cautery	13/21	1.998	0.513 to 7.789
Cut and sew	21/33	3.029	1.065 to 8.609
Unipolar RF ablation	3/5	3.631	0.517 to 25.473
Ultrasound	4/9	1.242	0.210 to 7.352

RF, radiofrequency.

Quality-adjusted life-years at 24 months

For this patient-centred primary outcome, QALYs could be estimated for 320 out of 352 patients (90.9%). The unadjusted and undiscounted mean QALYs over 2 years were 1.489 (95% CI 1.416 to 1.558) for the maze procedure arm and 1.485 (95% CI 1.403 to 1.559) for the control arm. In the primary complete-case, ITT analysis, adjusting for baseline covariates, the mean difference between the two arms (maze procedure vs. control) was –0.025 QALYs (95% CI –0.129 to 0.078 QALYs; $p = 0.6319$; see *Appendix 3*, *Table 38*). This difference corresponds to approximately 9 fewer days of life in perfect health for a maze procedure patient. A sensitivity analysis, which used different imputation methods for missing data and excluded any patients with outlying results, showed that these results were robust to model assumptions. Results did not vary substantially by surgeon; the average ICC across 40 multiple imputation samples was 0.001, indicating that only 0.1% of the total variation in 24-month QALYs was attributable to surgeon differences.

Sensitivity analysis

In a range of sensitivity analyses reflecting different assumptions about the missing data mechanism, the mean difference in QALYs changed only slightly, ranging from –0.029 (95% CI –0.135 to 0.078), for the last observation carried forward, to –0.010 (95% CI –0.119 to 0.100), for the analysis that adjusted for predictors of missingness (baseline EQ-5D-3L score, sex, diabetes and thoracic aorta surgery). Further details are available on request.

Subgroup analysis

The differences in QALYs at 2 years between the maze procedure and control arms are plotted in *Figure 4* for a range of subgroups (see also *Appendix 3*, *Table 39*). A number of subgroups had an estimated QALY difference above our predefined minimum clinically important difference of 0.083, in favour of the maze

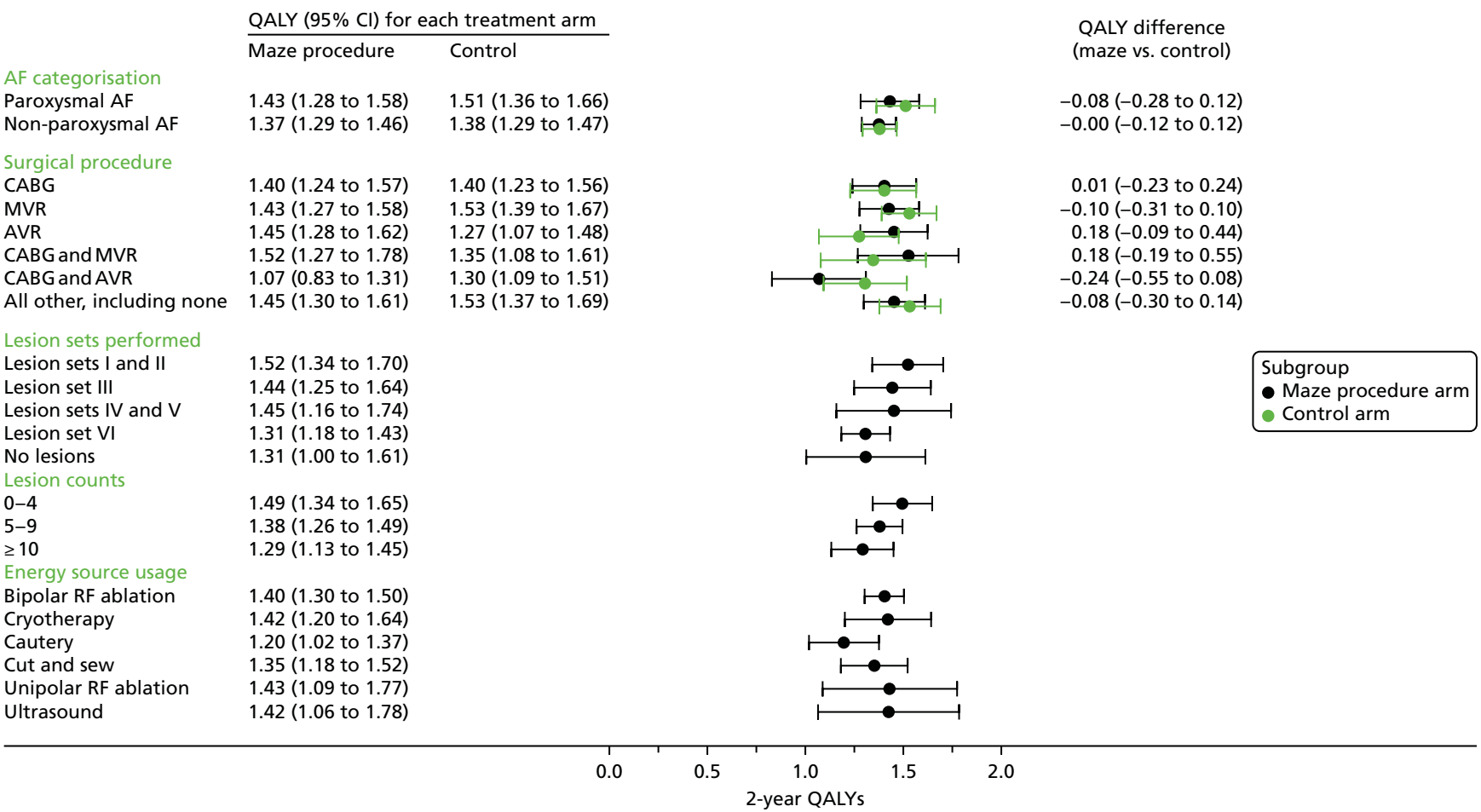


FIGURE 4 Forest plot showing the (adjusted) differences in QALYs at 2 years after randomisation for predefined subgroups (details of lesion sets are given in *Table 9*). RF, radiofrequency.

procedure arm (isolated AVR, CABG and MVR) or the control arm (isolated MVR, CABG and AVR), but no differences were statistically significant and these values probably reflect the variation observed in small subgroup analyses. Again, there were no apparent relationships between QALYs over 2 years and any of the following: lesion sets, number of lesions treated or ablation method (*Table 12*).

Secondary outcomes

Sinus rhythm at 24 months after surgery

At 24 months after surgery, 247 (70.2%) patients completed a 4-day ECG recording. In the maze procedure arm, 69 out of 118 (58.5%) completers were in SR compared with 47 out of 129 (36.4%) completers in the control arm. Thus, although the proportion of patients in SR decreased in both arms, the decrease was lower for the maze procedure arm. The baseline-adjusted OR for SR at 24 months was 3.24 (95% CI 1.76 to 5.96) in favour of the maze procedure arm (see *Appendix 3*, *Table 40* for the full model results). *Table 10* shows the number of people who had a change in SR between baseline and 24 months.

Survival and stroke-free survival

There were five (2.8%) postoperative deaths in the maze procedure arm and nine (5.1%) among the control patients (Fisher's exact test, $p = 0.4144$). This compares with the mean predicted in-hospital death rate of 6.79% (logistic EuroSCORE,[61] which is known to overestimate risk). Causes of death are listed in *Appendix 3*, *Table 41*. Between discharge and the planned 24-month follow-up date, there were 19 deaths in the maze procedure and nine in the control arm; note that three deaths in the maze procedure arm occurred just after the 24-month anniversary of their surgery.

TABLE 12 Two-year quality-adjusted survival among patients randomised to the maze procedure arm by subgroup (details of lesion sets are given in *Table 9*)

Level	Mean	SEM	95% CI
Lesion sets performed			
Lesion sets I and II	1.522	0.092	1.342 to 1.702
Lesion set III	1.443	0.100	1.247 to 1.639
Lesion sets IV and V	1.451	0.149	1.158 to 1.743
Lesion set VI	1.308	0.063	1.184 to 1.431
Lesion counts			
No lesions	1.308	0.155	1.005 to 1.611
0–4	1.494	0.078	1.341 to 1.646
5–9	1.377	0.060	1.260 to 1.494
≥ 10	1.290	0.082	1.130 to 1.450
Energy source usage			
Bipolar RF	1.402	0.052	1.301 to 1.503
Cryotherapy	1.420	0.112	1.200 to 1.640
Cautery	1.196	0.091	1.018 to 1.373
Cut and sew	1.351	0.087	1.181 to 1.521
Unipolar RF	1.429	0.175	1.086 to 1.772
Ultrasound	1.422	0.184	1.062 to 1.782

RF, radiofrequency; SEM, standard error of the mean.

Kaplan–Meier estimates of cumulative probability of death are plotted in *Figure 5*. This includes all 30 deaths in the maze procedure arm and all 25 deaths in the control arm at the end of the trial; the HR was 1.23 (95% CI 0.73 to 2.10; $p = 0.437$). Thus, the adjunct maze procedure did not significantly increase early or late death rates in this trial.

Figure 6 plots the cumulative incidence of death or stroke. During follow-up, 13 strokes were recorded in 10 (5.7%) maze procedure patients and 19 were recorded in 16 (9.1%) control patients; the difference of –3.4% (95% CI –14.1% to 7.3%) was not statistically significant (Fisher's exact test, $p = 0.3083$). Moreover, there was no significant difference in stroke-free survival between the two trial arms (HR 0.99, 95% CI 0.64 to 1.53; $p = 0.949$).

Anticoagulant and antiarrhythmic drug use

Table 13 shows that the number of patients requiring anticoagulant drug use was significantly lower in the maze procedure arm from 6 months after the procedure. Conversely, there were slightly more maze patients requiring antiarrhythmic drugs throughout follow-up, but the difference was not statistically significant at traditional levels.

Further cardioversions

There was no difference between the two treatment arms in the need for further cardioversion or permanent pacemaker implants. Sixty maze procedure patients (34.1%) required 65 cardioversions and 67 control patients (38.1%) required 72 cardioversions. The cardioversion success rates for these interventions were the same [48/65 (73.8%) for the maze arm and 54/72 (75.0%) for the control arm]. Fifteen maze procedure patients (8.6%) and 17 control patients (9.7%) required pacemaker implantation.

Additional results from the electrocardiogram recordings

At baseline, 27 maze procedure patients (15.3%) and 16 control patients (9.1%) had AF or tachycardia on at least 1 day of the ECG recordings. Corresponding numbers at 12 months were 29 (16.5%) and 26 (14.8%) for the maze procedure patients and control patients, respectively, which fell to 19 (10.8%) and 18 (10.2%), respectively, at 24 months. Junctional rhythm was observed for only one (maze procedure) patient at baseline, eight maze procedure patients at 12 months and six maze procedure patients and two control patients at 24 months.

Hospital admissions for haemorrhage

There were three admissions in three patients who had the maze procedure, and two admissions in two patients who did not have the maze procedure up to 2 years after randomisation.

New York Heart Association results

Figure 7 and *Appendix 3*, *Table 42* summarise the NYHA results for each treatment arm over the 2-year follow-up period. Among those who had complete data, there was some evidence that more patients in the maze procedure arm had symptoms of heart failure at 6 months after surgery (52.9% vs 42.7%). The difference was 10.2% (95% CI –1.4% to 21.5%; $p = 0.0995$), but the distribution of NYHA classes in each treatment arm was very similar thereafter.

Short Form questionnaire-36 items

Results for the eight dimensions of the SF-36 are shown in *Figure 8* and *Appendix 3*, *Table 43*. For all dimensions, the two treatment arms of the trial had very similar (baseline-adjusted) SF-36 results at all follow-up points. With the exception of the pain scale, which increased (improved) only slightly, all dimensions increased substantially for both treatment arms and to a similar extent.

For most 'physical' dimensions, the mean SF-36 score increased between baseline and 6 months, but did not increase substantially thereafter. However, the 'role limitations due to physical problems' scale continued to improve steadily over the 2-year follow-up period, as patients recovered from the procedure and gained confidence in their ability to carry out usual activities.

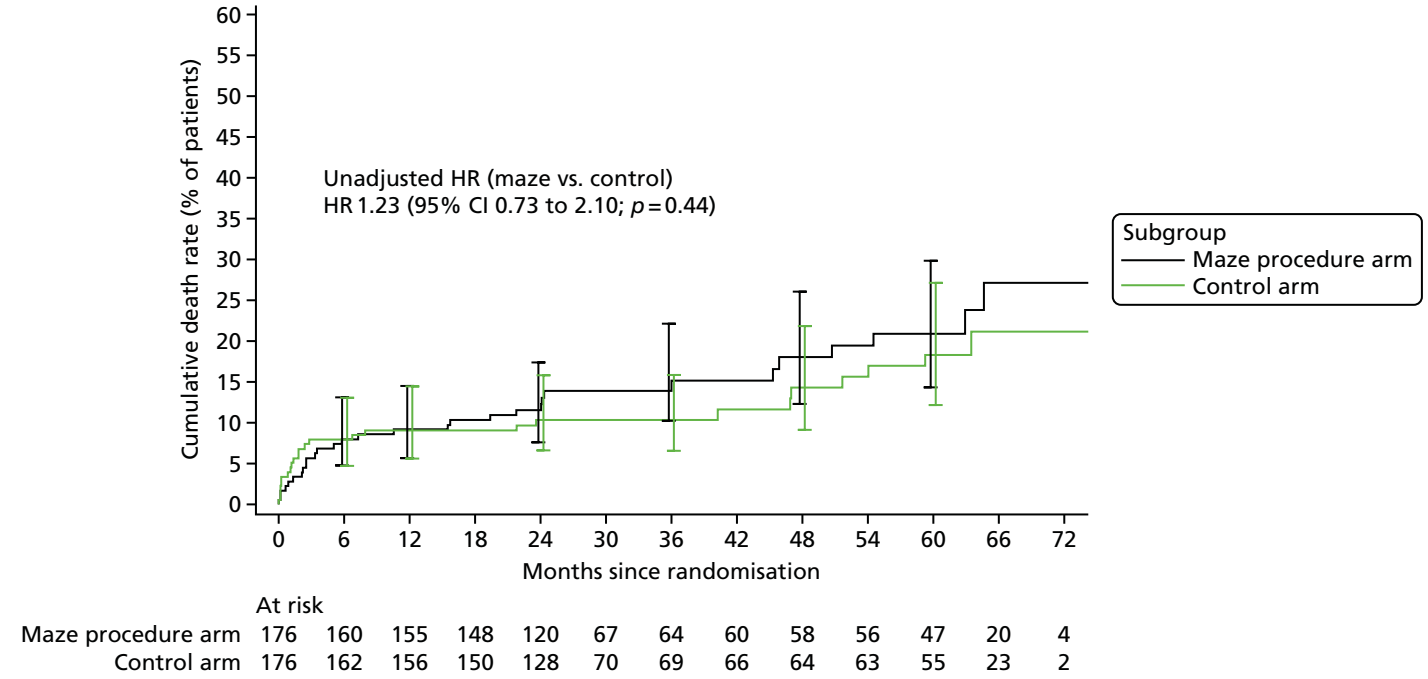


FIGURE 5 Kaplan–Meier estimates of the cumulative incidence of death throughout the trial.

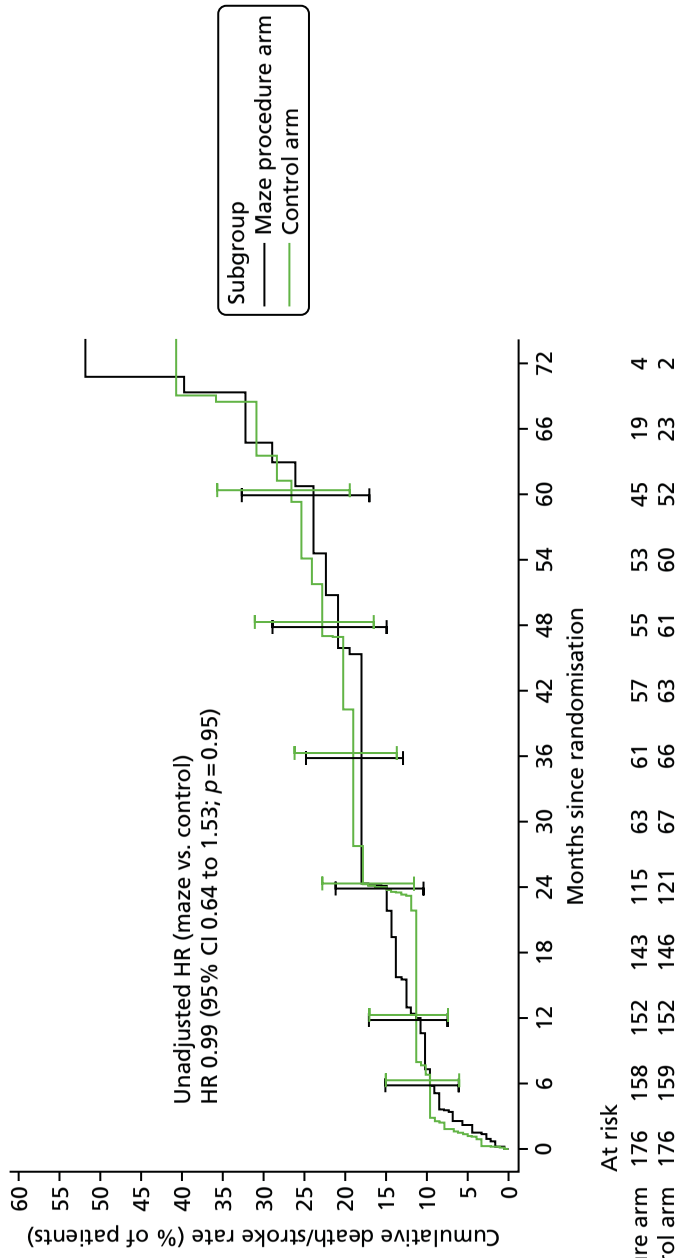


FIGURE 6 Kaplan–Meier estimates of the cumulative incidence of death or stroke.

TABLE 13 Anticoagulant and antiarrhythmic drug use

Time point for each type of drug	Treatment arm (*n*/*N*)		Adjusted OR maze procedure/control	95% CI	*p*-value
	Maze procedure	Control			
Anticoagulants					
Discharge	136/167	141/165	0.619	0.268 to 1.431	0.2616
6 weeks	130/160	138/161	0.483	0.206 to 1.131	0.0936
6 months	113/156	129/160	0.309	0.140 to 0.683	0.0037
12 months	94/149	106/149	0.381	0.178 to 0.818	0.0133
24 months	82/134	96/138	0.389	0.179 to 0.845	0.0171
Antiarrhythmias					
Discharge	138/167	134/165	1.323	0.623 to 2.806	0.4660
6 weeks	137/160	134/161	1.547	0.704 to 3.399	0.2775
6 months	131/156	127/160	1.603	0.751 to 3.423	0.2225
12 months	117/149	109/149	1.541	0.744 to 3.192	0.2446
24 months	102/134	97/138	1.613	0.778 to 3.344	0.1983

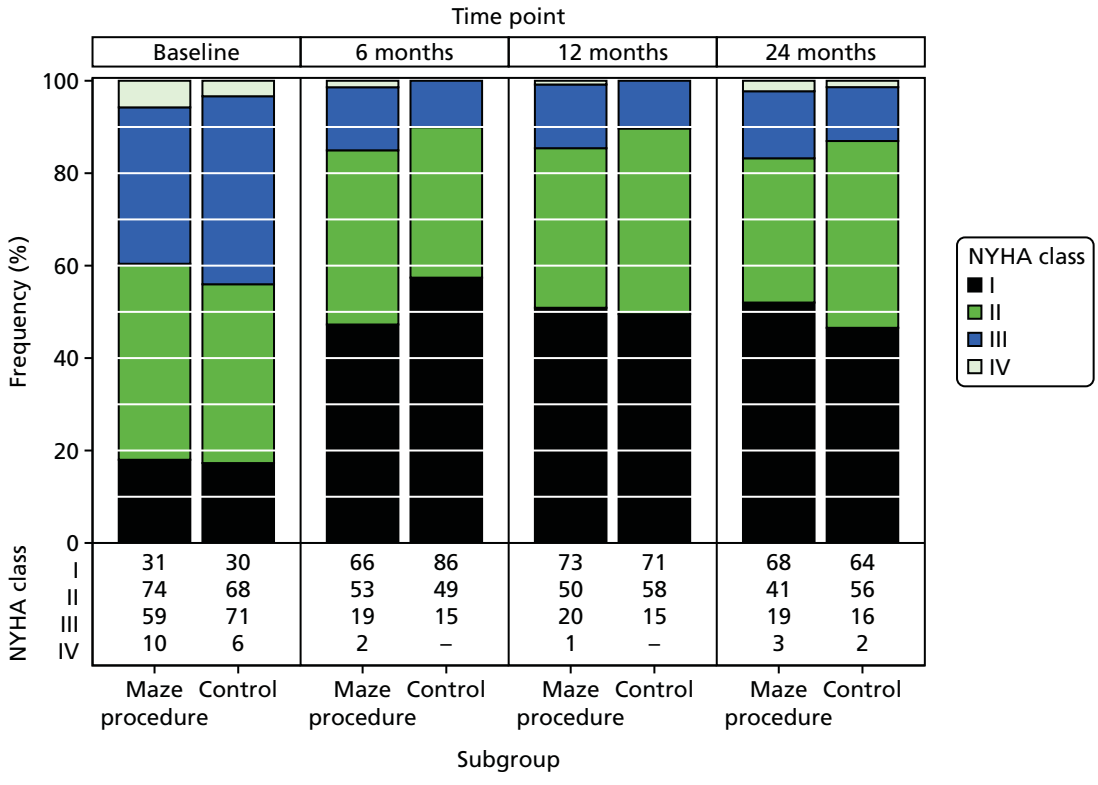


FIGURE 7 Comparison of NYHA classes over time.

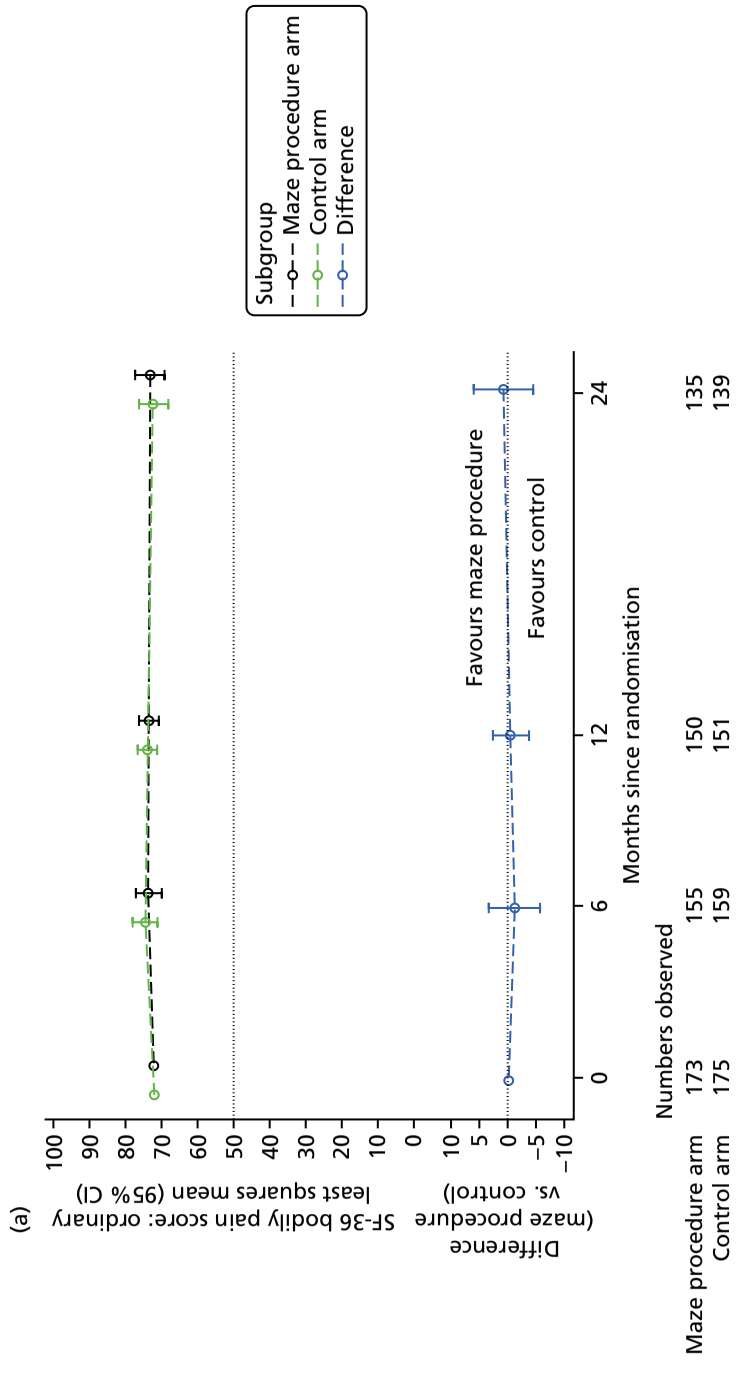


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

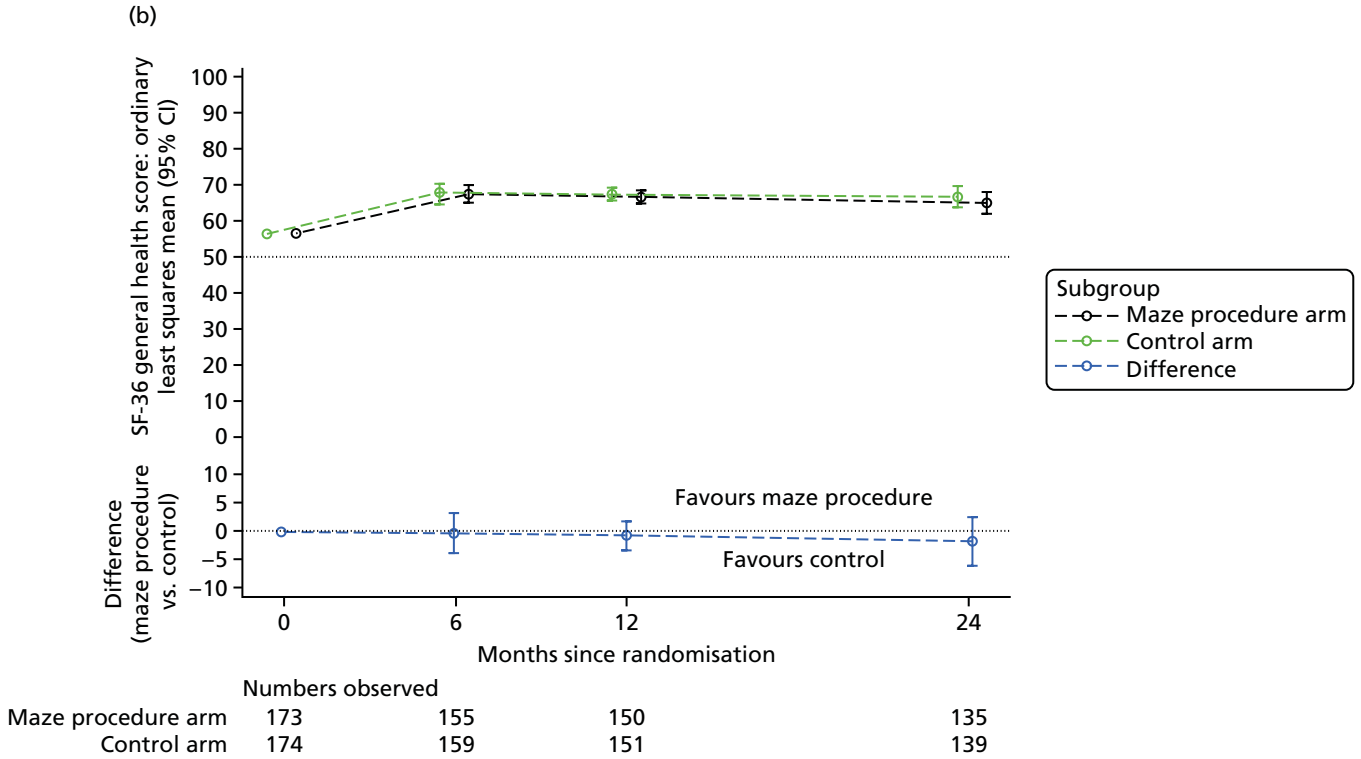


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

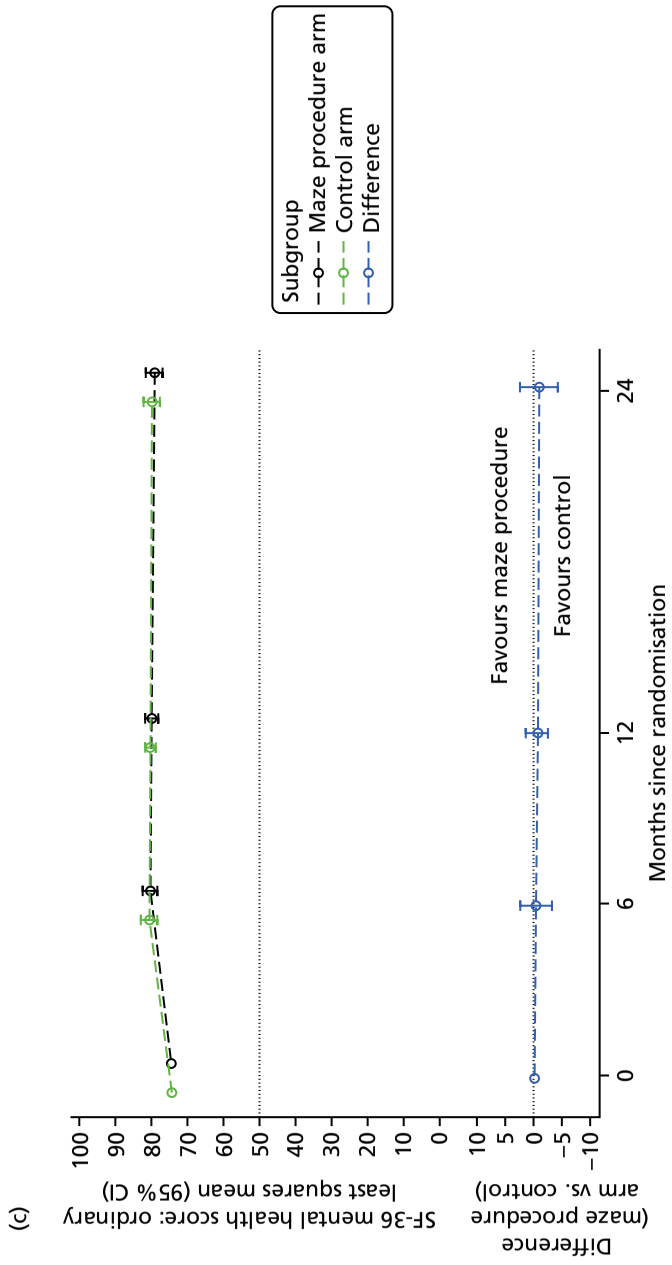


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

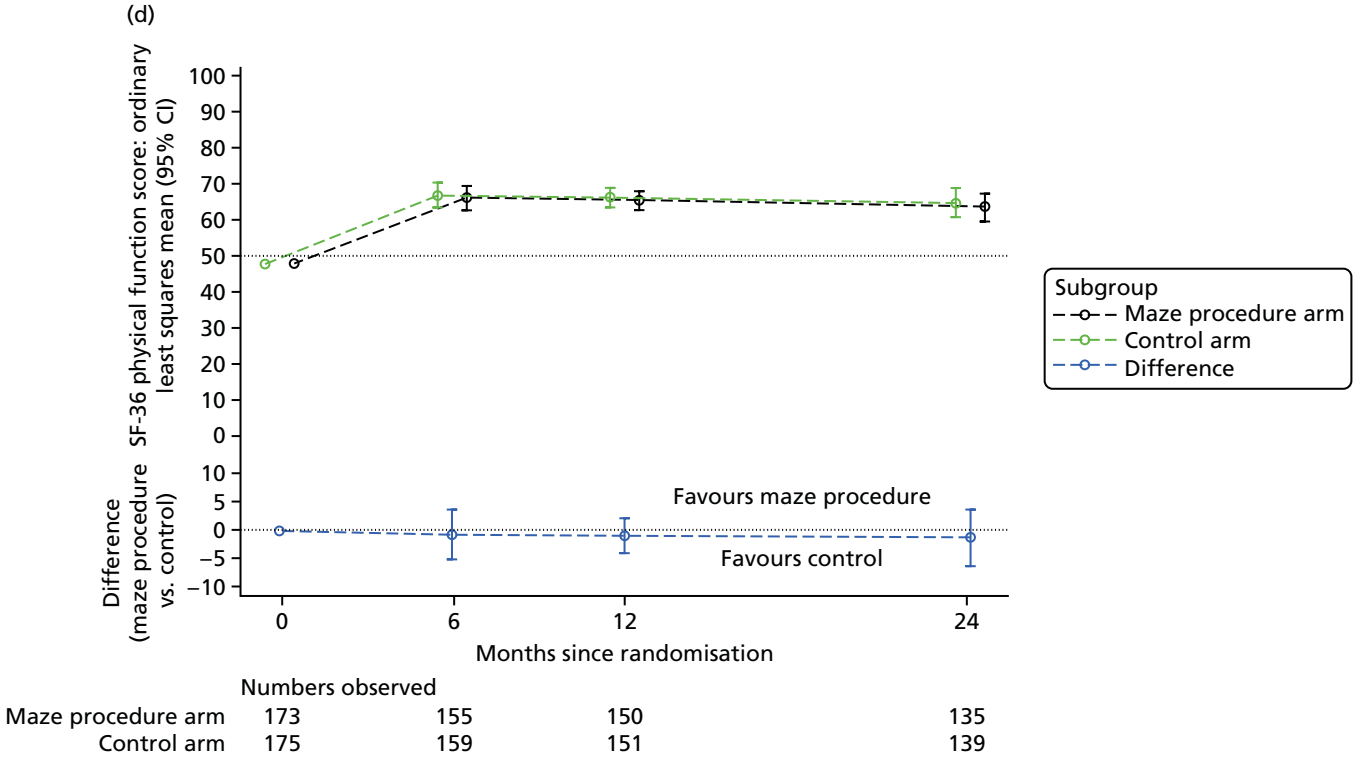


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

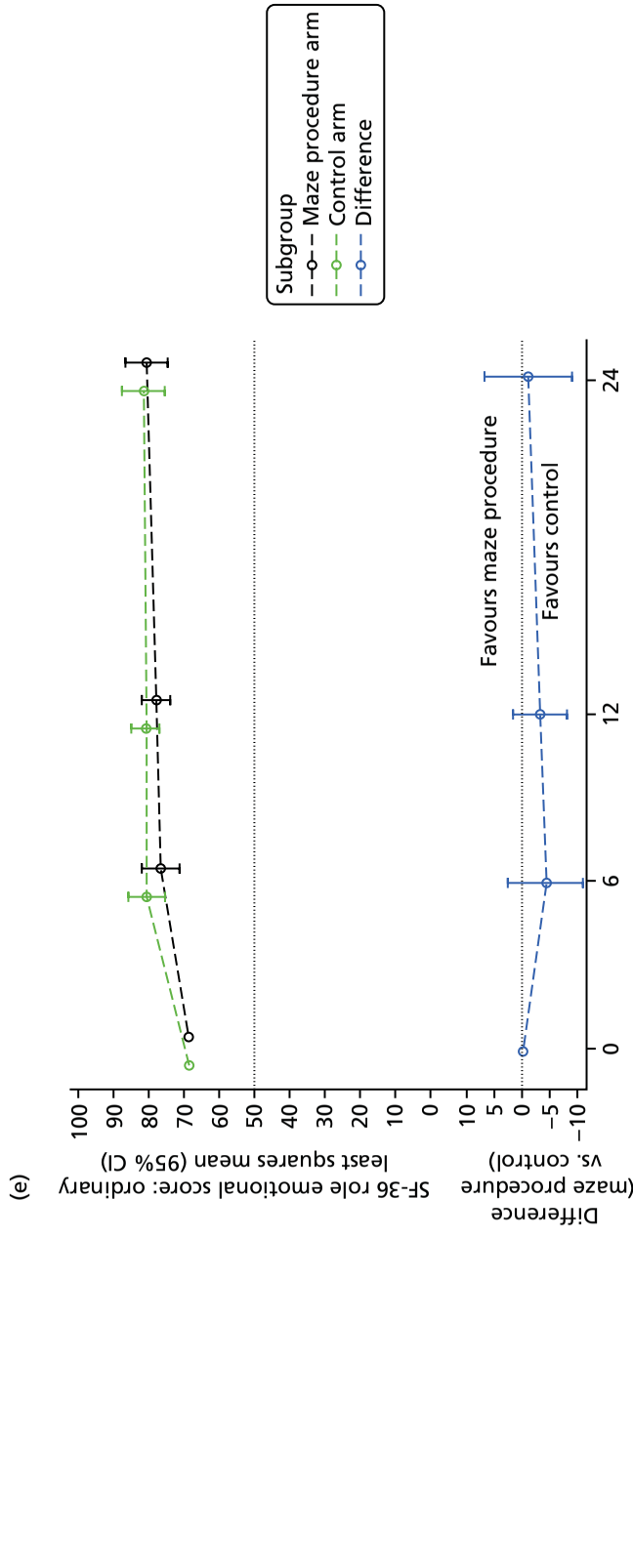


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

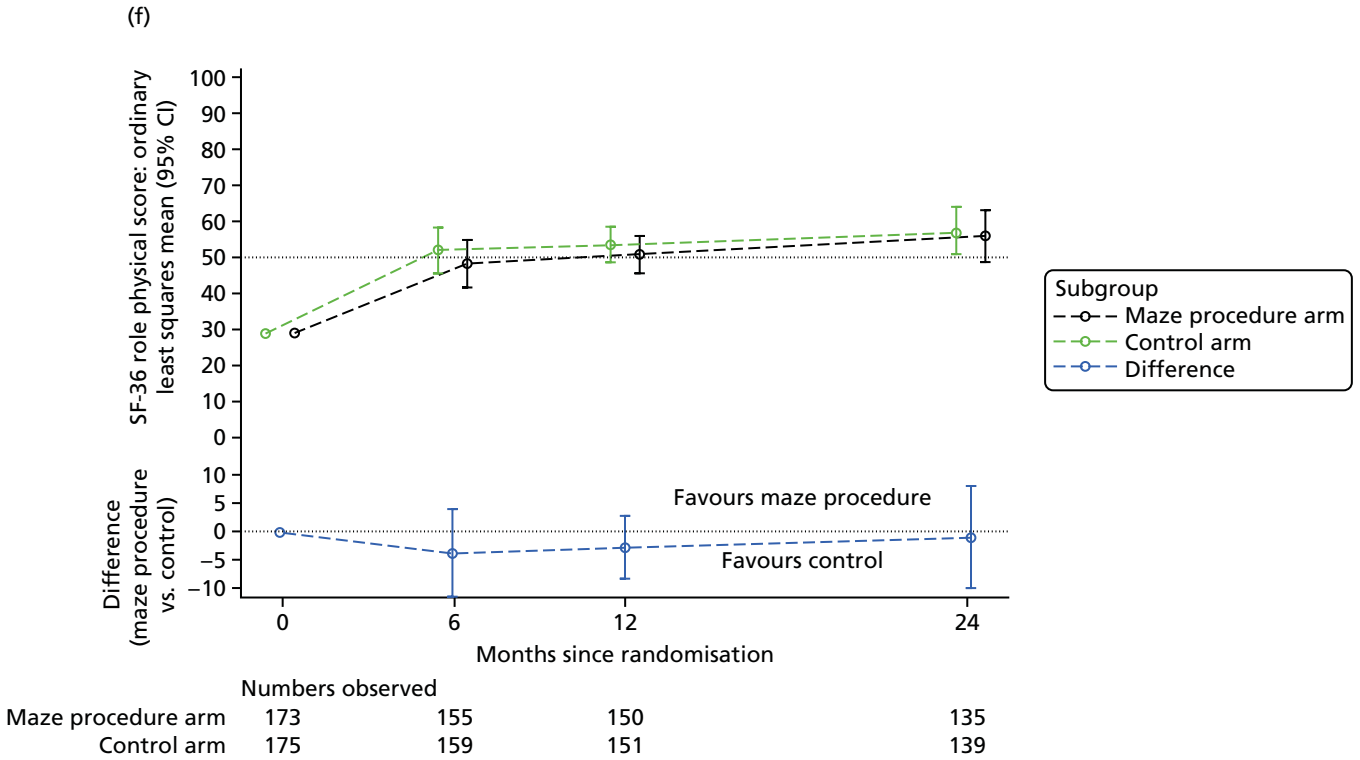


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

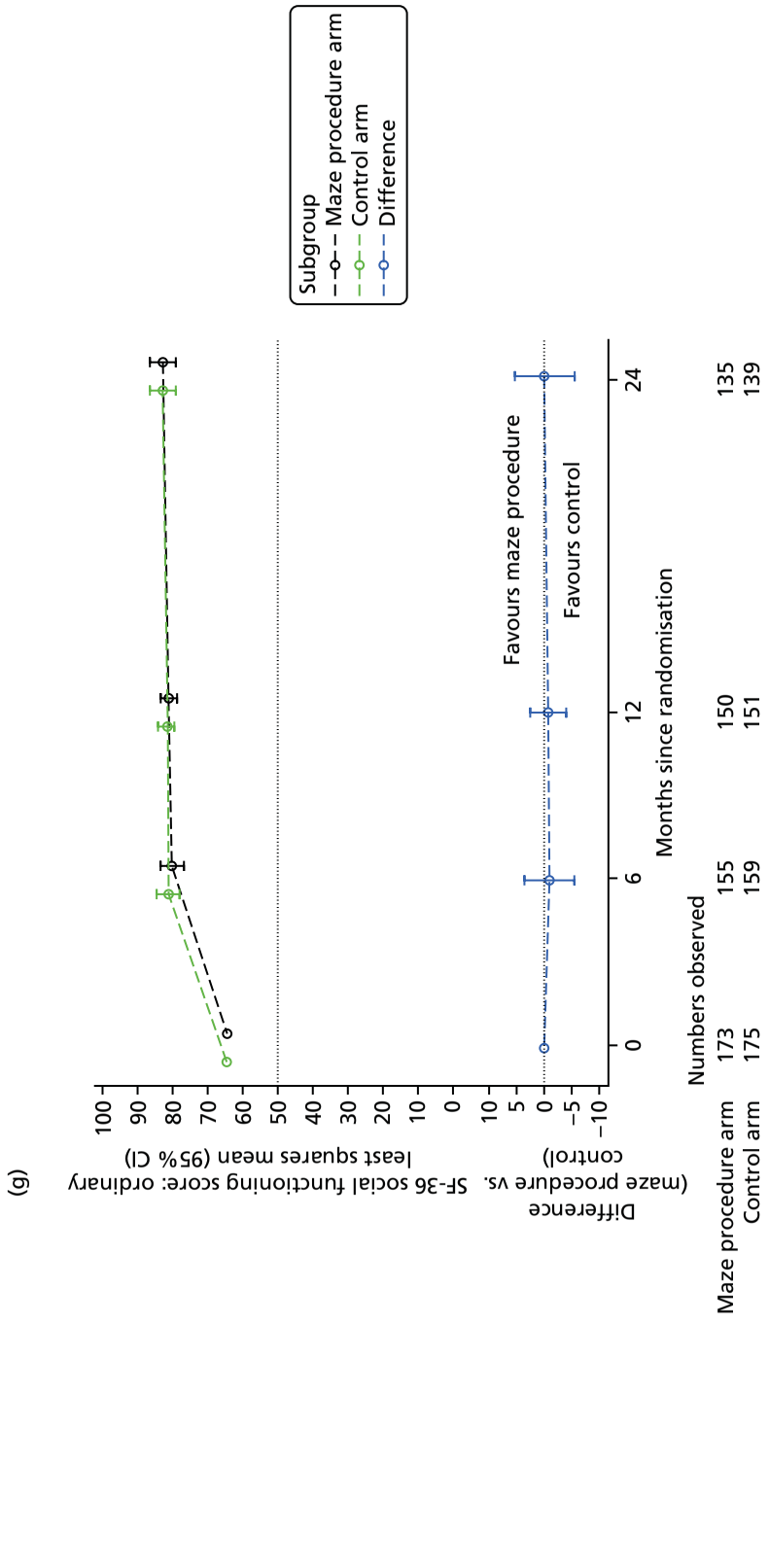


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time. (*continued*)

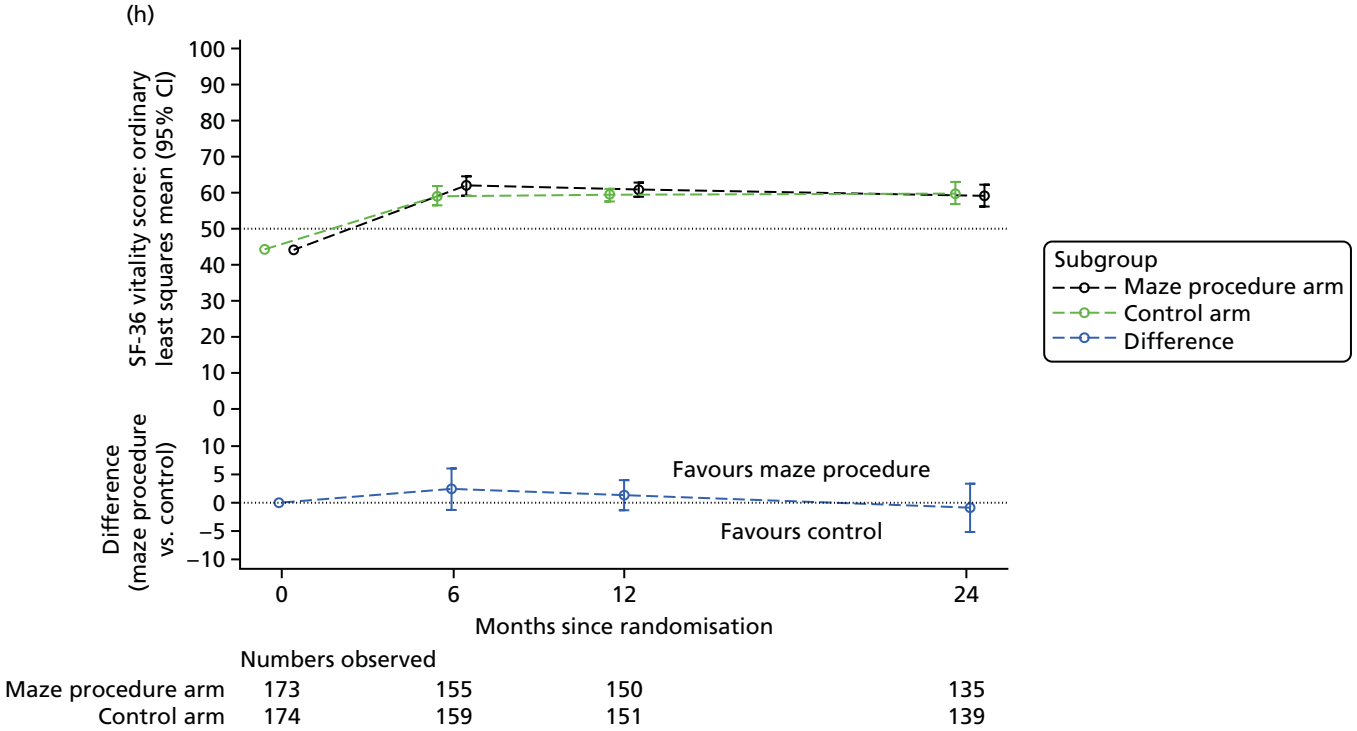


FIGURE 8 Random coefficient model: mean and 95% CIs for SF-36 scores over time. Random patient effect excluded, baseline score adjusted. (a) SF-36 bodily pain score over time; (b) SF-36 general health score over time; (c) SF-36 mental health score over time; (d) SF-36 physical function score over time; (e) SF-36 role emotional score over time; (f) SF-36 role physical score over time; (g) SF-36 social functioning score over time; and (h) SF-36 vitality score over time.

The physical dimension scales all remained somewhat lower than for the general population, which may be expected given that the patients all had heart problems. The baseline-adjusted PCS was around 38 or 39 points on average at all postoperative time points for both treatment arms, which was significantly below the population average of 50 points, but was greater than baseline (*Figure 9*; see also *Appendix 3*, *Table 44*; $p < 0.0001$ for all follow-up points and both treatment arms). The improvement from baseline for the PCS was lower than that for the individual physical scales, as a result of it being a weighted average over all eight individual dimensions, some of which changed only slightly.

For the three scales that measured social, emotional and mental health aspects of life, a similar pattern emerged, with the arms improving between baseline and 6 months to a similar extent, but changing only slightly beyond this point. The mean MCS was very close to the general population average at baseline, and this increased by 4 points on average at 6 months post randomisation, remaining at that level for 2 years for both arms of the trial. There were no significant differences between the two treatment arms at any time point.

Finally, patients reported that their general health status, measured using the SF-36 general health scale, improved in both treatment arms to a similar extent, with no significant differences between the treatment arms at any follow-up time point.

EuroQol-5 Dimensions, three-level version

The results of the EQ-5D-3L utility score for those patients who completed the EQ-5D-3L are summarised in *Appendix 3*, *Table 45* and *Figure 10*. As is common for surgical trials, there was a dip in estimated utility early after surgery, which was largely related to postoperative limitations in usual activities and self-care, and symptoms of pain/discomfort (data not shown). However, by 6 weeks after randomisation, the mean utility had increased in both treatment arms and was slightly higher than at baseline. By 6 months post randomisation, both treatment arms had a mean utility that was higher than the general population aged 65–74 years. There were no significant differences in baseline-adjusted EQ-5D-3L utility score at any follow-up time point. A similar pattern was observed for the visual analogue scale (data not presented).

Safety

The safety results were based on 167 completed maze procedures (165 patients randomised to the maze procedure arm and two patients randomised to the control arm) and 185 non-maze cardiac procedures (11 patients randomised to the maze procedure arm and 174 patients randomised to the control arm). AEs are expected in surgical trials as a result of the high-risk nature of procedures. In the Amaze trial, a total of 560 events (136 patients) were reported after maze procedures and 589 events (157 patients) were reported after control procedures; an overview is provided in *Table 14*, with details in *Appendix 3*, *Table 46*. The number of these events classed as SAEs were 330 (100 patients) and 333 (111 patients) in the maze procedure and control arms, respectively. The proportion of patients having a SAE in the two treatment arms was very similar (maze procedure 100/167; control 111/185; $p = 1.000$). Most events were mild in severity, but 71 maze procedure patients (42.5%) and 84 control patients (45.5%) had at least one event of moderate severity, and 31 maze procedure patients (18.6%) and 38 control patients (20.5%) had a severe event. Twenty-three events in 17 patients (10.2%) were possibly related to treatment in the maze procedure arm, compared with 28 events in 19 patients (10.3%) being possibly related to treatment in the control arm; one control patient (0.5%) was admitted to hospital for investigation of an atrial flutter, which was classified as definitely related to treatment. Overall, the safety profiles of these two treatment arms were similar.

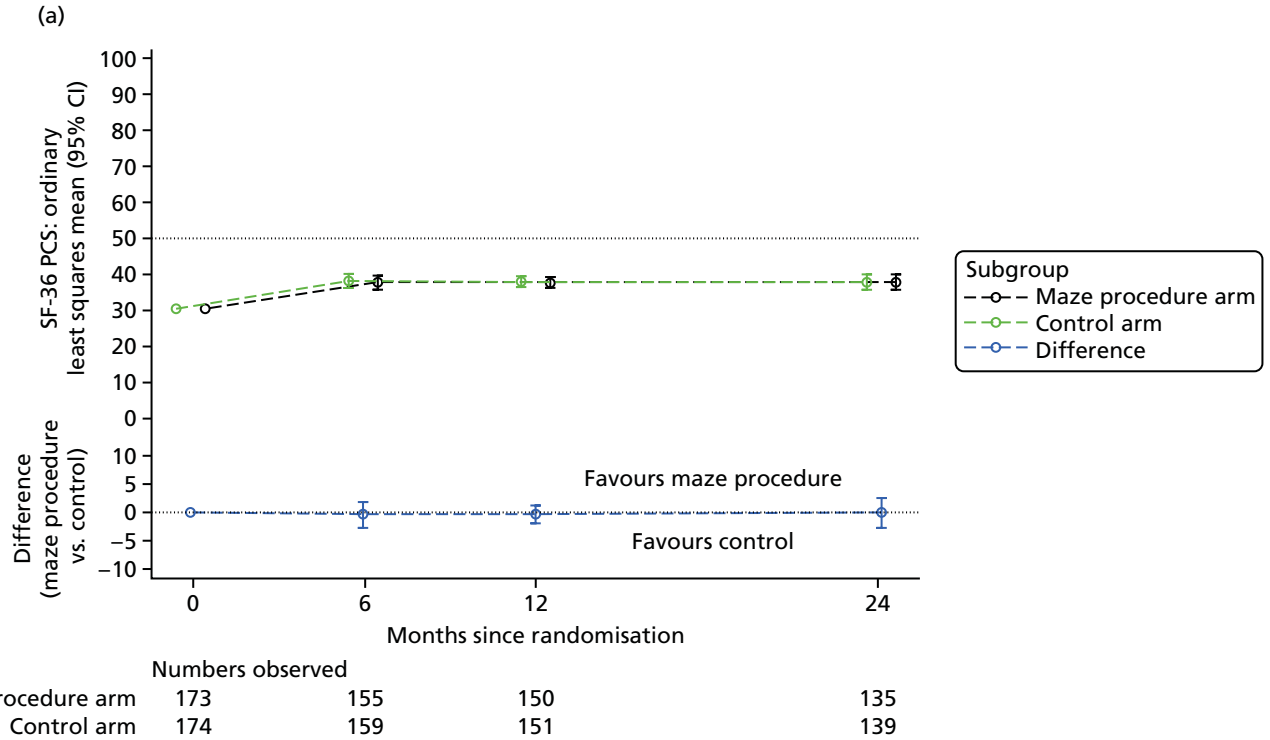


FIGURE 9 Random coefficients model: mean and 95% CIs for the SF-36 PCS and MCS. Random patient intercept, baseline score adjusted. (a) SF-36 PCS over time; and (b) SF-36 MCS over time. (*continued*)

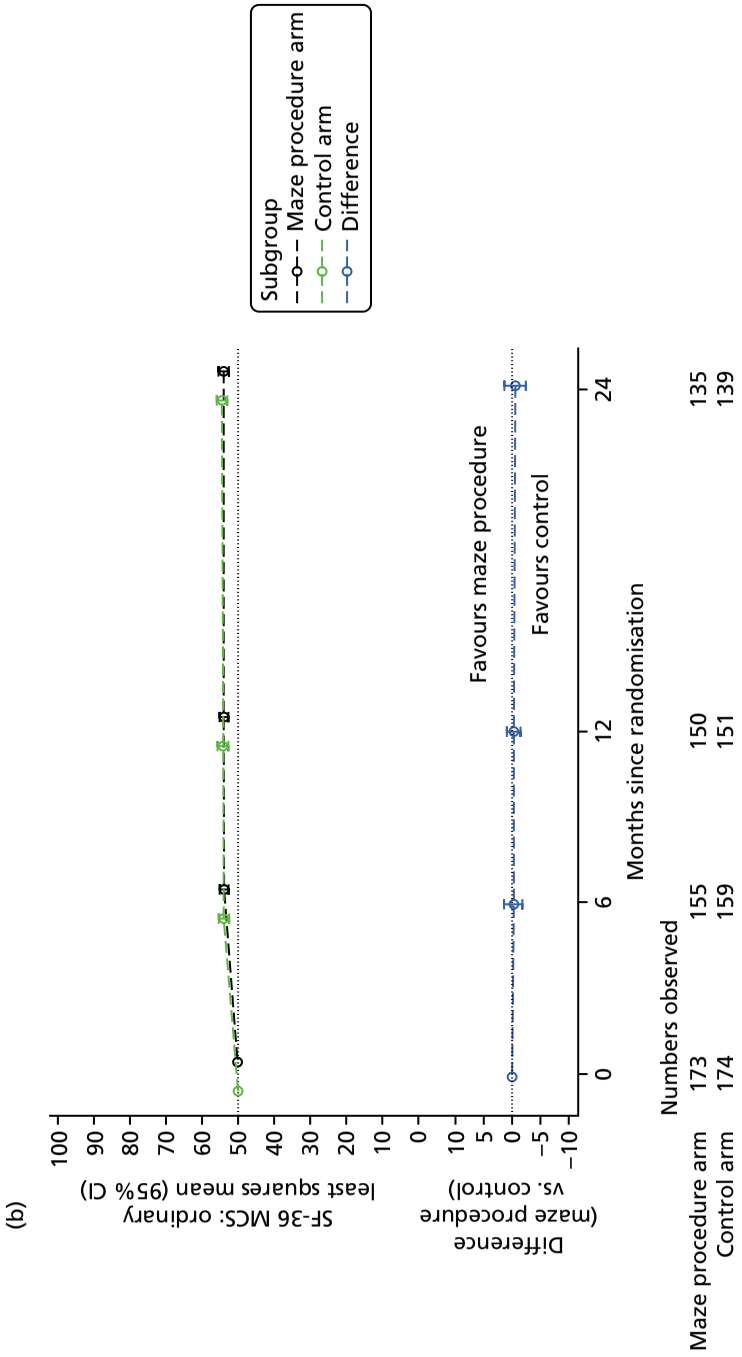


FIGURE 9 Random coefficients model: mean and 95% CIs for the SF-36 PCS and MCS. Random patient intercept, baseline score adjusted. (a) SF-36 PCS over time; and (b) SF-36 MCS over time.

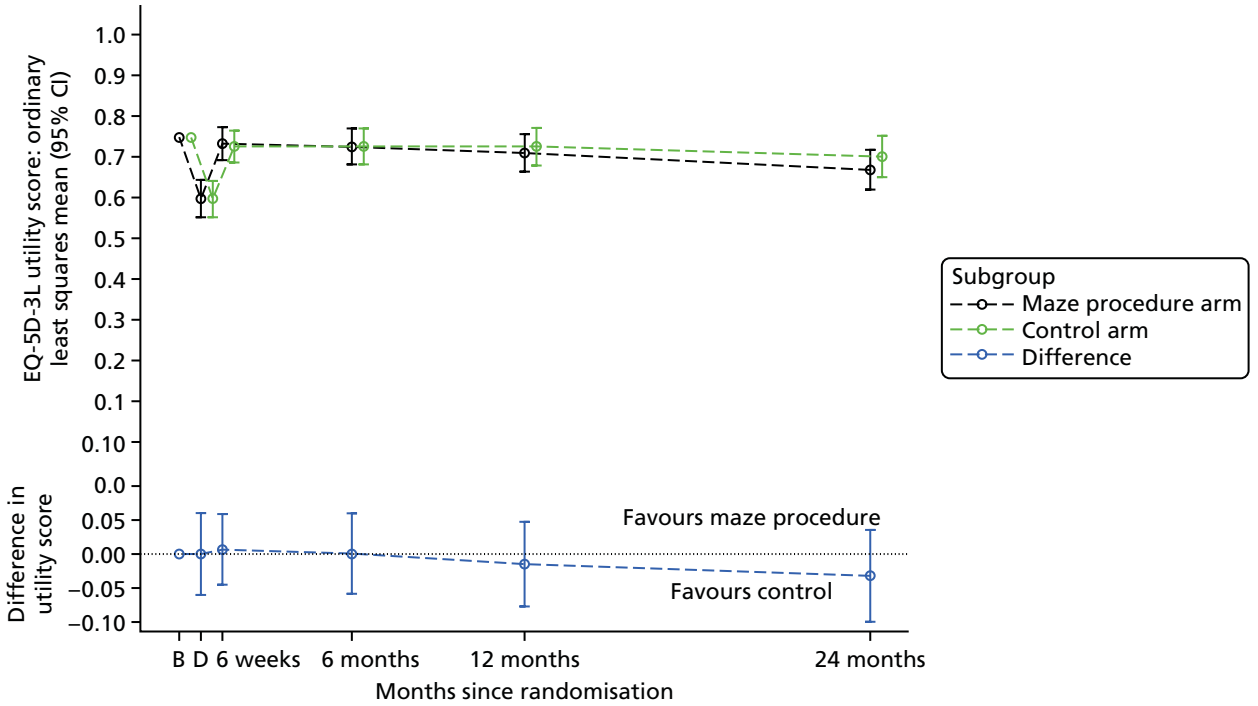


FIGURE 10 Mean and 95% CIs for the baseline-adjusted EQ-5D-3L utility score over time. B, baseline; D, discharge from hospital.

TABLE 14 Overview of the numbers of patients who reported AEs up to 2 years post randomisation

AE category	Treatment arm, *n* (%)		
	Maze procedure (*n* = 167)	Control (*n* = 185)	Total (*n* = 352), *n* (%)
Any AEs/SAEs/SUSARs reported	136 (81.4)	157 (84.9)	293 (83.2)
Any AEs reported	103 (61.7)	116 (62.7)	219 (62.2)
Any SAEs reported	100 (59.9)	111 (60.0)	211 (59.9)
Any SUSARs reported	0 (0.0)	0 (0.0)	0 (0.0)
Related events reported			
Definitely related	–	1 (0.5)	1 (0.3)
Possibly related	17 (10.2)	19 (10.3)	36 (10.2)
Unrelated	133 (79.6)	151 (81.6)	284 (80.7)
Severity			
Mild	109 (65.3)	120 (64.9)	229 (65.1)
Moderate	71 (42.5)	84 (45.4)	155 (44.0)
Severe	31 (18.6)	38 (20.5)	69 (19.6)

SUSAR, suspected unexpected serious adverse reaction.

Chapter 4 The 'Has Electrical Sinus Translated into Effective Remodelling?' substudy

Background

An important issue related to the maze procedure is that restoration of SR may not result in the return of atrial contractile function. Clinical consensus is that the increased risk of thromboembolic complications in AF is related to blood stagnation in parts of the left atrium, caused by the loss of contractile function. In addition, loss of atrial contraction also contributes to exacerbation of congestive heart failure and atrioventricular asynchrony.[65]

Published literature on atrial transport following the maze procedure is limited by small sample sizes and selection bias. In addition, most studies had no matched control patients for comparison. To date, there is little rigorous evidence that attempting to restore SR by treating AF with an ablation device during cardiac surgery restores atrial transport.

Objectives

The HESTER substudy aimed to assess whether or not patients in SR at least 1 year after an adjunct maze procedure had an equivalent active left atrial ejection fraction (ALAEF) to that of control patients who had undergone cardiac surgery and were in SR both prior to and at least 1 year after surgery.

Methods

Patients and matching

The HESTER substudy was a (1 : 1) matched cohort study: the maze procedure cohort were (trial or non-trial) patients with AF who underwent the maze procedure as an adjunct to routine cardiac surgery and who were in SR at least 1 year postoperatively; the control cohort were patients who were in SR both before and at least 1 year after routine cardiac surgery.

Eligibility criteria were patients aged > 18 years, who were able to give informed consent and attend investigations, and who had a presence of ECG-confirmed SR at least 1 year after cardiac surgery. Exclusion criteria were contraindications to cardiac MRI: cardiac pacemakers, surgical clips in the head, electronic inner-ear implants, ocular metal fragments, electronic stimulators, implanted pumps and severe claustrophobia.

Consecutive eligible participants in the Amaze trial, recruited at Papworth Hospital and Glenfield Hospital, as well as consecutive non-trial AF patients who received the maze procedure as part of their routine elective surgery and were in SR at their last hospital visit, were invited to participate in the HESTER substudy. After consenting, each participant had confirmatory ECG.

A list of potential control patients was compiled from Papworth Hospital databases by an independent member of the hospital audit team. Once a maze procedure patient had consented, a list of potential matched control patients was compiled, matched for time since procedure (± 6 months), age (± 5 years), sex, type of surgery, preoperative LVEF and predicted operative mortality (logistic EuroSCORE).[61] Matching was undertaken by researchers independent of data collection and analysis. Hospital records for potential control patients were reviewed to confirm SR before surgery and the matching criteria before inviting patients to take part.

Data collection

Eligible patients underwent ECG to confirm SR, transthoracic two- and three-dimensional echocardiography and MRI to evaluate left atrial function.

All echocardiography studies were performed using a Philips IE33 ultrasound machine (Philips UK Ltd, Guildford, UK) and the British Society of Echocardiography standard minimum data set. Analysis was by Philips QLab software, version 9.0 (Philips UK Ltd, Guildford, UK), and Xcelera, version 3.2.1.712-2011 (Philips UK Ltd, Guildford, UK), reporting system.

Magnetic resonance imaging studies used a 1.5-T MRI scanner (Siemens Avanto, Erlangen, Germany) and were evaluated using Argus cardiac software, version B17 (Siemens Healthcare, Erlangen, Germany). Images were acquired in standard planes, positioned either parallel to (horizontal and vertical long-axis planes) or perpendicular to the long axis of the heart (short-axis planes), using ECG-triggering steady-state gradient echo sequences (balanced fast-field echo). Left atrial volumes were measured with Argus software.

Each investigation was performed and interpreted by a single operator blinded to the patient identity.

Outcomes

The primary outcome was ALAEF, the measurement most directly related to active left atrial contractility. Maximum left atrial volume (LAV_{max}), minimum left atrial volume (LAV_{min}) and pre A-wave left atrial volume (LAV_{preA}) were measured by echocardiography. ALAEF was derived as a percentage:

$$\text{ALAEF} = 100 \times \frac{\text{LAV}_{\text{preA}} - \text{LAV}_{\text{min}}}{\text{LAV}_{\text{preA}}}. \quad (1)$$

Each volume was measured using both the two- and four-chamber views. The four-chamber assessment of ALAEF was our primary end point, with the two-chamber version being a key secondary end point.

Volume measurements and E/A ratio (an ECG marker of left ventricular function) were secondary end points. Derived secondary end points were active stroke volume ($LAV_{preA} - LAV_{min}$), passive stroke volume ($LAV_{max} - LAV_{preA}$) and left atrial ejection fraction (LAEF):

$$\text{LAEF} = 100 \times \frac{\text{LAV}_{\text{max}} - \text{LAV}_{\text{min}}}{\text{LAV}_{\text{max}}}. \quad (2)$$

Maximum left atrial volume and LAV_{min} were also measured using the MRI multiple-slice method and were, along with the LAEF derived from them, secondary end points.

Sample size

In AF, ALAEF is virtually zero, so that any measurable ALAEF after SR restoration is a marker of at least partial treatment success. In a previous study of subjects in SR, with a LAV_{max} of 50–70 ml, but without atrioventricular or intraventricular conduction abnormalities on a resting 12-lead ECG, the mean ALAEF (measured in four-chamber view echocardiography) was 43% with a SD of 18.2%.[66] Based on this, and the expert judgement of the investigators, the minimum clinically important difference in ALAEF was set at 1 SD, or 18.2%. Equivalence would therefore be concluded on estimating a treatment effect (maze procedure vs. control) with a two-sided 95% CI contained entirely in the interval from –18.2% to 18.2%. Based on this, 22 patients in each treatment arm would provide 80% power to demonstrate equivalence.

Statistical analyses

For the primary end point, a mixed-effects linear regression model was fitted, including fixed effects for treatment and matching variables, and random effects for the matched pairs. A variance components model was assumed for the covariance structure, with residuals at both levels assumed to be independent

and normally distributed. The estimated treatment coefficient from the model was taken as the mean difference in ALAEF between maze procedure patients and control patients. Model assumptions were checked using residual plots.

The same model was fitted to each secondary end point. If the estimated random-effects variance was zero for any end point, the pairing was included as a fixed effect. No equivalence margins were specified for the secondary end points.

Analysis was undertaken in SAS version 9.4, using the PROC MIXED procedure to fit mixed-effects models by the method of restricted maximum likelihood. There were no missing values for the primary end point. For secondary end points with missing values, complete-case analyses were used. The guidelines from the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement were followed.[67]

Results

Between 24 July 2013 and 8 July 2015, 22 eligible patients were recruited for each cohort and underwent echocardiography and MRI (*Figure 11*). All 22 control patients and 15 maze procedure patients were from Papworth Hospital, and seven maze procedure patients were from Glenfield Hospital. No patients experienced AEs during these tests.

Summaries of the variables used in matching are presented in *Table 15*. Matching by sex was exact. Two pairs differed in left ventricular function (maze procedure arm: LVEF of > 50%, control arm: LVEF of 30–50%; and maze procedure arm: LVEF 30–50%, control arm: LVEF of > 50%). Only two pairs differed in operation performed: one maze procedure patient who underwent mitral and tricuspid valve surgery was matched with a control patient who underwent mitral surgery alone; another who underwent CABG and atrial septal defect repair was matched with a control patient who underwent isolated CABG. For all but three of the pairs, the length of time since surgery was longer for the maze procedure patient than the control patient. Nine pairs differed by > 6 months in the length of time since surgery. Maze procedure patients were, on average, older and 16 were older than their matched control patients, including two age differences > 5 years. Maze procedure patients tended to have a slightly higher logistic EuroSCORE.[61]

Summaries of the four-chamber ALAEF assessment are in *Table 16*. The SD in both cohorts was 8%, but maze procedure patients had a lower mean ALAEF (18%) than control patients (26%). *Figure 12* shows that the ALAEF of the control patients was higher in all but three pairs. The maximum ALAEF was similar between treatment arms (41% for control patients and 39% for maze procedure patients), but the minimum for maze procedure patients (6%) was less than half that of the control patients (13%).

After adjusting for the matched design, the mean difference in ALAEF (four-chamber view) between maze procedure and control patients was –8.03 (95% CI –12.43 to –3.62; see *Figure 13* and *Appendix 3*, *Table 47*). The 95% CI was contained entirely in the interval (–18.2 to 18.2), so our predefined criterion for equivalence was met. However, the mean ALAEF was significantly lower in maze procedure patients than in control patients ($p = 0.0015$).

For the two-chamber measurements, the mean maze–control difference was –3.48 (95% CI –8.45 to 1.49), supporting the conclusions of the primary outcome analysis.

Excluding one patient in the maze procedure cohort with a high logistic EuroSCORE[61] of 16.9% did not affect the conclusion of equivalence (mean difference –7.90, 95% CI –12.55 to –3.25).

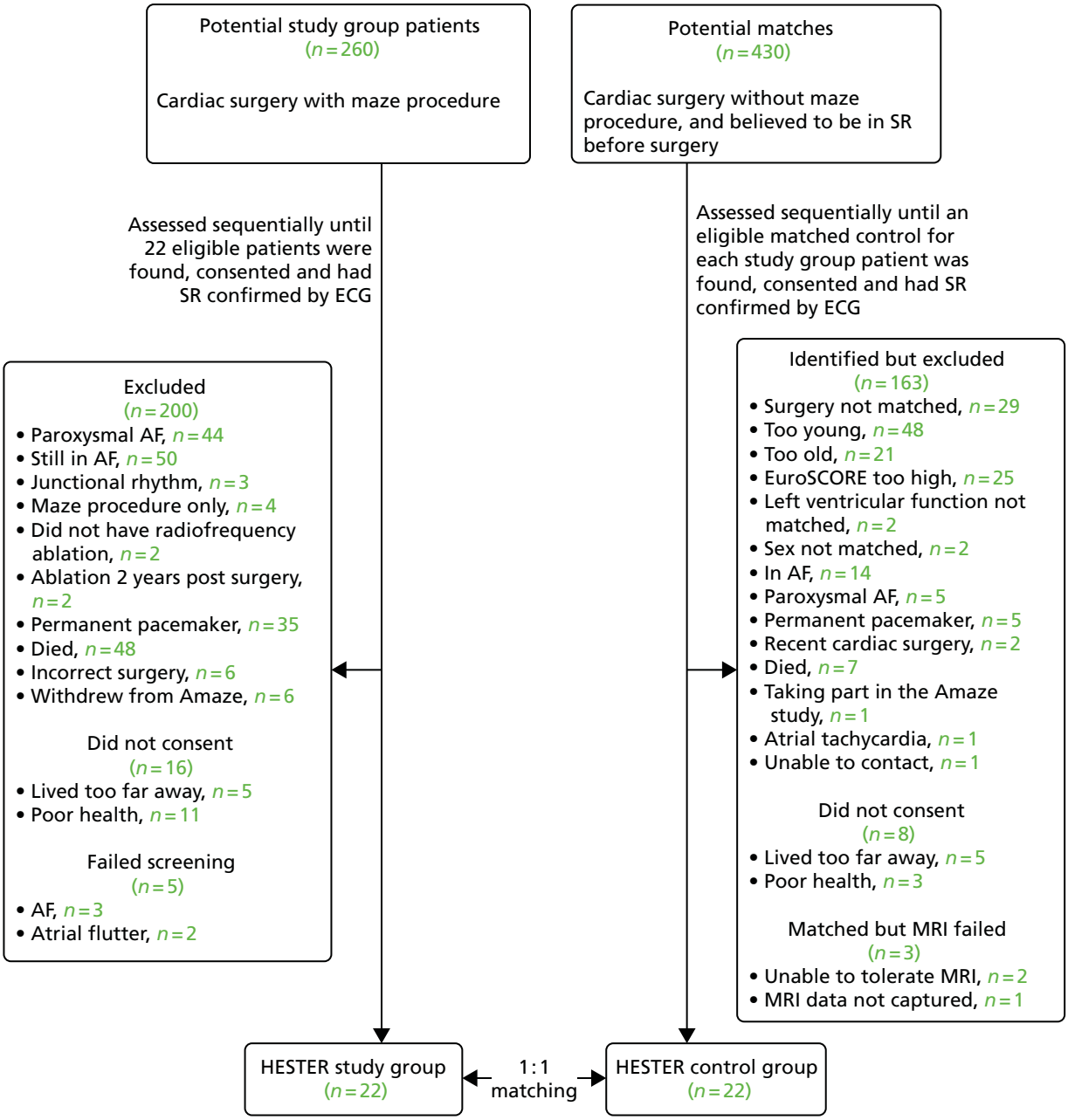


FIGURE 11 Recruitment and matching for the HESTER substudy.

Secondary outcomes are summarised in *Appendix 3*, *Tables 47–49*. Although the summaries for the four- and two-chamber echocardiography measurements were very similar, they were not consistent with the MRI summaries.

In regression analysis, the mean E/A ratio was significantly higher and the mean LAEF (four-chamber view and MRI) was significantly lower for maze procedure patients than those for control patients (*Figure 14*). There were no significant differences in the other end points.

TABLE 15 Summary statistics for the variables used in matching maze procedure and control patients in the HESTER substudy

	Treatment arm	
Variable	Maze procedure ($n = 22$)	Control ($n = 22$)
Age at surgery (years), mean (SD)	69 (9)	66 (12)
Years since surgery, mean (SD)	2.7 (1.0)	2.9 (1.1)
Logistic EuroSCORE (%), mean (SD)	4.3 (3.7)	3.7 (2.5)
Left ventricular function, *n* (%)		
Poor (LVEF of < 30%)	1 (5)	1 (5)
Moderate (LVEF of 30–50%)	5 (23)	5 (23)
Good (LVEF of > 50%)	16 (73)	16 (73)
Sex, *n* (%)		
Male	20 (91)	20 (91)
Female	2 (9)	2 (9)
Surgery, *n* (%)		
CABG	5 (23)	6 (27)
MVR	8 (36)	9 (41)
AVR	2 (9)	2 (9)
Combined procedures	6 (29)	4 (24)
ASD repair	1 (5)	1 (5)

ASD, atrial septal defect.

TABLE 16 Mean (SD) of all primary and secondary end points, measured using echocardiography (four-chamber view), in the HESTER substudy

	Treatment arm	
End point	Maze procedure ($n = 22$)	Control ($n = 22$)
ALAEF (%)	18 (8)	26 (8)
LAEF (%)	31 (9)	38 (7)
LAV_{min} (ml)	59 (20)	45 (14)
LAV_{preA} (ml)	72 (22)	60 (18)
LAV_{max} (ml)	86 (25)	72 (19)
Passive stroke volume (ml)	13 (7)	11 (4)
Active stroke volume (ml)	13 (7)	16 (7)
E/A ratio	1.8 (0.8)	0.9 (0.4)

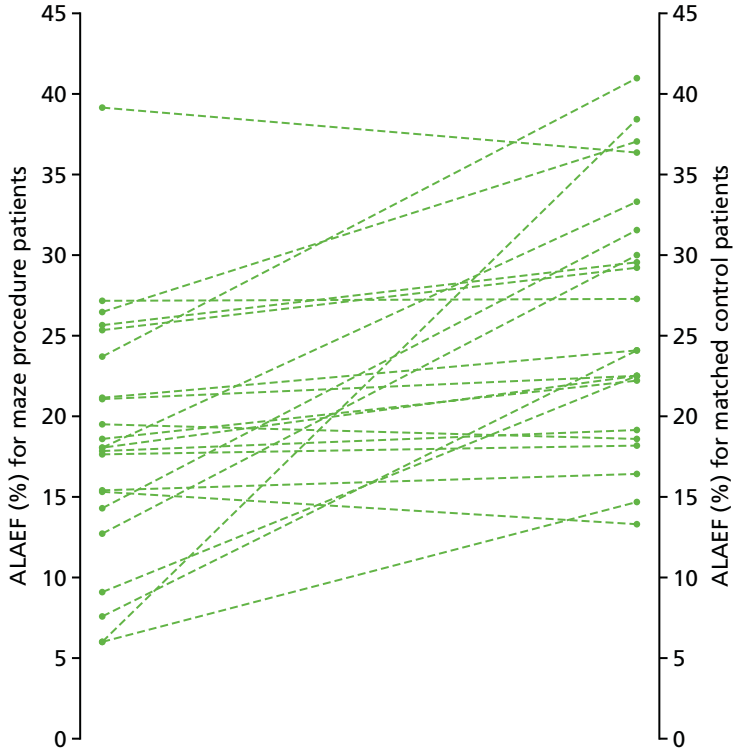


FIGURE 12 Four-chamber ALAEF measurements for individual patients undergoing the maze procedure and their matched control patients for the HESTER substudy.

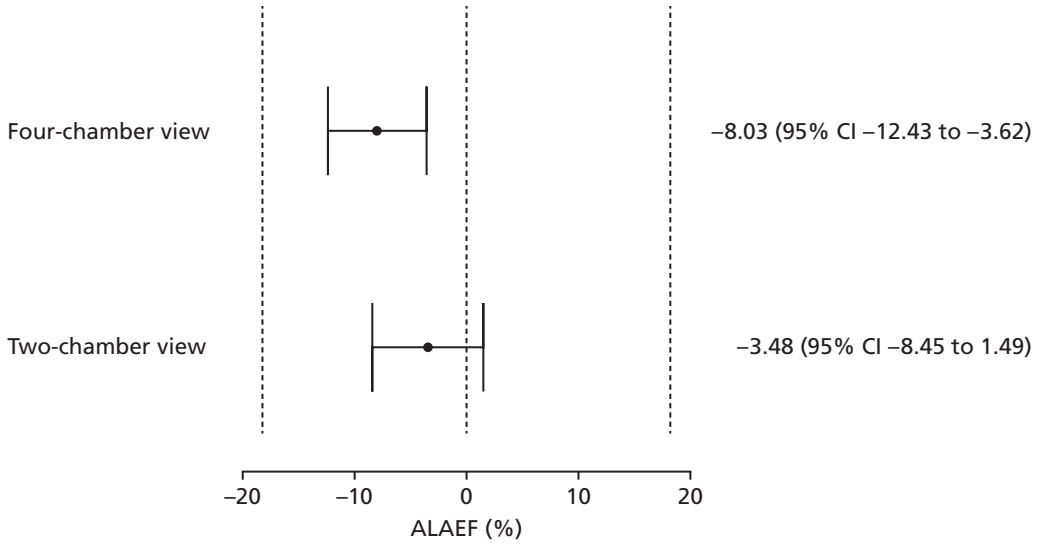
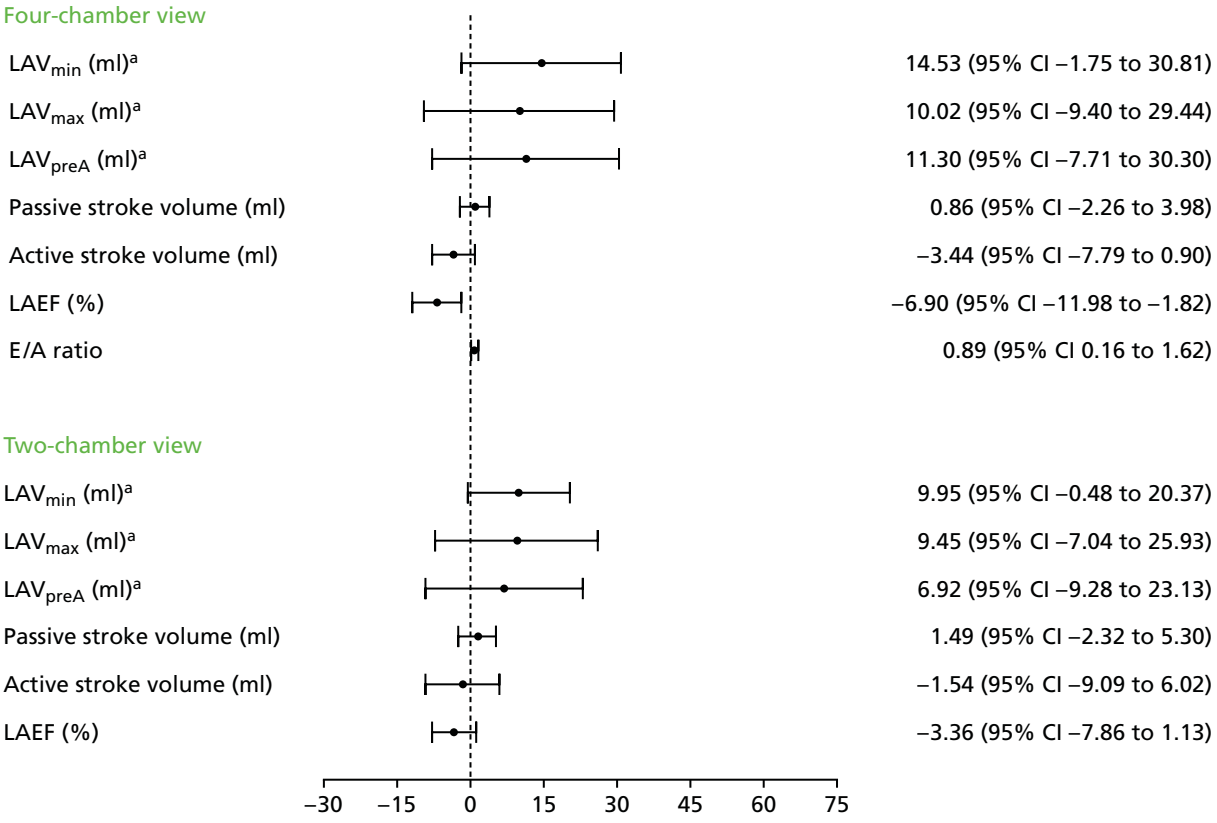


FIGURE 13 Forest plot showing estimated treatment effects (maze procedure vs. control) and 95% CIs for the four- and two-chamber measurements of ALAEF in the HESTER substudy. Equivalence limits (at ± 18.2%) are shown as dashed vertical lines.

Echocardiography

Four-chamber view



MRI

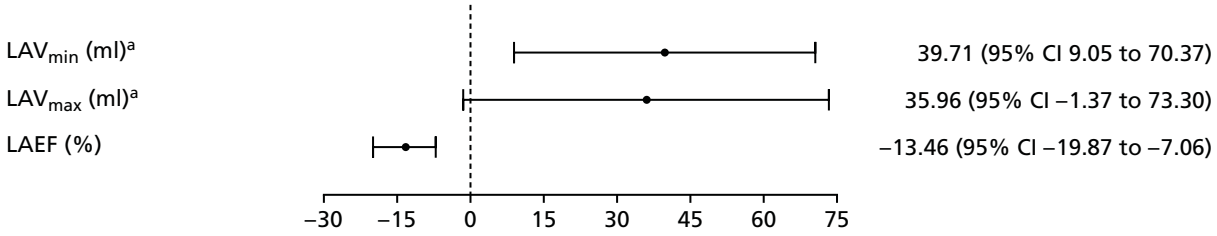


FIGURE 14 Forest plots showing estimated treatment effects (maze procedure vs. control) and 95% CIs for secondary end points in the HESTER substudy. a, For these variables, the models were fitted with patient pairing as a fixed effect.

Chapter 5 Trial results: cost-effectiveness

Data completeness

There were very few item non-responses for resource use during the primary admission (one control patient and two patients from the maze procedure arm had missing length of stay data). Half the patients who were referred to a rehabilitation centre (nine maze procedure patients and six control patients) or an acute hospital (three maze procedure patients) had missing final discharge dates and, therefore, values were imputed. Excluding patients who died, the proportion of missing resource-use items at each follow-up point ranged from 6.5% to 13.4% and the proportion lost to follow-up grew from 4% at 6 weeks to 6.5% at 12 months (*Table 17*). The proportion of incomplete EQ-5D-3L and SF-36 scores ranged from 0.9% to 10.5% across time. Patients with poorer initial health status, measured using the EQ-5D-3L, were significantly more likely to have missing responses.

Resource costs

Table 18 shows the length of stay for each stage of the primary admission, by intervention arm and control arm. Patients in the maze procedure arm had a longer stay in the ICU and the cardiac ward, and spent a longer time in an acute hospital, following referral, than those in the control arm. The distribution of resource use in the maze procedure arm included two outliers: one patient stayed in an ICU post operation

TABLE 17 Availability of resource-use follow-up data

		Treatment arm (*n*)		
Time point	Follow-up resource use	Maze procedure	Control	Total
6 weeks	Observations	157	156	313
	Dead	6	10	16
	Lost to follow-up	8	6	14
	Missing	5	4	9
6 months	Observations	152	154	306
	Dead[a]	14	14	28
	Lost to follow-up[a]	13	9	22
	Missing	5	5	10
12 months	Observations	147	146	293
	Dead[a]	16	16	32
	Lost to follow-up[a]	25	20	45
	Missing	1	3	4
24 months	Observations	130	134	264
	Dead[a]	23	18	41
	Missing	23	24	47

a The numbers presented here differ from the CONSORT flow diagram (see *Figure 2*). These missing numbers are based on completion of the resource-use CRF during follow-up and cumulative death at each time period.

TABLE 18 Length of stay in theatre, ICU and wards for primary admission

Resource use	Treatment arm	Number of observations	Mean	Min.	Max.
Theatre time (minutes)	Maze procedure	176	261.2	75.0	582.0
	Control	176	247.4	100.0	775.0
Cardiac ward (days)	Maze procedure	174	8.7	0.0	153.0
	Control	176	7.9	0.0	33.0
ICU (days)	Maze procedure	175	3.2	0.0	52.9
	Control	176	2.4	0.2	36.2
Rehabilitation centre referral (days)	Maze procedure	167[a]	0.3	0.0	19.0
	Control	170[b]	1.0	0.0	84.0
Acute hospital referral (days)	Maze procedure	173[c]	2.4	0.0	144.0
	Control	176[d]	0.4	0.0	45.0

max., maximum; min., minimum.
a Four maze procedure patients were referred.
b Eight control patients were referred.
c Eight maze procedure patients were referred.
d Four control patients were referred.

for 52.9 days and, following discharge directly to their home, died a few days later; and one patient was in the cardiac ward for 144 days.

The average total cost of resources in the maze procedure arm was £3839 more than that of the control arm, and this difference was statistically significant ($p = 0.0013$). *Table 19* shows that the primary admission had the highest cost (£14,880 in the maze procedure arm; £11,406 in the control arm) and resulted in a large proportion of the difference in cost between the trial arms (£3442; $p = 0.0016$). The primary admission contributed most to the variability in cost, as both arms had many high-cost outliers (*Figure 15*). The difference in cost between the trial arms during the primary admission stemmed largely from additional length of stay in each hospital location (surgery, ICU and cardiac ward, with the exception of rehabilitation centres), additional costs of surgical equipment to carry out the maze procedure and the influence of two outliers (see *Appendix 2*, *Tables 25* and *26*, for a detailed breakdown of resource use and total costs). *Table 19* shows that, after discounting, the average cost difference was £3841 (95% CI £1514 to £6167) higher per patient for the maze procedure arm than for the control arm ($p = 0.0013$).

The mean follow-up costs for the maze procedure arm were £460 more than those for the control arm, mainly because of a difference in readmissions; both treatment arms had large variations around the mean value (see *Table 19*). The mean cost of drugs was £63 less for the maze procedure arm than that for the control arm [the most expensive drugs used in the trial were the calcium channel blockers amlodipine and diltiazem hydrochloride (Dilzem® SR, Cephalon UK Limited, Castleford, UK), followed by acenocoumarol (Sinthrome®, Merus Labs Luxco S.a.R.L., Luxembourg City, Luxembourg)]. For more details of costs per patient, see *Appendix 2*, *Table 23*.

Comparison of *Tables 24* and *25* (*Appendix 2*) shows that, as a result of data completeness, there was very little impact of imputation during primary admission; resource use for both the intervention and control arms increased very slightly. Further details of the total costs for the two treatment arms are given in *Appendix 2*, *Table 26*.

TABLE 19 Mean (SD) of per-patient costs of resource use, with imputation

Health service use	Treatment arm, mean cost (£) per patient (SD)		Difference in cost (£) (maze procedure vs. control)
	Maze ($n = 176$)	Control ($n = 176$)	
Primary admission			
Theatre use	5225 (1594)	4949 (1863)	276
Ablation device	1212 (408)	14 (133)	1197
Adult critical care	4029 (7600)	3065 (5586)	964
Cardiac ward	3397 (4661)	3064 (2014)	333
Rehabilitation	48 (325)	148 (1082)	–100
Acute trust	937 (6105)	165 (1409)	772
Subtotal	14,847 (12,474)	11,404 (7194)	3443
Medication (whole trial period)	618 (1584)	681 (2765)	–63
Follow-up			
Readmissions	1650 (4192)	1220 (2994)	430
Tests	388 (376)	344 (283)	44
Health-care visits	1179 (1061)	1193 (1052)	–14
Subtotal	3217 (5629)	2757 (4329)	460
Grand total	18,681 (13,340)	14,842 (8295)	3841

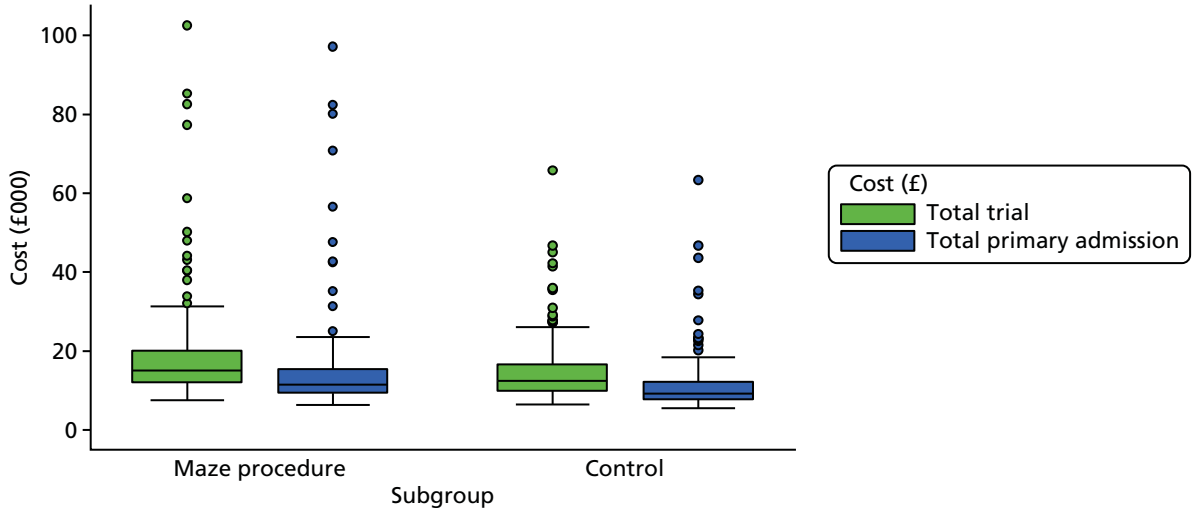


FIGURE 15 Comparison of the total primary admission cost with the total trial cost (with imputation).

Quality of life

Table 20 shows that, in both years, the average QALYs in the maze procedure arm were slightly lower than in the control arm for both the EQ-5D-3L and the SF-6D, although these differences were not statistically significant. The difference in mean EQ-5D-3L QALYs per patient over 2 years was –0.044 (95% CI –0.16 to 0.07 QALYs; $p = 0.44$), which is also less than our a priori estimate of a minimum clinically important difference. Comparison of *Tables 27* and *28* (*Appendix 2*) shows that imputation had no impact on these results.

TABLE 20 Mean (SD) of the QALYs per patient by year of follow-up, with imputation

QALY type	Year of QALY	Treatment arm, mean (SD)	
		Maze procedure ($n = 176$)	Control ($n = 176$)
EQ-5D-3L	QALY year 1	0.7160 (0.2583)	0.7235 (0.2640)
	QALY year 2	0.6896 (0.3066)	0.7274 (0.2964)
SF-6D	QALY year 1	0.6549 (0.2059)	0.6647 (0.2152)
	QALY year 2	0.6418 (0.2486)	0.6699 (0.2374)

Comparing costs and effects

Table 21 shows that the mean cost per patient (with imputation, but without adjustments) was statistically significantly higher for the maze procedure arm, and that mean incremental QALYs per patient were slightly lower (although not statistically significantly different) than those in the control arm. Therefore, the maze procedure arm was dominated by the control arm and, when valuing a QALY at £20,000, the INMB of the maze procedure arm was negative.

Cost-effectiveness

Table 22 shows the primary cost-effectiveness analysis from a seemingly unrelated regression analysis of costs and QALYs, adjusted for baseline differences and correlation between costs and QALYs, with discounting for year 2. The mean difference in the cost of the maze procedure, adjusting for age, sex, baseline EQ-5D-3L score and paroxysmal AF, was £3533 (95% CI £1321 to £5746) higher than for the control intervention, and the difference was statistically significant. The mean difference in QALYs between the maze procedure arm and the control arm, after also controlling for actual procedure used, was –0.022 (95% CI –0.1231 to 0.0791 QALYs), which was not statistically significant. The maze procedure arm was therefore dominated by the control arm.

TABLE 21 Unadjusted comparison of costs and QALYs per patient, with imputation

Cost-effectiveness parameter	Treatment arm, mean (SD)			
	Maze procedure ($n = 176$)		Control ($n = 176$)	
Total cost (£) in year 1	17,834	13,225	13,944	7954
Total cost (£) in year 2 (with discounting)	818	2185	868	1414
Total cost (£) over 2 years (present value)	18,653	–	14,812	–
Incremental cost (£) (maze procedure vs. control)	3841 (95% CI 1514 to 6167)			
EQ-5D-3L QALYs in year 1	0.7160	0.26	0.7235	0.26
EQ-5D-3L QALYs in year 2 (with discounting)	0.6663	0.30	0.7028	0.29
Total QALYs over 2 years (present value)	1.3823	–	1.4263	–
Incremental EQ-5D-3L QALYs (maze procedure vs. control)	–0.04398 (95% CI –0.1558 to 0.0678)			
ICER	Dominated			
INMB (£) at a WTP of £20,000 per QALY	–4720			
INMB (£) at a WTP of £30,000 per QALY	–5160			

TABLE 22 Coefficients and standard errors from seemingly unrelated regression models for costs and QALYs, adjusting for baseline covariates (primary cost-effectiveness analysis)

Dependent variable	Independent variable	Coefficient	SEM	*p*-value
EQ-5D-3L QALYs	Maze procedure arm	–0.0220	0.0516	0.67
	Male	–0.0836	0.0551	0.13
	Age (years)	–0.0045	0.0035	0.20
	Baseline EQ-5D-3L score	0.9369	0.1209	< 0.01
	Paroxysmal AF	–0.1053	0.0599	0.08
	Actual procedure	0.0006	0.0136	0.96
	Constant	1.3544	0.2855	< 0.01
Total cost (£) per patient	Maze procedure arm	3533	1129	< 0.01
	Male	–2131	1205	0.08
	Age (years)	255	75	< 0.01
	Baseline EQ-5D-3L score	–9367	2645	< 0.01
	Paroxysmal AF	2693	1300	0.04
	Constant	–1691	6247	0.79

SEM, standard error of the mean.

The plot of the estimated joint distribution of cost and QALY differences (*Figure 16*) shows that few estimates fall below the line representing £30,000 per QALY, which is currently considered the upper limit for cost-effective health interventions by NICE.[45] The cost-effectiveness acceptability curve (*Figure 17*) shows that, even at a WTP of £70,000 per QALY, the maze procedure has only a 10% probability of being cost-effective, and *Figure 18* shows a continuously declining net monetary benefit of the maze procedure from around –£3500.

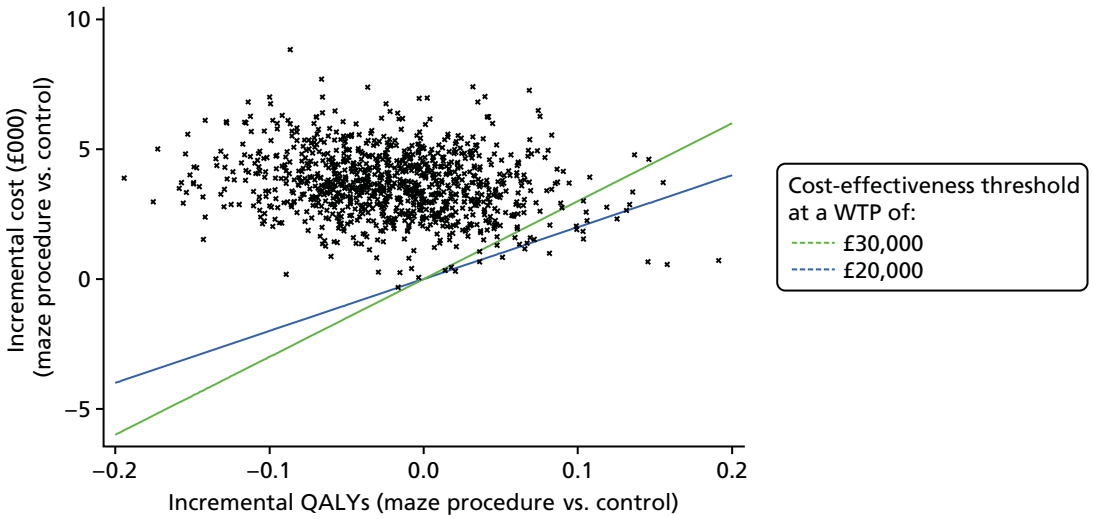


FIGURE 16 Incremental cost-effectiveness plane: the maze procedure arm compared with the control arm.

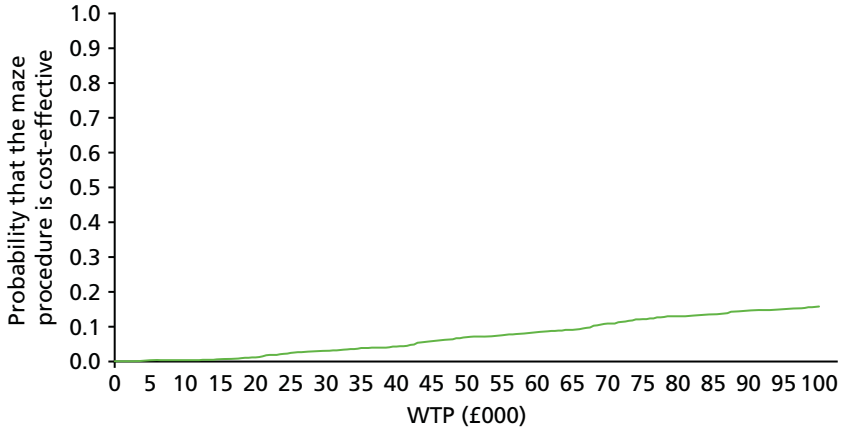


FIGURE 17 Cost-effectiveness acceptability curve for the maze procedure relative to the control arm.

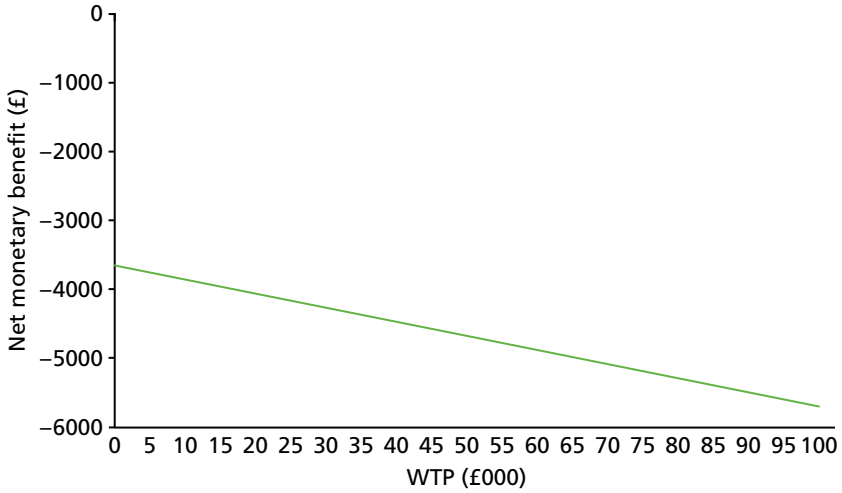


FIGURE 18 Net monetary benefit of the maze procedure arm relative to the control arm.

Tables 29 and *30* in *Appendix 2* show that none of the sensitivity analyses suggested that the maze procedure is likely to be cost-effective at £30,000 per QALY at 2 years. The closest the results came to this was limiting the analysis to patients randomised from April 2011, when the maze procedure had an ICER of £53,538 at 2 years. However, the INMB was negative and the analysis limiting patients to those randomised after April 2011 indicated that the probability of the maze procedure being cost-effective at £30,000 per QALY was < 30% (see the probabilistic sensitivity analyses in *Figure 19*). The final sensitivity analysis (see *Appendix 2*, *Table 31*) shows that, with imputation and controlling for baseline differences, the incremental cost per additional conversion from AF to SR was £25,220.

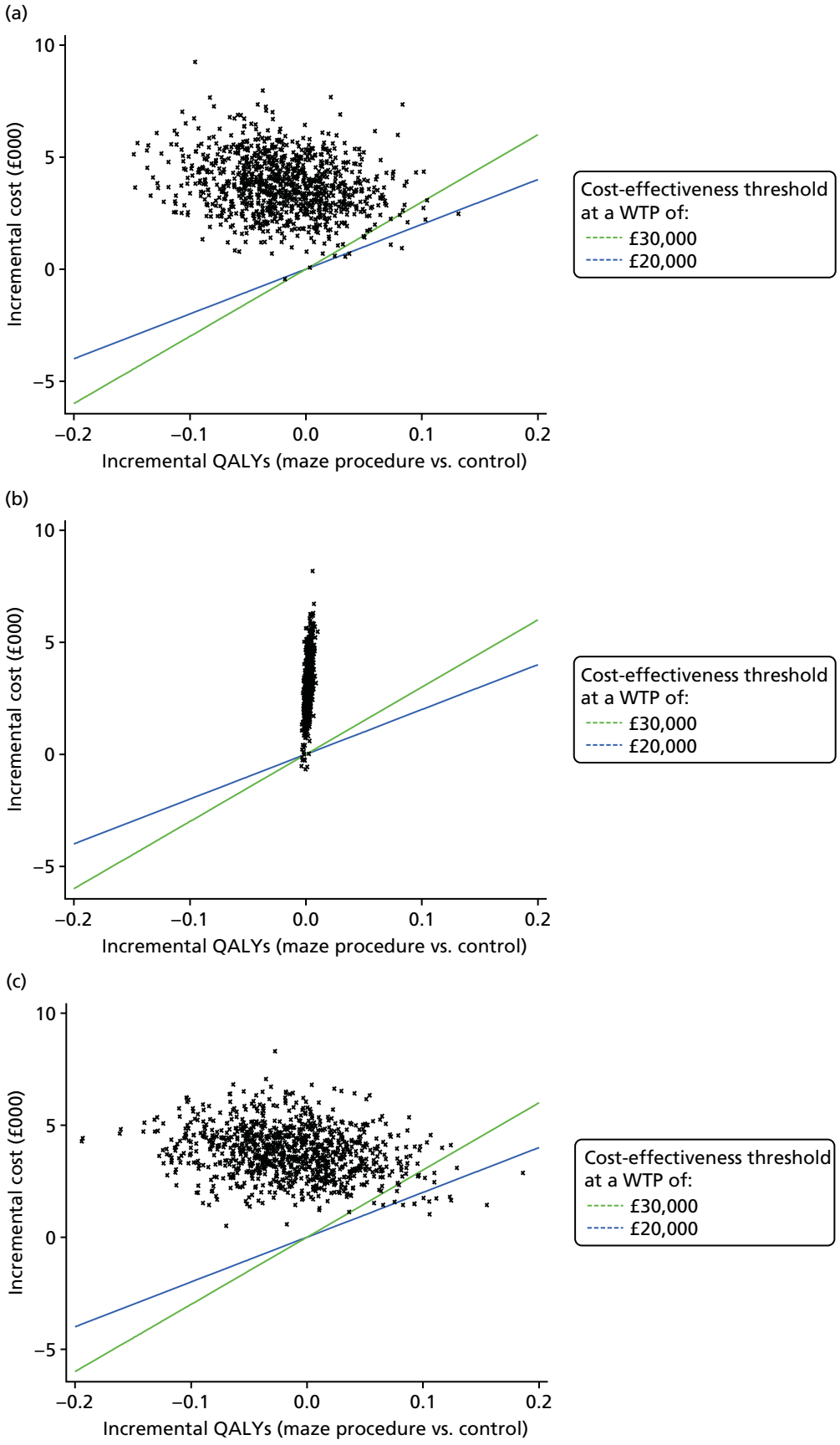


FIGURE 19 Probabilistic sensitivity analyses attached to varying deterministic conditions: (a) using the SF-6D; (b) using primary admission only; (c) using a cost imputation model; (d) excluding ablation device cost; (e) excluding resource-use outliers; (f) using complete-case analysis; and (g) excluding patients randomised before 1 April 2011. Incremental cost = maze procedure cost – control cost; and incremental QALY = maze QALY – control QALY. (*continued*)

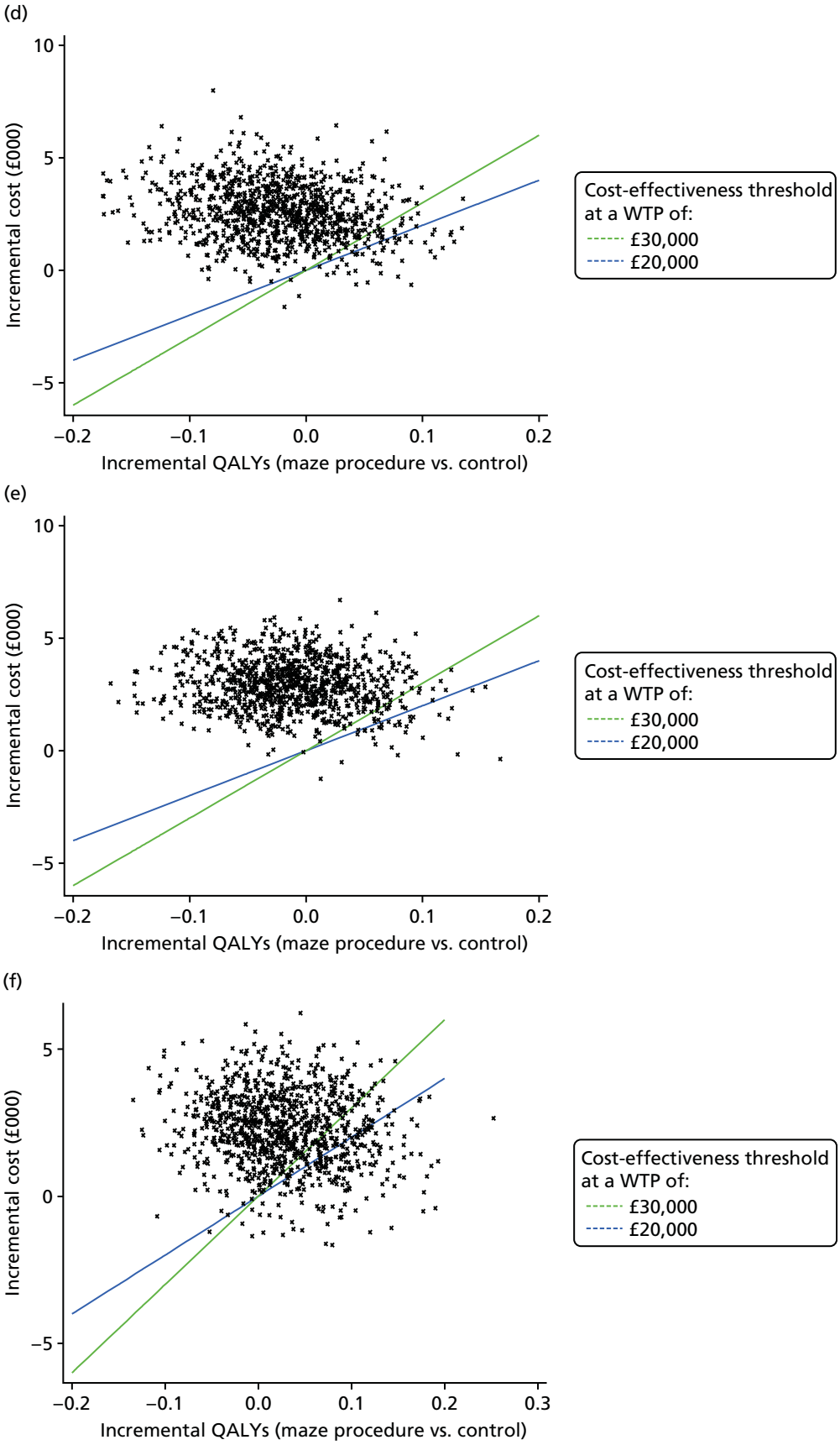


FIGURE 19 Probabilistic sensitivity analyses attached to varying deterministic conditions: (a) using the SF-6D; (b) using primary admission only; (c) using a cost imputation model; (d) excluding ablation device cost; (e) excluding resource-use outliers; (f) using complete-case analysis; and (g) excluding patients randomised before 1 April 2011. Incremental cost = maze procedure cost – control cost; and incremental QALY = maze QALY – control QALY. (*continued*)

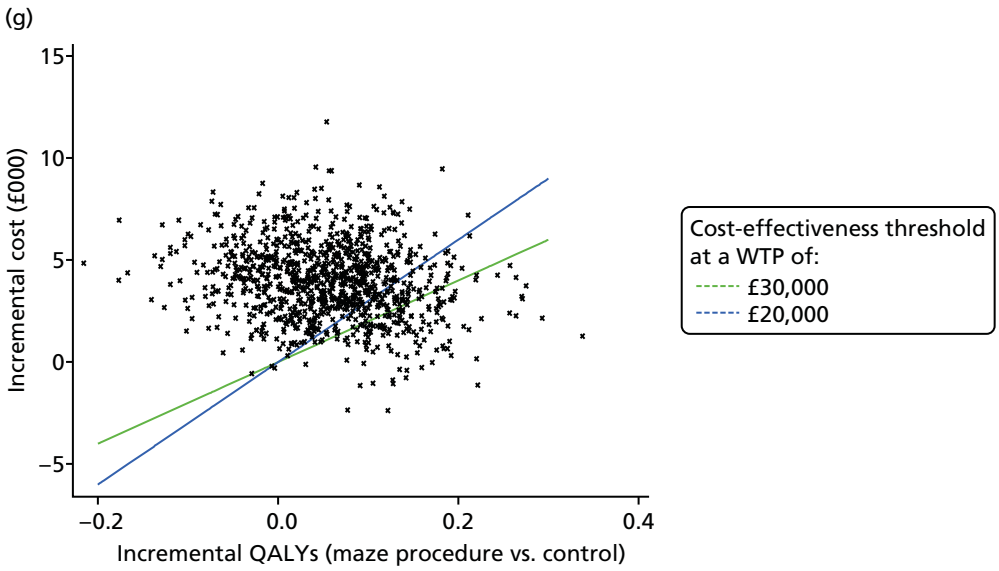


FIGURE 19 Probabilistic sensitivity analyses attached to varying deterministic conditions: (a) using the SF-6D; (b) using primary admission only; (c) using a cost imputation model; (d) excluding ablation device cost; (e) excluding resource-use outliers; (f) using complete-case analysis; and (g) excluding patients randomised before 1 April 2011. Incremental cost = maze procedure cost – control cost; and incremental QALY = maze QALY – control QALY.

Chapter 6 Discussion

Summary of clinical trial results

The objective of the maze procedure as an adjunct to conventional cardiac surgery is to restore SR and thereby improve contractile function. If atrial function is restored, the procedure may reduce the risk of death, stroke and, via reduction in anticoagulant medication, the risk of bleeding. Improvement in patient-reported HRQoL and cost-effectiveness should follow from a reduction in these clinical events.

The Amaze trial has demonstrated a clear and important increase in the clinical primary outcome and the proportion of cardiac patients who returned to SR at 12 months (OR 2.06, 95% CI 1.20 to 3.54; $p = 0.0091$), and that this was maintained and increased at 24 months (OR 3.24, 95% CI 1.76 to 5.96; $p = 0.0001$). Moreover, the odds of returning to SR were greater in maze procedure patients, irrespective of whether patients had paroxysmal or non-paroxysmal AF, for all concomitant cardiac surgery undertaken and did not differ between surgeons (although planned surgery outcomes varied by surgeon). Although these subgroups were small and within-subgroup analyses had low statistical power, this finding adds weight to the causal link between the maze procedure and return to SR. Note that, although the OR for being in SR increased over time after randomisation, this largely resulted from a greater proportion of control patients returning to AF, rather than an increase in prevalence of SR in the maze procedure arm.

Post hoc analysis of the primary outcome showed that the OR for return to SR in early trial patients was lower than for later patients. There may be a number of reasons for this. First, it may have occurred purely by chance. Second, there was some evidence that patients recruited to the trial were more likely to have COPD as the trial progressed, which may have decreased overall success rates. Third, it may be related to greater understanding of the operative technique as the trial progressed, especially the now accepted need for multiple applications of the energy source to guarantee a transmural lesion, which has evolved during the period of the trial.[68] However, detailed examination of changes in outcomes as the trial progressed showed that control patients recruited later in the study had lower rates of return to SR than control patients recruited in the early period. Conversely, there was little difference in return to SR rates throughout the trial for the maze procedure arm. The surgical factor that also changed through the trial was the frequency of left atrial appendage excision, which decreased in control patients and increased in maze procedure patients, with a weakly significant interaction found between operation order and treatment group (the OR for left atrial appendage excision for maze procedure patients, relative to control patients, was 1.34 per 50 patients recruited; $p = 0.0244$). This divergence in excision rates between the arms appeared to be more common for specific surgeons who recruited more patients in the second half of the trial. The implication may be that the maze procedure is equally effective at interrupting AF in all patients, but there is substantial heterogeneity in results for control patients. Assiduous attention to cointerventions, such as left atrial appendage excision and cardioversion, is also effective, and at least some of the effect of the maze procedure may be explained by the use of these cointerventions. We stress that these post hoc exploratory analyses should be interpreted with caution, as there was little power to detect changes over time, and many patient and operative characteristics are highly correlated.

The maze procedure did not result in an increase in AEs, although 11 procedures were not completed, mainly because of technical problems or concerns regarding patient safety. There was also evidence that maze procedure patients were more likely to have their anticoagulant drugs stopped, which may result in fewer adverse bleeding events in the future.

In this pragmatic trial, there was no mandated mode of delivery of the maze procedure; radiofrequency ablation, usually bipolar, was the method of choice for a large proportion of the surgeons in the trial.

The Amaze trial did not identify any clear differences in return to SR rates between methods, but the numbers of cases in these subgroups were small and these comparisons were not robust.

The lesion set treated was at the discretion of the operating surgeon, and there did not appear to be a clear relationship between the number or location of treatments and the odds of return to SR, although the power to detect these interactions was very low. The Amaze trial has confirmed that there appears to be still no practised consensus as to the optimal lesion set to be used in the adjunct maze procedure; the lesion set appeared to be more often determined by the individual surgeon and by the operative scenario than by firm guidelines and an operative plan. Nevertheless, the highest likelihood of SR restoration was observed after a left atrial ablation procedure that included the mitral isthmus lesion. Beyond that, we could discern no additional advantage of adding right atrial lesions.

In the HESTER substudy, the 95% CI for the mean difference in ALAEF, between patients in SR following the maze procedure and matched control patients who were in SR both before and after surgery, was within predefined clinically equivalent limits, although the ALAEF was statistically significantly lower in maze procedure patients. This is an important finding, as return to SR with no or minimal recovery of atrial function may not reduce the risk of thromboembolism. Absence of atrial contractility has been reported in around one-third of patients who return to SR following ablation (using a variety of surgical and catheter ablation techniques) and trials examining the return of atrial function have produced conflicting results.[65] Some reported the return of haemodynamically important function, but others suggested that the maze procedure is unlikely to achieve functional recovery, despite restoring SR.[69]

It is disappointing that we found no significant differences in survival and QoL between the maze procedure and control arms at 2 years. As all patients in the Amaze trial also had conventional cardiac surgery, marked improvements were observed in the NYHA status and QoL parameters in both treatment arms from 6 months after surgery, and most of these improvements must be attributable to the conventional cardiac operation. The fact that we found no additional HRQoL benefit attributable to the maze procedure may be because of the absence of such benefit, or because such benefit exists, but is small and requires a longer follow-up period. If the rate of SR restoration falls further in the control arm as time progresses, and if there is true recovery of atrial contractility in the restored SR group, then it is reasonable to conjecture that differences in QoL may be seen in future. Cardiac surgery alone is known to produce a marked increase in QoL within months of the procedure.[70] If AF surgery and restoration of SR have a positive impact on QoL, this may only be seen in the longer term, as patients no longer in AF are less exposed to the risks of stroke, anticoagulant drug complications, antiarrhythmic complications and progressive heart failure. The finding in the HESTER substudy of functional and contractile atria is an indication that such a benefit is possible and may be seen with longer-term follow-up. A long-term health economics model was not funded by the original grant. However, continued follow-up of clinical events and HRQoL is in progress and will inform a long-term analysis.

Comparison with existing evidence

Overall, our SR restoration rate after maze procedure was consistent with other RCTs, but lower than in some published series that represent selected practitioners treating selected patients.[9] The rate of SR restoration in the control arm was much higher than expected, and remained so, despite the substantial drop at 2 years. This is a sufficient indication for considering cardioversion in AF patients after conventional cardiac surgery alone, as over one-third of patients can be in SR at 2 years.

The two adverse features of asymptomatic AF are the impact on cardiac function and the risk of thromboembolism.[71] Both are directly related to atrial function. The HESTER substudy provided evidence that atrial contractile function increases as a result of the maze procedure; therefore, restoring SR with a maze procedure might provide clinical benefits in the future.

Return to SR with no or minimal recovery of atrial function may have no benefit to the patient with respect to the risk of thromboembolism. Left atrial transport function depends on atrial contractility, atrial synchrony and left ventricular diastolic function. Absence of atrial contractility has been reported in around one-third of patients who return to SR following ablation (using a variety of surgical and catheter ablation techniques).[65] Buber *et al.*[65] reported that the absence of left atrial contraction was associated with a significant increase in the risk of thromboembolic stroke after the maze procedure for patients in SR. Conversely, effective left atrial transport function may be associated with reduced morbidity after successful ablation of AF.[65]

Trials examining the return of atrial function have produced conflicting results. Some have reported the return of haemodynamically meaningful function, but others have suggested that the maze procedure is unlikely to achieve functional recovery, despite restoring SR.[69] In patients without atrial contraction, anticoagulant drugs significantly reduce stroke rate, but the incidence of major bleeding is increased.[72,73] Furthermore, patients with chronic AF have evidence of persistent left atrial dysfunction, even after restoration of SR by radiofrequency ablation. This suggests that global and regional atrial dysfunction may be attributable to a combination of injury from the ablation process and pre-existing disease.[66,74]

Patients want to know whether or not they can safely stop taking anticoagulants after SR is restored by a maze procedure. This requires long-term follow-up and stroke surveillance to be addressed with any certainty. However, the HESTER substudy results lend some support to those advocating or already practising anticoagulant drug withdrawal. The varying rates of left atrial functional recovery after maze procedure mean that it would be prudent to measure atrial function before considering withdrawal of anticoagulant drugs.

Summary and implications of cost-effectiveness results

Given the clinical results, it is perhaps not surprising that the per-patient costs over 2 years in the maze procedure arm were statistically significantly higher than those in the control arm (£3533, 95% CI £1321 to £5746) and that there was a small non-significant reduction in discounted QALYs (–0.022, 95% CI –0.1231 to 0.0791 QALYs). With higher costs and no improvement in QoL over the 2 years of the trial, the control arm, representing current practice, dominates the new maze procedure for patients with AF within a 2-year period.

Both the deterministic and probabilistic sensitivity analyses confirmed this conclusion, showing that, within a 2-year period, the probability that the maze procedure would be considered cost-effective was < 5%. Moreover, a variety of alternative and favourable assumptions would not change the overall conclusion that, compared with current practice in the control arm, the maze procedure is not a cost-effective intervention up to 2 years. A potential caveat to this was the result from an unplanned post hoc subgroup analysis that included only the later patients ($n = 200$), in which the incremental effect was positive and the ICER fell to £53,500.

To date, results from the wider literature on the cost per QALY of ablation surgery in patients with AF are somewhat mixed. This first trial-based economic evaluation found that, comparing an adjunct ablation surgery with cardiac surgery over a 1-year period, an additional cost of €4426 and a mean 0.06 QALY gain translated to an ICER of €73,359.[75] The associated probabilistic sensitivity analysis indicated that, at a WTP per QALY of €30,000, the probability that the add-on ablation surgery was cost-effective was around 10%, and, therefore, was not considered cost-effective at 1 year. Our 2-year results are therefore in line with these findings. Longer-term economic models have shown that, over 5 years, either a classic 'cut-and-sew' maze procedure or a high-intensity ultrasound-assisted surgical ablation procedure, in addition to scheduled CABG or valve surgery, was highly cost-effective compared with drug treatment.[76] Cost-effectiveness decreased slightly if the intervention was applied with a percutaneous procedure. However, economic modelling to inform Canadian decision-makers indicated that the cost-effectiveness

of AF ablation compared with antiarrhythmic medication was more favourable over longer time horizons (e.g. 10 years) and could dominate antiarrhythmic medication given time horizons of 20 years; this was not the case when considering a 5-year time horizon.[77] English decision-makers[78] concluded that, having reviewed McKenna *et al.*[79] and Rodgers *et al.*,[80] duration of benefit is a key determinant of cost-effectiveness and, therefore, rejected van Breugel *et al.*[75] as having too short a time period for appropriate decision-making. This suggests that costs and effects should be examined over a longer time period, and preferably for longer than 5 years.

A 2013 systematic review, which included each of these studies, concluded that there was insufficient evidence to support catheter ablation as a first-line treatment, and that there was mixed evidence for second-line ablation.[81] The authors specifically lament the lack of QoL data, especially in the long term, and highlight the critical importance of this in determining cost-effectiveness. They reflect on the shortcomings of existing economic studies and, without hard evidence on such key outcomes in the medium term (1–5 years), successfully counselled the Belgian government not to extrapolate beyond a 5-year term, because of uncertainties. This highlights the importance of the planned long-term follow-up of patients from the Amaze trial and the relevance of these future data to extending economic analysis into the long term. Interestingly, an industry-sponsored, 5-year economic model published in 2014 reported that, although catheter ablation and the convergent procedure were argued to be cost-effective relative to medical management for non-paroxysmal AF, this was dependent on key assumptions about long-term maintenance of SR beyond 2 years and its relationship to future QoL.[82]

Strengths and weaknesses

The Amaze trial was ground-breaking in being a relatively large, multicentre Phase III trial in a surgical setting, in which randomisation was completed during surgery, and patients, investigators and clinical staff (with the exception of the surgical team) were blinded to treatment allocation. Details of the randomised arm were sealed and released only to the clinical teams in the event of a SAE that required unblinding. Complete blinding was achieved for patients and HRQoL assessors, and only 13 patients had their treatment allocation open to the clinical teams caring for the patients, thus supporting the integrity of the trial methodology. A further strength was the longer follow-up period, which incorporated quality-adjusted survival as a co-primary outcome.

Screening logs were completed assiduously only at the co-ordinating centre, and reasons for exclusion from the trial were not always transparent, which affected the interpretation of the generalisability of the results. However, comparison of baseline characteristics with those reported in annual audit statistics and registries suggests that trial patients were slightly older and more likely to be female, and that they had a slightly lower average EuroSCORE, but otherwise were broadly representative of NHS-treated cardiac surgery patients.[62]

In common with other cardiac surgical trials, prolongation of hospital stay and readmissions were common, and this is reflected in the number of SAEs recorded. There were no differences between the maze procedure and non-maze procedure patients in the number of events, their severity or relation to treatment, or in the number of patients who had events. Slow recruitment is a widespread problem in RCTs in general, and surgical trials in particular, and it is unfortunate that the Amaze trial did not achieve the target recruitment level.

In the study design, recruitment and randomisation of 400 patients were expected to be complete in 18 months. Delays in obtaining trust approvals at the recruitment sites ranged from 5 to 16 months, so that centre set-up took much longer than anticipated. In addition, a high number of staff changes across many sites, including withdrawal of Comprehensive Local Research Network support for the study at Papworth Hospital, had a significant impact on the recruitment rate.

During the trial, it became apparent that most local centres significantly overestimated their surgical activity and/or the recruitment rate for the Amaze trial. Moreover, there was heightened awareness of the maze procedure among both patients, as a result of their own personal research, and clinicians, which affected equipoise of potential recruits and decreased the number of patients approached to participate in the trial. The age distribution of the population that would be eligible for the trial also changed over time, with the proportion of elderly patients (aged $\geq$ 80 years) increasing. These patients were generally supportive of the research, but often declined to participate because of the need for attendance at follow-up visits and the presence of comorbidities. This was particularly true for patients living in rural areas, despite the provision of alternative methods of travel, such as taxis.

Recruitment was carefully monitored by the project management team throughout the trial. At every TSC meeting, a recruitment recovery plan was presented by the project team and reviewed. Specific measures taken to improve recruitment were as follows: (1) there was an increase in the number of sites, from 8 to 11; (2) all sites were encouraged to include all consultants who had the required experience of the maze procedure in recruitment; (3) contracts were revised to encourage 'competitive recruitment' across sites; and (4) extensive trial promotional material was produced to raise the profile and to encourage 'ownership' of the trial. Despite these measures, target recruitment was not reached within the funded time period.

Around 26% of the randomised patients ($n = 92$) had paroxysmal AF at baseline, and we found that being in SR during baseline ECG monitoring was strongly related to being in SR at annual follow-ups. Although all patients had a documented history of AF at baseline and satisfied the inclusion criteria for the trial, and the number of patients with paroxysmal AF at baseline were equally distributed between the two treatment arms, this could have diluted the treatment effect to some extent.

The decision by the independent DMEC to recommend that the TSC consider stopping the trial on the basis of futility meant that unblinded trial summaries were available to TSC members. We cannot rule out bias as a result of the knowledge that treatment effects were not promising at this time. Individual patients and the investigators recording primary outcomes remained unaware of group allocation, so we believe that any bias was minor. However, it is possible that clinicians drifted towards recruitment of patients with more severe AF, which in turn resulted in a larger treatment effect in the second half of the trial.

There were three follow-up data collection points (6, 12 and 24 months). Having longer gaps over which patients are asked to recall resource use has been shown to link to under-reporting of frequent events, severity of illness and those using services very intensively.[83] Despite this, neither Johnston *et al.*[83] nor Ridyard and Hughes[84] recommended a specific interval between data collection points, but they did suggest compromising between respondent burden and collecting data on the incidence of events that drive resource use. Recent evidence published by Seidl *et al.*[85] reported that, in a 1-year trial evaluating management of acute MI in patients aged > 65 years, in one German hospital, three data collection points gave similar results to four data collection points.[85] In the Amaze trial, almost 95% of the difference in follow-up costs between trial arms related to admissions; as these are major and infrequent events, it is unlikely that their costs have been underestimated. There was lower resource use for primary and community care resources in the second year than in the first year, and although this could be a function of the frequency of data collection, it could also be a function of higher expected use of outpatient services following initial admission. Even so, it is not clear this would affect the incremental cost difference and, therefore, change the findings.

There was no statistically significant effect for the patient-centred primary outcome, undiscounted QALYs over 2 years, and the 95% CI for the difference between the treatment arms (maze procedure vs. control) excluded our prespecified minimum clinically important difference of 0.083 QALYs. More specifically, based on the 95% CI, we could rule out differences in QALYs of $\geq$ 0.083 in favour of the maze procedure arm. A long-term economic model may demonstrate a delayed effect on patient-centred outcomes, but it was outside the scope of this report.

The HESTER substudy was small and not randomised, so the results regarding restoration of contractile function were less robust. Echocardiography measurements of left atrial volumes, despite being well correlated with MRI measurements, tend to be underestimations; therefore, estimated treatment effects from MRI measurements were larger and had wider CIs than the corresponding estimates from echocardiography measurements.[86,87] Implications for the analyses of ALAEF, for which we do not have MRI measurements, are difficult to assess. Our a priori definition of clinical equivalence in the HESTER substudy was based on the SD (18.2%) of normal ALAEF reported in a published study, along with the investigators' clinical judgement.[88] However, the SD in our sample was lower (< 8%). This may have resulted from an underestimation of the population SD in our cohort, but could be a consequence of our sample being a more homogeneous group of patients, which may limit the generalisability of our results.

There were 49 maze procedures at Papworth Hospital outside the trial; these patients either had severely symptomatic AF, so that surgeons were not in equipoise for these patients, or the patients themselves or their surgeon had a strong preference for the treatment, despite a lack of robust evidence. Thus, our results do not necessarily apply to these patients.

Implications for service

- Ablation can be safely practised in a NHS routine cardiac surgical setting and will increase the number of patients who return to SR after surgery.
- There is some support for anticoagulant drug withdrawal, but the varying rates of left atrial functional recovery after a maze procedure suggest that measurement of atrial function before considering withdrawal of anticoagulant drugs would be prudent.
- The improvement in SR restoration as the study progressed suggests better patient selection and better maze procedures, including cointerventions (e.g. excision of the left atrial appendage), were performed in the later stages. This may reflect greater understanding of the lesion set and greater success in achieving truly transmural lesions. Surgeons performing the procedure may wish to audit their SR restoration rate and modify their practice if necessary.

Implications for further research

- Continued monitoring of clinical events (deaths, strokes and haemorrhaging), resource use (especially readmissions) and HRQoL to inform a long-term health economics model is required, and for at least 5 years after surgery.
- Further study of the relative effectiveness of different ablation techniques would clarify the choice of methodology.

Conclusions

In the Amaze trial, the maze procedure, as an adjunct to routine cardiac surgery, was safe and effective in increasing the probability of restoration of SR in patients with pre-existing paroxysmal or non-paroxysmal (persistent, longstanding or chronic) AF. The odds of being in SR approximately doubled at 12 months, and more than tripled at 24 months, after surgery. There was evidence from a small, non-randomised substudy that this was associated with improved atrial contractile function that was lower, but within the limits of clinical equivalence, when compared with patients who were in SR before surgery. The additional cost of the maze procedure per person over 2 years, relative to routine cardiac surgery, was £3500. In the 2-year follow-up period, these clinical effects had not resulted in additional improvement in HRQoL; however, given the known association between AF and subsequent neurological events and the reduced use of anticoagulants in the maze procedure arm, continued follow-up and analysis is warranted.

Acknowledgements

The trial was funded by the NIHR HTA programme, and was carried out while the first author was employed as a programme leader at the Medical Research Council Biostatistics Unit and, later, as a Division Director at the Leeds Institute of Clinical Trials Research.

We are grateful to the independent members of the TSC [Dr Dai Rowlands (chairperson), Consultant Cardiologist, Peterborough Hospital (retired); Mr Brian Elliott, independent lay member; and Professor Paul Kinnersley, School of Medicine, Cardiff University, Cardiff, UK] and the DMEC [Mr James Roxburgh (chairperson), Consultant Cardiothoracic Surgeon, Guy's and St Thomas' Hospital, London, UK; Dr Derick Todd, Consultant Cardiac Electrophysiologist, Liverpool Heart and Chest Hospital, Liverpool, UK; Professor Mark Sculpher, Health Economist, University of York, York, UK; and Dr Fay Cafferty, Statistician, Wolfson Institute of Preventative Medicine, London, UK] for their guidance throughout the trial.

Furthermore, we are grateful to Matt Glover for the initial health economics input, Sheridan Carter for the ECG analysis, Donna Alexander for conducting the trial, Helen Holcombe and Hannah Munday for project co-ordination and all research nurses, theatre and ward staff and patients.

The trial was conducted in collaboration with the Papworth Trials Unit Collaboration, the staff of which supported the trial design, conduct and analysis, including data management (Christine Mills, Clinical Project Manager), clinical trial operations and quality assurance (Christine Mills and Victoria Stoneman, Clinical Project Managers).

Contributions of authors

Linda Sharples (Professor of Medical Statistics) was a grant applicant, who designed the original study, led the statistical and trial methodology, oversaw the statistical analysis and drafted the manuscript. She is the guarantor for the statistical analysis.

Colin Everett (Senior Trial Statistician) analysed the Amaze trial data and contributed to the drafting of the manuscript.

Jeshika Singh (Research Fellow, Health Economics) carried out the cost-effectiveness analysis and contributed to the drafting of the manuscript.

Christine Mills (Clinical Project Manager, Project Delivery) contributed to the design, patient recruitment and project management of the trial at all centres, and also contributed to the drafting of the manuscript.

Tom Spyt (Consultant Cardiac Surgeon) contributed to the design, trial protocol, patient recruitment and management and manuscript review.

Yasir Abu-Omar (Consultant Cardiac Surgeon) contributed to patient recruitment and management and also to the drafting of the chapter on the HESTER substudy and manuscript review.

Simon Fynn (Consultant Cardiologist) contributed to the design, trial protocol, patient recruitment and management, and led the analysis of the ECG results and reviewed the manuscript.

Benjamin Thorpe (Research Methods Fellow, Statistics) analysed the HESTER substudy data and contributed to the drafting of the manuscript.

Victoria Stoneman (Clinical Project Manager, Project Delivery) contributed to the grant application and protocol for the HESTER substudy, the trial management of the HESTER substudy and the Amaze trial and the drafting of the manuscript.

Hester Goddard (Trial Manager, retired) was a grant applicant, who contributed to the protocols, managed the trial during set-up and the first 4 years and reviewed the manuscript.

Julia Fox-Rushby (Professor of Health Economics) led the health economics analysis, reviewed the methodology and contributed to the drafting of the manuscript. She is the guarantor for the cost-effectiveness analysis.

Samer Nashef (Chief Investigator, Consultant Cardiac Surgeon) was a grant applicant, who conceived and designed the original study, led the surgical methodology, oversaw surgical trial activity across all centres and drafted the manuscript. He is the overall guarantor for the study.

All authors reviewed the final version of the monograph.

Publications

Abu-Omar Y, Thorpe BS, Freeman C, Mills C, Stoneman VEA, Gopalan D, *et al.* Recovery of left atrial contractile function after maze surgery in persistent longstanding atrial fibrillation. *J Am Coll Cardiol* 2017;**70**:2309–11.

Nashef SAM, Fynn S, Abu-Omar Y, Spyt TJ, Mills C, Everett CC, *et al.* Amaze: a randomised controlled trial of adjunct surgery for atrial fibrillation [published online ahead of print 17 April 2018]. *Eur J Cardiothorac Surg* 2018. https://doi.org/10.1093/ejcts/ezy165

Data sharing statement

Data sets collected and/or analysed during the Amaze trial will be available from the corresponding author on reasonable request, after all trial publications have been accepted, provided that the TSC agrees and existing patient consent is consistent with the request.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: https://understandingpatientdata.org.uk/data-citation.

References

1. Mann DL, Zipes DP, Libby P, Bonow RO, Braunwald E. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. 10th edn. Philadelphia, PA: Saunders; 2014.
2. Camm CF, Nagendran M, Xiu PY, Maruthappu M. How effective is cryoablation for atrial fibrillation during concomitant cardiac surgery? *Interact Cardiovasc Thorac Surg* 2011;**13**:410–14. https://doi.org/10.1510/icvts.2011.271676
3. Naccarelli GV, Varker H, Lin J, Schulman KL. Increasing prevalence of atrial fibrillation and flutter in the United States. *Am J Cardiol* 2009;**104**:1534–9. https://doi.org/10.1016/j.amjcard.2009.07.022
4. Hobbs FD, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, *et al.* A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study. *Health Technol Assess* 2005;**9**(40). https://doi.org/10.3310/hta9400
5. Sundt TM, Gersh BJ. Making sense of the maze: which patients with atrial fibrillation will benefit? *JAMA* 2005;**294**:2357–9. https://doi.org/10.1001/jama.294.18.2357
6. National Institute for Health and Care Excellence (NICE). *Atrial Fibrillation: Management*. Clinical Guideline [CG180]. London: NICE; 2006.
7. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991;**22**:983–8. https://doi.org/10.1161/01.STR.22.8.983
8. Calkins H, Brugada J, Packer DL, Cappato R, Chen SA, Crijns HJ, *et al.* HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for personnel, policy, procedures and follow-up. A report of the Heart Rhythm Society (HRS) Task Force on catheter and surgical ablation of atrial fibrillation. *Heart Rhythm* 2007;**4**:816–61. https://doi.org/10.1016/j.hrthm.2007.04.005
9. Huffman MD, Karmali KN, Berendsen MA, Andrei AC, Kruse J, McCarthy PM, *et al.* Concomitant atrial fibrillation surgery for people undergoing cardiac surgery. *Cochrane Database Syst Rev* 2016;**22**:CD011814. http://dx.doi.org/10.1002/14651858.CD011814.pub2.
10. de Lima GG, Kalil RA, Leiria TL, Hatem DM, Kruse CL, Abrahao R, *et al.* Randomized study of surgery for patients with permanent atrial fibrillation as a result of mitral valve disease. *Ann Thorac Surg* 2004;**77**:2089–94. https://doi.org/10.1016/j.athoracsur.2003.11.018
11. Doukas G, Samani NJ, Alexiou C, Oc M, Chin DT, Stafford PG, *et al.* Left atrial radiofrequency ablation during mitral valve surgery for continuous atrial fibrillation: a randomized controlled trial. *JAMA* 2005;**294**:2323–9. https://doi.org/10.1001/jama.294.18.2323
12. Gillinov AM, Gelijns AC, Parides MK, DeRose JJ, Moskowitz AJ, Voisine P, *et al.* Surgical ablation of atrial fibrillation during mitral-valve surgery. *N Engl J Med* 2015;**372**:1399–409. https://doi.org/10.1056/NEJMoa1500528
13. Haïssaguerre M, Jaïs P, Shah DC, Takahashi A, Hocini M, Quiniou G, *et al.* Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med* 1998;**339**:659–66. https://doi.org/10.1056/NEJM199809033391003
14. Cox JL, Boineau JP, Schuessler RB, Ferguson TB, Cain ME, Lindsay BD, *et al.* Successful surgical treatment of atrial fibrillation. Review and clinical update. *JAMA* 1991;**266**:1976–80. https://doi.org/10.1001/jama.1991.03470140088029

15. Shemin RJ, Cox JL, Gillinov AM, Blackstone EH, Bridges CR. Workforce on Evidence-Based Surgery of the Society of Thoracic Surgeons. Guidelines for reporting data and outcomes for the surgical treatment of atrial fibrillation. *Ann Thorac Surg* 2007;**83**:1225–30. https://doi.org/10.1016/j.athoracsur.2006.11.094

16. Malaisrie SC, Lee R, Kruse J, Lapin B, Wang EC, Bonow RO, *et al.* Atrial fibrillation ablation in patients undergoing aortic valve replacement. *J Heart Valve Dis* 2012;**21**:350–7.

17. Badhwar V, Rankin JS, Damiano RJ, Gillinov AM, Bakaeen FG, Edgerton JR, *et al.* The Society of Thoracic Surgeons 2017 clinical practice guidelines for the surgical treatment of atrial fibrillation. *Ann Thorac Surg* 2017;**103**:329–41. https://doi.org/10.1016/j.athoracsur.2016.10.076

18. Lall SC, Melby SJ, Voeller RK, Zierer A, Bailey MS, Guthrie TJ, *et al.* The effect of ablation technology on surgical outcomes after the Cox–maze procedure: a propensity analysis. *J Thorac Cardiovasc Surg* 2007;**133**:389–96. https://doi.org/10.1016/j.jtcvs.2006.10.009

19. Schuetz A, Schulze CJ, Sarvanakis KK, Mair H, Plazer H, Kilger E, *et al.* Surgical treatment of permanent atrial fibrillation using microwave energy ablation: a prospective randomized clinical trial. *Eur J Cardiothorac Surg* 2003;**24**:475–80. https://doi.org/10.1016/S1010-7940(03)00377-4

20. Abreu Filho CA, Lisboa LA, Dallan LA, Spina GS, Grinberg M, Scanavacca M, *et al.* Effectiveness of the maze procedure using cooled-tip radiofrequency ablation in patients with permanent atrial fibrillation and rheumatic mitral valve disease. *Circulation* 2005;**112**(Suppl. 9):I20–5.

21. Akpinar B, Sanisoglu I, Guden M, Sagbas E, Caynak B, Bayramoglu Z. Combined off-pump coronary artery bypass grafting surgery and ablative therapy for atrial fibrillation: early and mid-term results. *Ann Thorac Surg* 2006;**81**:1332–7. https://doi.org/10.1016/j.athoracsur.2005.09.074

22. Deneke T, Khargi K, Grewe PH, Laczkovics A, von Dryander S, Lawo T, *et al.* Efficacy of an additional MAZE procedure using cooled-tip radiofrequency ablation in patients with chronic atrial fibrillation and mitral valve disease. A randomized, prospective trial. *Eur Heart J* 2002;**23**:558–66. https://doi.org/10.1053/euhj.2001.2841

23. McCarthy PM, Manjunath A, Kruse J, Andrei AC, Li Z, McGee EC Jr, *et al.* Should paroxysmal atrial fibrillation be treated during cardiac surgery? *J Thorac Cardiovasc Surg* 2013;**146**:810–23. https://doi.org/10.1016/j.jtcvs.2013.05.015

24. Calkins H, Kuck KH, Cappato R, Brugada J, Camm AJ, Chen SA, *et al.* 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. *Europace* 2012;**14**:528–606. https://doi.org/10.1093/europace/eus027

25. Cherniavsky A, Kareva Y, Pak I, Rakhmonov S, Pokushalov E, Romanov A, *et al.* Assessment of results of surgical treatment for persistent atrial fibrillation during coronary artery bypass grafting using implantable loop recorders. *Interact Cardiovasc Thorac Surg* 2014;**18**:727–31. https://doi.org/10.1093/icvts/ivu016

26. Jessurun ER, van Hemel NM, Defauw JJ, Brutel De La Rivière A, Stofmeel MA, Kelder JC, *et al.* A randomized study of combining maze surgery for atrial fibrillation with mitral valve surgery. *J Cardiovasc Surg* 2003;**44**:9–18.

27. Van Breugel HN, Nieman FH, Accord RE, Van Mastrigt GA, Nijs JF, Severens JL, *et al.* A prospective randomized multicenter comparison on health-related quality of life: the value of add-on arrhythmia surgery in patients with paroxysmal, permanent or persistent atrial fibrillation undergoing valvular and/or coronary bypass surgery. *J Cardiovasc Electrophysiol* 2010;**21**:511–20. https://doi.org/10.1111/j.1540-8167.2009.01655.x

28. Akpinar B, Guden M, Sagbas E, Sanisoglu I, Ozbek U, Caynak B, *et al.* Combined radiofrequency modified maze and mitral valve procedure through a port access approach: early and mid-term results. *Eur J Cardiothorac Surg* 2003;**24**:223–30. https://doi.org/10.1016/S1010-7940(03)00258-6

29. Albrecht A, Kalil RA, Schuch L, Abrahao R, Sant'Anna JR, de Lima G, *et al.* Randomized study of surgical isolation of the pulmonary veins for correction of permanent atrial fibrillation associated with mitral valve disease. *J Thorac Cardiovasc Surg* 2009;**138**:454–9. https://doi.org/10.1016/j.jtcvs.2009.04.023

30. Blomstrom-Lundqvist C, Johansson B, Berglin E, Nilsson L, Jensen SM, Thelin S, *et al.* A randomized double-blind study of epicardial left atrial cryoablation for permanent atrial fibrillation in patients undergoing mitral valve surgery: the SWEDish Multicentre Atrial Fibrillation study (SWEDMAF). *Eur Heart J* 2007;**28**:2902–8. https://doi.org/10.1093/eurheartj/ehm378

31. Budera P, Straka Z, Osmancik P, Vanek T, Jelinek S, Hlavicka J, *et al.* Comparison of cardiac surgery with left atrial surgical ablation vs. cardiac surgery without atrial ablation in patients with coronary and/or valvular heart disease plus atrial fibrillation: final results of the PRAGUE-12 randomized multicentre study. *Eur Heart J* 2012;**33**:2644–52. https://doi.org/10.1093/eurheartj/ehs290

32. Chevalier P, Leizorovicz A, Maureira P, Carteaux JP, Corbineau H, Caus T, *et al.* Left atrial radiofrequency ablation during mitral valve surgery: a prospective randomized multicentre study (SAFIR). *Arch Cardiovasc Dis* 2009;**102**:769–75. https://doi.org/10.1016/j.acvd.2009.08.010

33. Jönsson A, Lehto M, Ahn H, Hermansson U, Linde P, Ahlsson A, Laczkovics A. Microwave Ablation in Mitral Valve Surgery for Atrial Fibrillation (MAMA). *J Atr Fibrillation* 2012;**5**:432.

34. Khargi K, Deneke T, Haardt H, Lemke B, Grewe P, Müller KM, *et al.* Saline-irrigated, cooled-tip radiofrequency ablation is an effective technique to perform the maze procedure. *Ann Thorac Surg* 2001;**72**:S1090–5. https://doi.org/10.1016/S0003-4975(01)02940-X

35. Knaut M, Kolberg S, Brose S, Jung F. Epicardial microwave ablation of permanent atrial fibrillation during a coronary bypass and/or aortic valve operation: prospective, randomised, controlled, mono-centric study. *Appl Cardiopulm Pathophysiol* 2010;**14**:220–8.

36. Pokushalov E, Romanov A, Corbucci G, Cherniavsky A, Karaskov A. Benefit of ablation of first diagnosed paroxysmal atrial fibrillation during coronary artery bypass grafting: a pilot study. *Eur J Cardiothorac Surg* 2012;**41**:556–60. https://doi.org/10.1093/ejcts/ezr101

37. Srivastava V, Kumar S, Javali S, Rajesh TR, Pai V, Khandekar J, *et al.* Efficacy of three different ablative procedures to treat atrial fibrillation in patients with valvular heart disease: a randomised trial. *Heart Lung Circ* 2008;**17**:232–40. https://doi.org/10.1016/j.hlc.2007.10.003

38. Vasconcelos JT, Scanavacca MI, Sampaio RO, Grinberg M, Sosa EA, Oliveira SA. Surgical treatment of atrial fibrillation through isolation of the left atrial posterior wall in patients with chronic rheumatic mitral valve disease. A randomized study with control group. *Arq Bras Cardiol* 2004;**83**:211–18, 203–10.

39. von Oppell UO, Masani N, O'Callaghan P, Wheeler R, Dimitrakakis G, Schiffelers S. Mitral valve surgery plus concomitant atrial fibrillation ablation is superior to mitral valve surgery alone with an intensive rhythm control strategy. *Eur J Cardiothorac Surg* 2009;**35**:641–50. https://doi.org/10.1016/j.ejcts.2008.12.042

40. Wang X, Song Y, Hu S, Wang W. Efficiency of radiofrequency ablation for surgical treatment of chronic atrial fibrillation in rheumatic valvular disease. *Int J Cardiol* 2014;**174**:497–502. https://doi.org/10.1016/j.ijcard.2014.03.153

41. Nashef S, Fynn S, Sharples L, Buxton M, Griffith G, Roberts P, *et al. The Amaze Trial: A Randomised Controlled Trial to Investigate the Clinical and Cost-Effectiveness of Adding an Ablation Device-Based Maze Procedure as a Routine Adjunct to Elective Cardiac Surgery for Patients with Pre-Existing Atrial Fibrillation (Af): Study Protocol*. 2015. URL: www.papworthhospital.nhs.uk/research/data/uploads/ptuc/protocol-v4-may-20151.pdf (accessed 15 December 2017).

42. Nashef S, Abu-Omar Y, Gopalan D, Bushra R, Sharples L, Stoneman V, Mills C. *The HESTER Study: Has Electrical Sinus Translated Into Effective Remodelling: Study Protocol*. 2014. URL: www.papworthhospital.nhs.uk/research/data/uploads/ptuc/hester-study-protocol-3.pdf (accessed 15 December 2017).

43. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c332. https://doi.org/10.1136/bmj.c332

44. The EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. https://doi.org/10.1016/0168-8510(90)90421-9

45. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013*. Process and Methods [PMG9]. London: NICE; 2013.

46. Dolan P, Gudex C, Kind P. *A Social Tariff for EuroQol: Results from a UK General Population Survey*. York: University of York, Centre for Health Economics; 1995.

47. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;**21**:271–92. https://doi.org/10.1016/S0167-6296(01)00130-8

48. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health* 1999;**53**:46–50. https://doi.org/10.1136/jech.53.1.46

49. Ware JE. *SF-36 Health Survey: Manual and Interpretation Guide*. Lincoln, RI: QualityMetric Inc; 2000.

50. Sharples L, Hughes V, Crean A, Dyer M, Buxton M, Goldsmith K, *et al.* Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial. *Health Technol Assess* 2007;**11**(49). https://doi.org/10.3310/hta11490

51. Campbell HE, Tait S, Sharples LD, Caine N, Gray TJ, Schofield PM, *et al.* Trial-based cost-utility comparison of percutaneous myocardial laser revascularisation and continued medical therapy for treatment of refractory angina pectoris. *Eur J Health Econ* 2005;**6**:288–97. https://doi.org/10.1007/s10198-005-0310-5

52. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;**357**:1191–4. https://doi.org/10.1016/S0140-6736(00)04337-3

53. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. London: Chapman & Hall; 1993. https://doi.org/10.1007/978-1-4899-4541-9

54. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–99. https://doi.org/10.1002/sim.4067

55. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 1987. https://doi.org/10.1002/9780470316696

56. NHS. *NHS Prescription Services Electronic Drug Tariff*. 2016. URL: www.drugtariff.nhsbsa.nhs.uk/ (accessed 10 January 2018).

57. Joint Formulary Committee. *British National Formulary* (online). London: BMJ Group and Pharmaceutical Press. URL: www.evidence.nhs.uk/formulary/bnf/current (accessed 10 January 2018).

58. Curtis L. *Unit Costs of Health and Social Care 2015*. Canterbury: PSSRU, University of Kent; 2015. URL: www.pssru.ac.uk/project-pages/unit-costs/2015/ (accessed 10 January 2018).

59. Department of Health and Social Care. *NHS Reference Costs 2014 to 2015*. London: Department of Health and Social Care; 2015. URL: www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015 (accessed 10 January 2018).

60. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *PharmacoEconomics* 2014;**32**:1157–70. https://doi.org/10.1007/s40273-014-0193-3

61. Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J* 2003;**24**:881–2. https://doi.org/10.1016/S0195-668X(02)00799-6

62. Cosgriff R, Hickey G, Grant S, Bridgewater B. *UK Heart Surgery: What Patients Can Expect from their Surgeons*. London: National Institute for Cardiovascular Outcomes Research; 2013.

63. Kind P, Hardman G, Macran S. *UK Population Norms for EQ-5D*. York: University of York, Centre for Health Economics; 1999.

64. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes* 2009;**7**:27. https://doi.org/10.1186/1477-7525-7-27

65. Buber J, Luria D, Sternik L, Raanani E, Feinberg MS, Goldenberg I, *et al.* Left atrial contractile function following a successful modified Maze procedure at surgery and the risk for subsequent thromboembolic stroke. *J Am Coll Cardiol* 2011;**58**:1614–21. https://doi.org/10.1016/j.jacc.2011.05.051

66. Boyd AC, Schiller NB, Ross DL, Thomas L. Differential recovery of regional atrial contraction after restoration of sinus rhythm after intraoperative linear radiofrequency ablation for atrial fibrillation. *Am J Cardiol* 2009;**103**:528–34. https://doi.org/10.1016/j.amjcard.2008.10.021

67. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;**370**:1453–7. https://doi.org/10.1016/S0140-6736(07)61602-X

68. Melby SJ, Lee AM, Zierer A, Kaiser SP, Livhits MJ, Boineau JP, *et al.* Atrial fibrillation propagates through gaps in ablation lines: implications for ablative treatment of atrial fibrillation. *Heart Rhythm* 2008;**5**:1296–301. https://doi.org/10.1016/j.hrthm.2008.06.009

69. Yoshitake M, Takakura H, Sasaki T, Hachiya T, Onoguchi K, Taguchi S, *et al.* Electron beam cine CT-based evaluation of left atrial function after the maze procedure for mitral valve regurgitation. *Ann Thorac Cardiovasc Surg* 2010;**16**:91–8.

70. Grady KL, Lee R, Subačius H, Malaisrie SC, McGee EC, Kruse J, *et al.* Improvements in health-related quality of life before and after isolated cardiac operations. *Ann Thorac Surg* 2011;**91**:777–83. https://doi.org/10.1016/j.athoracsur.2010.11.015

71. Pozzoli A, Taramasso M, Coppola G, Kamami M, La Canna G, Della Bella P, *et al.* Maze surgery normalizes left ventricular function in patients with persistent lone atrial fibrillation. *Eur J Cardiothorac Surg* 2014;**46**:871–6. https://doi.org/10.1093/ejcts/ezu034

72. Martinez-Comendador J, Gualis J, Marcos-Vidal JM, Buber J, Martin CE, Gomez-Plana J, *et al.* Efficacy of oral anticoagulation in stroke prevention among sinus-rhythm patients who lack left atrial mechanical contraction after cryoablation. *Tex Heart Inst J* 2015;**42**:430–7. https://doi.org/10.14503/THIJ-14-4572

73. Ad N, Henry L, Shuman DJ, Holmes SD. A more specific anticoagulation regimen is required for patients after the cox-maze procedure. *Ann Thorac Surg* 2014;**98**:1331–8. https://doi.org/10.1016/j.athoracsur.2014.05.088

74. Thomas L, Boyd A, Thomas SP, Schiller NB, Ross DL. Atrial structural remodelling and restoration of atrial contraction after linear ablation for atrial fibrillation. *Eur Heart J* 2003;**24**:1942–51. https://doi.org/10.1016/j.ehj.2003.08.018

75. van Breugel N, Bidar E, Essers B, Neiman F, Accord R, Severens J, *et al.* Cost-effectiveness of ablation surgery in patients with atrial fibrillation undergoing cardiac surgery. *Interact Cardiovasc Thorac Surg* 2011;**12**:394–8. https://doi.org/10.1510/icvts.2010.249482

76. Lamotte M, Annemans L, Bridgewater B, Kendall S, Siebert M. A health economic evaluation of concomitant surgical ablation for atrial fibrillation. *Eur J Cardiothor Surg* 2007;**32**:702–10. https://doi.org/10.1016/j.ejcts.2007.07.027

77. Assasi N, Blackhouse G, Xie F, Gaebel K, Robertson D, Hopkins R, *et al. Ablation Procedures for Rhythm Control in Patients with Atrial Fibrillation: Clinical and Cost-Effectiveness Analyses*. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health (CADTH); 2010.

78. National Institute for Health and Care Excellence (NICE). *Atrial Fibrillation: The Management of Atrial Fibrillation*. London: NICE; 2014. URL: www.nice.org.uk/guidance/cg180/evidence/atrial-fibrillation-update-full-guideline-pdf-243739981 (accessed 11 January 2018).

79. McKenna C, Palmer S, Rodgers M, Chambers D, Hawkins N, Golder S, *et al.* Cost-effectiveness of radiofrequency catheter ablation for the treatment of atrial fibrillation in the United Kingdom. *Heart* 2009;**95**:542–9. https://doi.org/10.1136/hrt.2008.147165

80. Rodgers M, McKenna C, Palmer S, Chambers D, Van Hout S, Golder S, *et al.* Curative catheter ablation in atrial fibrillation and typical atrial flutter: systematic review and economic evaluation. *Health Technol Assess* 2008;**12**(34). https://doi.org/10.3310/hta12340

81. Neyt M, Van Brabandt H, Devos C. The cost-utility of catheter ablation of atrial fibrillation: a systematic review and critical appraisal of economic evaluations. *BMC Cardiovasc Disord* 2013;**13**:78. https://doi.org/10.1186/1471-2261-13-78

82. Anderson LH, Black EJ, Civello KC, Martinson MS, Kress DC. Cost-effectiveness of the convergent procedure and catheter ablation for non-paroxysmal atrial fibrillation. *J Med Econ* 2014;**17**:481–91. https://doi.org/10.3111/13696998.2014.911185

83. Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials. *Health Technol Assess* 1999;**3**(6).

84. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment program. *Value Health* 2010;**13**:867–72. https://doi.org/10.1111/j.1524-4733.2010.00788.x

85. Seidl H, Meisinger C, Wende R, Holle R. Empirical analysis shows reduced cost data collection may be an efficient method in economic clinical trials. *BMC Health Serv Res* 2012;**12**:318. https://doi.org/10.1186/1472-6963-12-318

86. Rodevan O, Bjornerheim R, Ljosland M, Maehle J, Smith HJ, Ihlen H. Left atrial volumes assessed by three- and two-dimensional echocardiography compared to MRI estimates. *Int J Card Imaging* 1999;**15**:397–410. https://doi.org/10.1023/A:1006276513186

87. Whitlock M, Garg A, Gelow J, Jacobson T, Broberg C. Comparison of left and right atrial volume by echocardiography versus cardiac magnetic resonance imaging using the area-length method. *Am J Cardiol* 2010;**106**:1345–50. https://doi.org/10.1016/j.amjcard.2010.06.065

88. Anwar AM, Geleijnse ML, Soliman OI, Nemes A, ten Cate FJ. Left atrial Frank–Starling law assessed by real-time, three-dimensional echocardiographic left atrial volume changes. *Heart* 2007;**93**:1393–7. https://doi.org/10.1136/hrt.2006.099366

89. Lovibond K, Jowett S, Barton P, Caulfield M, Heneghan C, Hobbs FD, *et al.* Cost-effectiveness of options for the diagnosis of high blood pressure in primary care: a modelling study. *Lancet* 2011;**378**:1219–30. https://doi.org/10.1016/S0140-6736(11)61184-7

90. National Institute for Health and Care Excellence (NICE). *Quality and Outcomes Framework (QOF) Indicator Development Programme. Cost Impact Statement: Hypertension*. London: NICE; 2013. URL: www.nice.org.uk/Media/Default/standards-and-indicators/qof%20indicator%20key%20documents/NM66%20Cost%20statement.pdf (accessed 10 January 2018).

91. Auguste P, Barton P, Hyde C, Roberts TE. An economic evaluation of positron emission tomography (PET) and positron emission tomography/computed tomography (PET/CT) for the diagnosis of breast cancer recurrence. *Health Technol Assess* 2011;**15**(18). https://doi.org/10.3310/hta15180

Appendix 1 Statistical analysis plan

Clinical Trials Research Unit (CTRU) University of Leeds Statistical Analysis Plan

AMAZE

Version 1.0

15th March 2016

Produced by:

CTRU Statistician Colin Everett

Supervising Statistician Prof Linda Sharples

1 Introduction

1.1 Background

Atrial fibrillation (AF) is the most common disturbance of heart rhythm. With a UK prevalence of 7.2% in patients aged 65 and over and 10.3% in patients aged 75 and over [1] AF has considerable impact on quality of life and NHS resources[2][3]. Treatment of AF and its consequences (anti-arrhythmic & anti-coagulant drugs, hospital monitoring & stroke treatment) are expensive for the NHS and implementation of the recent NICE guidelines (June 2006)[2] on management of AF is estimated to cost £21.86m per year[3]. The NHS devotes 5% of its budget to strokes and 15% of these are associated with AF[1]. Routine anticoagulation is used to reduce the risk of stroke, however this incurs an increased risk of bleeding and the burden of monitoring treatment falls on general practice, anticoagulant clinics and haematology laboratories.

AF ablation devices are a new and costly technology being marketed to treat this condition. Their use is increasing in NHS practice despite the lack of good research evidence to support adoption. Although there are instances of their use as a stand-alone procedure, they are already in use within the NHS as an adjunct procedure for patients having cardiac surgery for other problems.

1.1.1 Existing research

The current basis for treatment and management of AF is dealt with in a UK NICE Guideline (2006)[2] European Guidelines[4] and a Cochrane review[5]. International recommendations on surgical and catheter ablation of AF were published in 2007 jointly by the Heart Rhythm Society, European Heart Rhythm Association and European Cardiac Arrhythmia Society in their Expert Consensus statement[6]. The maze procedure can be performed in two ways:

1. The traditional cut-and-sew technique, known as the Cox-maze with its many modifications, is reliable in restoring sinus rhythm in the majority of patients (references cited in Calkins et al.[6]). Despite being available since 1987, this procedure has signally failed to achieve widespread use. The main reason for this is that it is technically demanding and adds substantially to the operative burden of a heart operation. It is currently in very limited use by a few surgeons in a few centres and tends to be reserved for otherwise fit patients with severely symptomatic AF who are prepared to take the risk of a major intervention to relieve their symptoms.

2. The ablation device maze procedure uses an energy source (heat, cold, radiofrequency or microwave) to replicate the lesion set of the Cox-maze. As a rule, the

procedure is safe, well tolerated and only adds minimal increase in time and burden of the operation.

Common sense suggests that treating AF at the time of cardiac surgery is advantageous to the patient. However the only evidence supporting this comes from 5 small randomised controlled trials of ablation devices as adjuncts to surgery[7][8][9][10][11]. These trials found that SR was restored in 44-94% of treated patients compared to 5-33% of controls. The trials were small and follow-up was short. Success was mostly defined on the basis of a single ECG recording. No trial looked at patient-centred outcomes or cost effectiveness. Despite this lack of robust evidence, an increasing number of patients with AF having open heart surgery are now being offered concomitant ablation maze procedures (cited in Calkins et al.[6]).

The 2007 Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation[6] developed by Heart Rhythm Society in partnership with European Heart Rhythm Association and European Cardiac Arrhythmia Society launched a call for high quality prospective multicentre trials to adopt consistent definitions of procedural success in long term assessments of the safety and efficacy of ablation.

The relevance of the Amaze trial is highlighted by the 2010 consensus statement from international cardiothoracic surgeons[12]. This statement emphasises the urgent need for adequately powered and properly designed and reported randomised trials to measure clinically relevant outcomes (e.g. stroke, symptom relief, QoL, long term mortality) and resource use[13]. In other words, dimensions of methodological quality and clinical relevance that are incorporated into the Amaze protocol.

1.1.2 Purpose of Amaze Trial

This trial responds to this call and will inform patients, clinicians and the NHS about the routine adoption of this technology. The study will test the hypothesis that treating AF by incorporating a modified maze procedure (using an ablation device) into elective cardiac surgery will promote a return to SR and improve quality of life as well as being cost-effective from an NHS perspective.

1.2 Study design

The Amaze trial is a pragmatic, multicentre, prospective, double blind, randomised controlled trial to compare clinical, patient-based and cost outcomes for patients with pre-existing AF who undergo routine cardiac surgery either with or without an adjunct device-based ablation procedure.

The trial is double blind to the extent that neither the patient nor the cardiologist who analyses the 4 day ECG, nor the quality of life assessor should be aware which group the patient has been allocated to.

Eligible patients are randomised (in a 1:1 ratio) to receive either their routine cardiac surgery with no additional procedure or their routine cardiac surgery with an additional device-based AF ablation procedure.

1.3 Study aims and objectives

1.3.1 Intermediate Primary objective

To compare two groups for the rate of return to stable SR at 12 months as well as quality-adjusted survival over 2 years, 4-day ECG monitors will be used to assess the predominant rhythm (SR or AF) and the AF load, i.e. the percentage of time that the patient is in AF if the predominant rhythm is SR.

1.3.2 Final Primary Objective

To compare Quality-Adjusted Survival in terms of Quality-Adjusted Life Years over two years between the two groups.

1.3.3 Secondary Objectives

1. To determine whether the adjunct maze procedure improves the rate of return to stable SR at 24 months after surgery.

2. To determine whether the adjunct maze procedure decreases thromboembolic neurological complications (eg. stroke).

3. To determine whether the adjunct maze procedure enables anticoagulant treatment to be withdrawn safely.

4. To determine whether the adjunct maze procedure enables safe reduction or withdrawal of antiarrhythmic medication.

5. To determine whether the adjunct maze procedure is cost effective compared to the routine procedure.

1.4 Sample size and recruitment

1.4.1 Sample size calculation

Sample size calculations are based on both primary endpoints.

Return to SR at 12 months

Published RCTs of ablation as an addition to cardiac surgery have reported rates of return to SR at 12 months [7][8][9][10][11] ranging from 44% to 87% in the trial arms and 5% to 33% in the control arms. If we take a conservative estimate of the difference between the groups (45% vs. 30%) then we would have 80% power to detect this difference with a sample size of 176 in each group, total 352 (2-sided significance 5%). With recruitment of 400 patients this would allow for deaths or loss to follow up at 12 months of approximately 15%.

Clinical effectiveness measured as quality-adjusted survival over 2 years

The emphasis in cost-effectiveness studies is on estimation rather than hypothesis testing so that formal sample size calculations are less important. However, we provide a power calculation based on the effectiveness measure QALY. We could find no studies reporting comparative QALYs in similar patients undergoing ablation and cardiac surgery. From previous studies of patients undergoing angiography for suspected ischaemic heart disease[14] and patients with refractory angina[15] the standard deviation of QALYs over 12 and 18 months is at most 0.3. Over follow up of 2 years the minimum clinically important improvement is considered to be one extra month of quality-adjusted life, or 0.083 QALYs. With a sample of 200 patients per group, total 400, we would have approximately 80% power to detect a difference of 0.083 QALYs, (2-sided significance 5%). If the accepted threshold for cost effectiveness were in the range £20-30,000 per QALY and we could demonstrate a significant increase in QALYs of 0.0833, then the procedure would be cost-effective for an incremental cost of at most £2,500.

1.5 Randomisation

Patients who fulfil the eligibility and have given written informed consent and have sufficient time for discussion and consideration, are randomised (in a 1:1 ratio) to one of two groups to receive either their routine cardiac surgery with no additional procedure or their routine cardiac surgery with an adjunct maze procedure.

Patient allocations are computer generated by the trial statistician and are in random permuted blocks of variable lengths, stratified by surgeon and by planned cardiac procedure (CABG, aortic valve, mitral valve, combined procedure).

1.6 Eligibility

Eligible patients are consecutive elective cardiac surgical patients undergoing major cardiac surgery (such as coronary, valve or combined operations) with a history of paroxysmal, persistent or chronic AF beginning more than 3 months before the date of the operation.

Paroxysmal AF is defined as recurrent AF (> 2 episodes) that terminates spontaneously within 4 days (Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation)[6].

Persistent AF is defined as AF which continues beyond 4 days.

Chronic or longstanding AF is persistent AF beyond 1 year.

Inclusion criteria:

age over 18.

elective cardiac surgery planned (Coronary surgery, valve surgery, combined coronary and valve surgery, any other Cardiac surgery requiring cardiopulmonary bypass.)

history of documented atrial fibrillation (chronic, persistent or paroxysmal) beginning more than 3 months before entry into the study.

written informed consent to participation.

Exclusion criteria:

previous cardiac operations.

emergency or salvage cardiac operations.

surgery without cardiopulmonary bypass.

unlikely to be available for follow-up over a two-year period.

deemed not competent to provide consent.

All randomised participants will be included in the final intention-to-treat analysis, except those for whom consent to use data was withdrawn, or written informed consent was not received. Deviations from these criteria will be summarised and reported.

2 Outcome

2.1 Primary outcomes

Rate of return to stable SR at 12 months- 4 day ECG monitors are being used to assess the predominant rhythm (SR or AF) and the AF load ie the percentage of time that the patient is in AF if their predominant rhythm is SR.

Clinical effectiveness quality adjusted survival over 2 years

2.2 Secondary outcomes

Clinical endpoints of SR at 24 months after surgery, overall survival and stroke-free survival, incidence of anticoagulant-related haemorrhage.

Health-related QoL measured by the EuroQoL, SF-36 and NYHA.

Resource use and cost-effectiveness of the adjunct maze procedure.

Anticoagulant and antiarrhythmic drug usage.

2.3 Missing data

Data management will focus on the consenting process, participant eligibility, safety, date consistency and test outcomes. Attempts will be made to retrieve missing data on these areas via a thorough data cleaning process.

The levels of missing data and reasons for missingness will be investigated for the consenting process, participant eligibility, safety, date consistency and test outcomes. The quantity of missing data will be monitored by treatment group, and a summary of the number of patients with missing primary endpoint data and the quantity of missing data by treatment group will be reported.

For the intermediate primary endpoint (return to sinus rhythm at 12 months), if a patient withdraws consent or is lost to follow-up for further participation within 12 months, multiple imputation used to impute missing outcome (AF or SR at 12 months) as a function of the baseline heart rhythm, surgeon, surgical procedure and treatment group. Rubin's rules will be used to combine imputed datasets. If more than 5% of patients are withdrawn or lost to follow-up before 12 months, then the outcome will be imputed under the alternative patterns as sensitivity analyses. (See section 5.4.1)

For the final primary endpoint (Quality-Adjusted Survival at 2 years), where a patient is deceased before the end of follow-up, the utility value of 0 will be imputed for all

subsequent assessments. If the response is missing, and the patient is not known to be deceased, the missing value will be imputed using the method of multiple imputation. A "Last Observation Carried Forward" approach will be used as an alternative imputation technique for imputing other (non-death) missing values in a sensitivity analysis.

The primary analysis model will only require the baseline rhythm, surgeon and surgical procedure which are immediately recorded when patients attend the preadmission clinic and has consented to participate or during the period of surgery, so there is little concerns about missing data arising in this model. If any missing values in the covariates are reported, they will be imputed using a function of the known covariates and primary outcomes of interest.

3 Population

3.1 Intention-to-treat Analysis

An intention-to-treat analysis will be the primary method for analysing and summarising the trial data. The intention-to-treat population is defined as all randomised patients, regardless of eligibility, withdrawal, compliance with the protocol, loss to follow-up or actual treatment received. Only patients who have withdrawn their consent for their data to be used in the study, or for whom written informed consent has not been received, will be excluded in this population. These patients will be analysed and summarised according to the intervention they were randomised to receive.

If more than 5% tests or trial conduct constitute a major protocol violation such as cross-over to the other arm or cancelling surgeries, the Complier Average Casual Effect analysis will be considered.

3.2 Quality of life populations

A separate quality of life population will be formed for the analysis of each questionnaire (SF36). Each population will comprise all patients who return an analysable baseline questionnaire, regardless of subsequent questionnaire return: patients without analysable baseline questionnaires will be excluded from the analysis, regardless of subsequent questionnaire return.

3.3 Safety Population

All patients will be included in the safety population if they underwent one of the two procedures. Patients will be included in the arm corresponding to the intervention received. If no intervention was received, then the patient will be summarised separately from the other intervention groups.

4 Data Collection

4.1 Methods

The data is collected on to a web-based system designed and coordinated by the Data Scan and Quality Officer at Papworth Hospital. The Clinical Research Nurse (CRN) at each centre enters the data directly on to the database. Surgical data will be recorded by a designated member of the surgical team either directly or via a paper form. All paper data collection forms are returned to the R&D Unit at Papworth. The Trial Coordinator (TC) are responsible for data monitoring and quality control. The whole process is overseen by the Trial Manager situated in the co-ordinating centre at Papworth Hospital.

4.2 Baseline data collection

We adhere to the ACC/AHA/ECS 2006 Guidelines which recommend that the initial patient description includes demographics, type and duration of AF and the planned cardiac procedure.

The first 4-day ECG recording starts after the patient attends the preadmission clinic and has consented to participate. All other baseline measurements are recorded on the day of admission for surgery. Once these measurements have been taken, the patient is registered with the co-ordinating centre's R&D unit and randomised as described in Section 1.5.

4.3 Data collection during and after surgery

Data collection is based on the recommendations of Shemin et al.[13] and includes procedural details-including the lesion set in the experimental group. Data collected after surgery includes: mortality, stroke/thromboembolic events, medications, EuroQoL, health-related quality of life, cardioversion plan if appropriate, 4-day ECG recordings, resource use, adverse events.

Data are collected during surgery, at discharge, 6 weeks after surgery (at a routine service visit), at 6, 12 and 24 months after surgery during out-patient research visits and annually thereafter by telephone follow-up.

4.4 Analysis of ECG recordings

All 4 day continuous ECG recordings will be analysed centrally at Papworth Hospital. Participating centres forward the SD cards from the ECG recorders to Papworth Hospital. Analysis using the proprietary automated software package, together with

manual checking of the recording in its entirety, will be done. Total time spent in sinus rhythm and in AF (AF burden) during the 4 day recording will be calculated, with only those episodes of AF lasting greater than 60 seconds duration included in the analysis. Episodes of atrial flutter will be noted and included in the AF burden.

Occurrences of Atrial Flutter or Atrial Tachycardia ("Organised Atrial Arrhythmia") in patients experiencing AF and Junctional Rhythm in patients reportedly in Sinus Rhythm will be reported.

4.5 ONS tracking

All patients enrolled in this trial (with their consent) are registered with the Office of National Statistics (ONS) Tracking System to allow long term follow up of survival. ONS tracking data is not expected to form part of the primary analysis of outcomes up to 24 months. Instead, this will be used to follow-up patients over a longer period if longer-term follow-up analyses are required.

4.6 Data validation

Data management will focus on the data associated with the consenting process, participant eligibility, safety, date consistency and test outcomes and this section refers to the cleaning of these items. The Data Management Assistant (DMA)/ Trial Co-ordinator (TC) will carry out initial validation of the forms in accordance with the trial-specific Data Management Work Instructions. This will ensure that data is complete, consistent, and up-to-date. The Data Clarification Form (DCF) will be sent to sites to highlight missing data items and queries associated with data collected on CRFs to date. Reasons should be obtained when data is unobtainable.

The database will validate most data in line with validation rules and highlight any issues that need further investigation i.e. with the site. Manual checks on all entered data will be performed prior to the validations being implemented. Data items collected relating to the safety and rights of individual patients are to be highlighted via priority validations and dealt with as a data management priority. Periodic batch validation will also be carried out to detect any data queries that may be missed if case record forms (CRFs) are entered in an order that does not allow real time validation checks to work.

A key data items list drawn up by the Trial Statistician that will include all data items that are required for the analysis of the primary endpoint. All key data items will be checked manually for completeness and accuracy by the DMA/TC, in addition to any automatic checks raised on the database. Data automatically generated through the 24-hour randomisation system will be checked by the Trial Statistician.

The Trial Statistician will also perform checks to identify any missing or inconsistent data and liaise with the Trial-Coordinator to resolve any queries.

The data will be validated and checked using SAS in the following steps:

The data will be read into permanent SAS data sets.

A random sample of 5 patients from each SAS dataset will be checked against the data as seen on the database to ensure that the data transfer has been successful. The names and contents of the variables can be found on the annotated final database specification reports in the Statistician's Trial File.

Data checks will include:-

Eligibility checks

Sequential dates

Checks for unusual and outlying data

Inconsistency in data between forms

Checks for missing data (are there variables which are systematically missing/do specific variables have a large amount of missing data, particularly key outcome data)

Other checks as deemed appropriate

Any inconsistent data will be noted and an e-mail sent to the trial co-ordinator responsible for the study. A copy of this e-mail will be kept in the statistician's trial file. All queries will be resolved and the outcome documented.

5 Data analysis

It is expected that the final analysis of the data will be performed when all patients have completed 24 months of follow-up.

5.1 General calculations

All statistical analyses and reporting will comply with CONSORT guidelines where possible.

Confidence intervals for a single proportion shall be calculated using the Exact method. Confidence intervals for a difference between independent proportions shall be calculated using Exact intervals (Method 8 of [16]).

All percentages will be calculated using the total number of patients within the specified analysis population, percentages will be reported to 1 decimal place. All statistical tests will be 2-sided and performed at the 5% significance level. All analyses will be carried out using SAS.

For summary statistics, the number of non-missing items, the means, standard deviations, medians, upper and lower quartiles and minima and maxima will be summarised to one more decimal place than the data are collected.

5.2 General principles

Multivariable analyses will not be `built' following a model-fitting strategy. Instead, all variables specified for inclusion will be added to the model, and the significance of each factor will be reported. Where one categorical variable has more than one `factor level' then the significance of overall effect of including all factor levels will be tested, rather than those for each individual factor level. For all factor levels, suitable point and interval estimates of effect size will be presented.

If any analysis requires the use of simulation and / or re-sampling methods, the initial `seed' value for the random number generation will be 0471346543. The same seed will be used at the start of every such analysis.

5.3 Baseline data and surgery data

Patient baseline data and surgery data as recorded on the baseline assessment or during surgery will be tabulated using frequencies and summary statistics by treatment group, for each randomising centre and in total, for the intention-to-treat population (and safety populations if appropriate). No statistical testing will be carried out on these data.

5.4 Analysis of Primary Endpoints

Quality-Adjusted Survival

For the final primary outcome, QALYs will be estimated from serial measurements of the EQ-5D for each patient up to 2 years using interpolation. The Health Economics analysis will be given in Section 6.

At baseline, on discharge and at 6 weeks, 6 months, 12 months and 24 months post-surgery all patients complete the EuroQol questionnaire, including the EQ-5D. The social tariff for the EQ-5D, as estimated by Dolan et al. [17] will be applied to each patient's self-reported classification in order to calculate utility values. Using actual rather than nominal times of assessment, and assuming a linear change in values between time points, patient-specific utility curves up to 24-months post randomisation will be calculated. A value of zero will be applied at the date of death for those patients who died.

The QALYs experienced by each patient to 24-months post randomization are calculated as the area under their utility curve to 24 months or time of death, whichever occurs first, where the true test dates rather than nominal test dates will be utilized in plotting the utility curve. In order to adjust for differences in baseline utilities a linear regression will be fitted to the utilities post treatment, with baseline utility and treatment group as explanatory variables. The linear regression will also include a random effect for surgeon if it is feasible to fit, and yields a positive variance component for the surgeon effect. Adjusted treatment effects will be taken from the treatment group coefficient of this regression. For patients who do not complete all EuroQoL measurements and are censored the methods of Willan and Lin[18] will be used to estimate mean QALYs and costs. The adequacy of model fit for the linear regression model at each timepoint will be assessed by examining distributions of standardised residuals, association with the predicted values, as well as identifying influential observations by referring to leverage statistics.

The differences in Quality-adjusted Survival will be presented. A confidence interval for the true difference will be formed using a non-parametric bootstrap resampling approach: [19]

A simple random sample with replacement will be drawn from the full analysis dataset of the same size as the full dataset. (ie some patients may be drawn more than once)

The difference in QALYs between the two treatment groups will be estimated for this bootstrap sample.

Steps (1) and (2) will be repeated 1000 times.

The 95% confidence interval will be formed as the interval between the 2.5% and the 97.5% percentile of the differences computed in these bootstrap samples.

Return to SR

The intermediate primary outcome, whether patient returns to SR at 12 months, will be summarized according to the group to which they were randomised. The comparison between the heart rhythm (AF or SR) at the baseline and 12 months will be also summarized.

For the primary endpoint analysis, the rate in SR for Routine treatment group will be compared to that for Routine+Maze treatment group using a binary logistic regression model, which will include surgeon (as a random effect), surgical procedure and baseline heart rhythm as fixed effects. The odds ratio for the rate of return to SR at 12 months of Routine treatment group against Routine+Maze treatment group will be reported with 95% confidence interval and p-value for the data seen, under the null hypothesis that the rates of return to SR at 12 months are no different between the pairs of groups.

The adequacy of the logistic regression models for the primary endpoint will be assessed by examining the following statistics and relevant graphical summaries:

Pearson Residuals/Deviance (Half-normal plots)

Leverage values.

Cook's Distance.

Cross validation probabilities (the probability of a particular observation, conditional on the remaining observations).

L-statistics (the influence of an observation on the difference in deviance due to fitting an the treatment effect).

Actual percentage time in AF across the 4 days of monitoring at baseline and at 12 months, i.e. the percentage of time that the patient is in AF if their predominant rhythm is SR, is reported by treatment group. Based on the interim report, the percentage time

in AF is almost dichotomized at 0% and 100%. Therefore, only summary statistics are reported in this case.

5.4.1 Sensitivity Analysis

For the final primary endpoint (Quality-adjusted Survival), we will impute as zero any EQ5D utility value that is missing due to the patient death. Any remaining missing values will be imputed using multiple imputation. For the sensitivity analysis, a Last Observation Carried Forward approach will be used to impute missing (nonzero) utility values at 24 months, and any missing intermediate utility values. Alternative imputation techniques will be considered. We will additionally consider a sensitivity analysis using SF6D-derived utility values and other valuation methods as appropriate.

If missing, the intermediate primary endpoint (heart rhythm) will be estimated by multiple imputation technique. If more than 5% of patients are withdrawn or lost to follow-up before the 12 months is reported, then the following methods will be used as sensitivity analyses to estimate the primary endpoint:

1. a `death=AF, censored=AF' strategy: If a patient is withdrawn or lost to follow-up within the 12 months, he will be assumed to be in AF.

2. a `death=AF, censored=OMIT' strategy: If a patient dies for any cause within 12 months, he will be assumed to be in AF at 12 months. If a patient is withdrawn or lost to follow-up but is alive within 12 months, his record will not be included into the sensitivity analysis.

5.5 Subgroup Analysis

Subgroup analysis will include those patients for whom measurements are available.

The first objective of subgroup analysis is within the whole dataset, to account for potential variation in the treatment effects between

patients with paroxysmal and non-paroxysmal AF: Paroxysmal AF vs. Persistent Chronic or longstanding AF.

individual centres (as a random effect to allow for heterogeneities in small centres

different cardiac surgical procedures

different surgeons.

different lesion sets.

The different lesion sets are to be defined as follows:

minimal LA lesion set: pulmonary vein isolation only ± LA appendage line

more extensive LA lesion set excluding mitral annulus

more extensive LA only lesion set including mitral annulus

minimal LA lesion set + RA lesion set

more extensive LA lesion set excluding mitral annulus + RA lesion set

more extensive LA lesion set including mitral annulus + RA lesion set

In the event that patients are too sparsely-distributed across the 6 categories, the categories will be combined into 4, by combining (i) with (ii) and combining (iv) with (v). If this is still too sparse to facilitate comparison, then the lesion set subgroup will be reduced to a comparison of category (vi) to all other groups.

For the Quality-Adjusted Survival Endpoint, a linear regression model will be fitted to the Area Under the Utility Curve, with baseline EQ5D score, surgeon, surgical procedure, treatment group, subgroup variable and the subgroup-by-treatment interaction variable. The treatment modifying effect will be reported with a 95% confidence interval. The Area under the utility curve will be appropriately transformed prior to performing subgroup analyses, and the results back-transformed where necessary.

For the Return to SR Endpoint, a logistic regression model will be fitted to heart rhythm at 12 months with the baseline heart rhythm, surgeon, surgical procedure, treatment group, subgroup variable of interest and its interaction term with treatment group. The odds ratio of the interaction term between treatment and subgroup variable will be reported with 95% confidence interval and p-value for the data seen, under the null hypothesis that there is no difference in the treatment effects on patients within different subgroup of interest.

The second objective is within the Routine+Maze treatment group, to account for variation in the treatment effects between

different ablation devices

completeness of lesion sets – both on a continuous scale and categorised as 0-4, 5-9, 10+.

In each subgroup analysis listed above, the regression model will include only patients in the Routine+Maze treatment group. The interaction effect will then be tested in the same manner as for the previous interaction effects. The odds ratio or parameter estimate of the subgroup variable will be reported with 95% confidence interval and p-value for the data seen, under the null hypothesis that in the Routine+Maze treatment group, there is no difference in the treatment effects on patients within different subgroup of interest.

5.6 Key Secondary Endpoint Analysis

The key secondary endpoint, the rate to return stable SR at 24 months will be analysed in a similar way to the rate of return to SR at 12 month, but will include only those patients for whose measurements are available. The rate in SR at 24 months for Routine treatment group will be compared to that for Routine+Maze treatment group using a binary logistic regression model including baseline heart rhythm, surgeon, surgical procedure and treatment group. The odds ratio for the rate of return to SR at 24 months of Routine treatment group relative to the Routine+Maze group will be reported with 95% confidence interval and p-value for the data seen, under the null hypothesis that the rates of return to SR at 24 months are no different between the pairs of groups.

5.7 SF36

The SF-36 consists of eight scaled scores (vitality, physical functioning, bodily pain, general health perceptions, physical role functioning, emotional role functioning, social functioning and mental health), which are the weighted sums of the questions in their section. Each scale is directly transformed into a 0-100 scale and a higher score represents a better health for that domain. Standardised physical and mental health scores are then calculated which for a general UK population are expected to be approximately normally distributed with mean 50 and standard deviation 10. [20]

The SF-36 questionnaire will be administered at baseline, 6 months, 12 months and 24 months. Then the scores for the summary measures of SF36, (Physical score and mental health score) will be given at each timepoint.

SF-36 component summary scores will be analysed using a linear regression model, adjusting for time point, treatment group, time by treatment group interaction, baseline SF-36 scores (all modelled as fixed effects) and allowing random intercepts for patients.

5.8 Additional Secondary Endpoint Analyses

To investigate whether the adjunct maze procedure decreases thromboembolic neurological complications (e.g. stroke), the patients who have suffered a stroke will be summarized within 12 months of surgery and the overall proportion of stroke events will be calculated by treatment group, using the total number of patients participating in the trial as the denominator. The absolute differences between the proportions for Routine treatment group and Routine+Maze group will be tested by Fisher's exact test, and reported along with 95% confidence intervals for differences in proportions.

The number of recruited patients who use anti-arrhythmic drugs will be tabulated by time points (at baseline, discharge, 6 weeks, 6 months, 12 months and 24 months) and drug categories (Sotalol, Amiadarone and Flecainade). Logistic Regression for the outcome of each patient (1=have one or more drugs during time period t, 0=have no drug during time period t) will be fitted, including drug category, time period using drug, baseline drug usage and treatment group as independent variables. The odds ratio for using anti-arrhythmic drug in the Routine treatment group relative to the Routine+Maze group will be reported with 95% confidence interval and p-value for the data seen, under the null hypothesis that the usage of anti-arrhythmic drugs is no different between the groups.

In the similar way, the number of recruited patients who use anti-coagulant drugs will be tabulated by time point (at baseline, discharge, 6 weeks, 6 months, 12 months and 24 months) and drug categories (Warfarin, Sinthrome and other anticoagulants). The logistic regression for the outcome of each patient (1=have one or more drugs during time period t, 0=have no drug during time period t) will be fitted, including the drug category, time period using drug, baseline drug usage and treatment group. The odds ratio of using anti-coagulant drug in the Routine treatment group relative to the Routine+Maze group will be reported with 95% confidence interval and p-value for the data seen, under the null hypothesis that the usage of anticoagulants drug is no different between the pairs of groups.

The occurrence of atrial flutter and atrial tachycardia ("organised atrial arrhythmia" - OAA) and junctional rhythm (JR) will be summarised by arm. An exploratory analysis will look at the relation between the completeness of the lesion set and the occurrences of OAA and JR.

5.9 Safety Analysis

A listing of total number of adverse events in each category as well as the deaths from any cause will be presented, and summarised by treatment group, corresponding to the intervention received. Events will be summarised according to whether they meet the

criteria of Serious Adverse Events, and whether they are thought to be related to the procedure.

Adverse Events are not planned to be categorised. However, a number of pre-specified Adverse Event categories have been specified.

6 Economic Analysis

Health Economic Analysis will be performed by the Health Economic Analysis team. A separate analysis plan has been written, and should be referred to for a description of the planned analyses.

References

[1] F. R. Hobbs, D. Fitzmaurice, J. Mant, E. Murray, S. Jowett, S. Bryan, J. Raftery, M. Davies, and G. Lip, "A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. the safe study," Health Technology Assessment, vol. 9, no. 40, p. 93pp, 2005.

[2] "Atrial fibrillation: the management of atrial fibrillation," NICE clinical guideline, no. 36, pp. 1–48, 2006.

[3] "Atrial fibrillation: the management of atrial fibrillation. costing report. implementing nice guidance in england," NICE clinical guideline, no. 36, pp. 1–39, 2006.

[4] J. Dunning, T. Treasure, M. Versteegh, and S. A. Nashef, "Guidelines on the prevention and management of de novo atrial fibrillation after cardiac and thoracic surgery," European journal of cardio-thoracic surgery, vol. 30, no. 6, pp. 852–872, 2006.

[5] E. Crystal, M. S. Garfinkle, S. Connolly, T. Ginger, K. Sleik, and S. Yusuf, "Interventions for preventing post-operative atrial fibrillation in patients undergoing heart surgery," Cochrane Database Syst Rev, vol. 4, 2004.

[6] H. Calkins, J. Brugada, D. L. Packer, R. Cappato, S.A. Chen, H. J. Crijns, R. J. Damiano, D. W. Davies, D. E. Haines, M. Haissaguerre, et al., "Hrs/ehra/ecas expert consensus statement on catheter and surgical ablation of atrial fibrillation: Recommendations for personnel, policy, procedures and follow-up a report of the heart rhythm society (hrs) task force on catheter and surgical ablation of atrial fibrillation developed in partnership with the European heart rhythm association (ehra) and the European cardiac arrhythmia society (ecas); in collaboration with the American college of cardiology (acc), American heart association (aha), and the society of thoracic surgeons (sts). endorsed and approved by the governing bodies of the American college of cardiology, the American heart association, the European cardiac arrhythmia society, the European heart rhythm association, the society of thoracic surgeons, and the heart rhythm society.," Europace, vol. 9, no. 6, pp. 335–379, 2007.

[7] T. Deneke, K. Khargi, P. Grewe, A. Laczkovics, S. Von Dryander, T. Lawo, K.M. Müller, and B. Lemke, "Efficacy of an additional maze procedure using cooled-tip radiofrequency ablation in patients with chronic atrial fibrillation and mitral valve disease. a randomized, prospective trial," European Heart Journal, vol. 23, no. 7, pp. 558–566, 2002.

[8] A. Schuetz, C. Schulze, K. Sarvanakis, H. Mair, H. Plazer, E. Kilger, B. Reichart, and S. Wildhirt, "Surgical treatment of permanent atrial fibrillation using microwave energy ablation: a prospective randomized clinical trial," European Journal of Cardio-Thoracic Surgery, vol. 24, no. 4, pp. 475–480, 2003.

[9] G. Doukas, N. J. Samani, C. Alexiou, M. Oc, D. T. Chin, P. G. Stafford, L. L. Ng, and T. J. Spyt, "Left atrial radiofrequency ablation during mitral valve surgery for continuous atrial fibrillation: a randomized controlled trial," Jama, vol. 294, no. 18, pp. 2323–2329, 2005.

[10] C. A. Abreu Filho, L. A. Lisboa, L. A. Dallan, G. S. Spina, M. Grinberg, M. Scanavacca, E. A. Sosa, J. A. F. Ramires, and S. A. Oliveira, "Effectiveness of the maze procedure using cooled-tip radiofrequency ablation in patients with permanent atrial fibrillation and rheumatic mitral valve

disease," Circulation, vol. 112, no. 9 suppl, pp. I–20, 2005.

[11] B. Akpinar, I. Sanisoglu, M. Guden, E. Sagbas, B. Caynak, and Z. Bayramoglu, "Combined off-pump coronary artery bypass grafting surgery and ablative therapy for atrial fibrillation: early and mid-term results," The Annals of thoracic surgery, vol. 81, no. 4, pp. 1332–1337, 2006.

[12] N. Ad, D. C. Cheng, J. Martin, E. E. Berglin, B.C. Chang, G. Doukas, J. S. Gammie, T. Nitta, R. K. Wolf, and J. D. Puskas, "Surgical ablation for atrial fibrillation in cardiac surgery: a consensus statement of the international society of minimally invasive cardiothoracic surgery (ismics) 2009," Innovations: Technology and Techniques in Cardiothoracic and Vascular Surgery, vol. 5, no. 2, pp. 74–83, 2010.

[13] R. J. Shemin, J. L. Cox, A. M. Gillinov, E. H. Blackstone, and C. R. Bridges, "Guidelines for reporting data and outcomes for the surgical treatment of atrial fibrillation," The Annals of thoracic surgery, vol. 83, no. 3, pp. 1225–1230, 2007.

[14] V. Fuster, L. E. Rydén, D. S. Cannom, H. J. Crijns, A. B. Curtis, K. A. Ellenbogen, J. L. Halperin, J.Y. Le Heuzey, G. N. Kay, J. E. Lowe, et al., "Acc/aha/esc 2006 guidelines for the management of patients with atrial fibrillation: full text a report of the American college of cardiology/American heart association task force on practice guidelines and the European society

of cardiology committee for practice guidelines (writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation) developed in collaboration with the European heart rhythm association and the heart rhythm society," Europace, vol. 8, no. 9, pp. 651–745, 2006.

[15] L. Sharples, V. Hughes, A. Crean, M. Dyer, M. Buxton, K. Goldsmith, and D. Stone, Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial. Gray Pub., 2007.

[16] R. G. Newcombe, "Interval estimation for the difference between independent proportions: comparison of eleven methods," Statistics in medicine, vol. 17, no. 8, pp. 873–890, 1998.

[17] P. Dolan, C. Gudex, P. Kind, and A. Williams, A social tariff for EuroQol: results from a UK general population survey. Centre for Health Economics University of York, UK, 1995.

[18] A. Willan, D. Lin, R. Cook, and E. Chen, "Using inverse-weighting in cost-effectiveness analysis with censored data," Statistical methods in medical research, vol. 11, no. 6, pp. 539–551, 2002.

[19] B. Efron and R. J. Tibshirani, An introduction to the bootstrap, vol. 57. CRC press, 1994.

[20] Brazier J E, Harper R, Jones N M, O'Cathain A, Thomas K J, Usherwood T, Westlake L. Validating The SF-36 Health Survey Questionnaire: New Outcome Measure For Primary Care. British Medical Journal 1992; 305: 160-4.SF-36

Appendix 2 Additional tables from the health economics analysis

TABLE 23 Summary of unit costs

Primary admission cost	Source	Consultation time/code	Mean cost (£) (SD) for 2014/15
Theatre use	Papworth Hospital estimate		20.00 (4.00)
Adult critical care	*NHS Reference Costs 2014 to 2015*[59]	Total/weighted average	1274.92 (583.33)
Cardiac ward	*NHS Reference Costs 2014 to 2015*[59]	Code: SD01A	387.96 (77.59)
General ward	*NHS Reference Costs 2014 to 2015*[59]	Code: SD03A	103.01 (20.60)
Rehabilitation	PSSRU 2015,[58] 1.3		158.57 (31.71)
Acute (specialised ward)	*NHS Reference Costs 2014 to 2015*[59]	Code: SD01A	387.96 (77.59)
Follow-up admission			
ICU	*NHS Reference Costs 2014 to 2015*[59]	Total/weighted average	1274.92 (583.33)
Cardiac ward	*NHS Reference Costs 2014 to 2015*[59]	Code: SD01A	387.96 (77.59)
General ward	*NHS Reference Costs 2014 to 2015*[59]	Code: SD03A	103.01 (20.60)
Day case	*NHS Reference Costs 2014 to 2015*[59]	Code: DC	720.78 (144.16)
Follow-up tests			
Angiography	*NHS Reference Costs 2014 to 2015*[59]	Code: EY43F	260.00 (52.00)
MUGA	*NHS Reference Costs 2014 to 2015*[59]	Code: RN22Z	192.39 (38.48)
Echocardiography	*NHS Reference Costs 2014 to 2015*[59]	Simple echocardiogram	83.94 (16.79)
PET	*NHS Reference Costs 2014 to 2015*[59]	Code: RN07A	524.77 (104.95)
TOE	*NHS Reference Costs 2014 to 2015*[59]	Complex echocardiogram	128.49 (25.70)
Computerised tomography scan	*NHS Reference Costs 2014 to 2015*[59]	Total/weighted average	122.31 (48.86)
Echocardiogram stress	*NHS Reference Costs 2014 to 2015*[59]	Complex echocardiogram	128.49 (25.70)
MRI scan	*NHS Reference Costs 2014 to 2015*[59]	Total/weighted average	146.15 (56.64)

continued

TABLE 23 Summary of unit costs (*continued*)

Primary admission cost	Source	Consultation time/code	Mean cost (£) (SD) for 2014/15
Exercise test	*NHS Reference Costs 2014 to 2015*[59]	Field exercise testing	287.08 (57.42)
ECG	*NHS Reference Costs 2014 to 2015*[59]	ECG monitoring	52.13 (10.43)
24-hour ECG	*NHS Reference Costs 2014 to 2015*[59]	ECG monitoring	169.26 (33.85)
> 24-hour ECG	*NHS Reference Costs 2014 to 2015*[59]	ECG monitoring	169.26 (33.85)
24-hour blood pressure monitoring	Lovibond *et al.*[89]		61.47 (12.29)
< 24-hour blood pressure monitoring	NICE Cost Statement 2013[90]		38.34 (7.67)
Left heart catheterisation	Papworth Hospital estimate		1267.00 (253.40)
Radiography (chest)	Auguste *et al.*[91]		3.46 (0.69)
Follow-up health-care visits			
GP visits	PSSRU 2015,[58] 10.8b	Per-patient contact lasting 17.2 minutes	65.00 (13.00)
GP home visits	PSSRU 2015,[58] 10.8b	Per-patient contact lasting 11.7 minutes	45.00 (9.00)
Nurse visits	PSSRU 2015,[58] 10.6	Per-patient contact 15.5 minutes	14.47 (2.89)
Nurse home visits	PSSRU 2015,[58] 10.4	Per-patient contact 17.2 minutes	19.38 (3.88)
Cardiology clinic	*NHS Reference Costs 2014 to 2015*[59]	Code: WF01A	123.02 (24.60)
AF clinic	*NHS Reference Costs 2014 to 2015*[59]	Code: WF01A (cardiology clinic)	123.02 (24.60)
Pacemaker	*NHS Reference Costs 2014 to 2015*[59]	Code: EY08E	76.32 (15.26)
Physiotherapy	*NHS Reference Costs 2014 to 2015*[59]	Code: WF01A	14.32 (2.86)
Occupational therapy	*NHS Reference Costs 2014 to 2015*[59]	Code: WF01A	21.41 (4.28)
A&E visit	*NHS Reference Costs 2014 to 2015*[59]	Total/weighted average	140.59 (141.05)
Wound clinic	*NHS Reference Costs 2014 to 2015*[59]	Code: N25AF/AN	54.93 (10.99)
Cardiac rehabilitation	*NHS Reference Costs 2014 to 2015*[59]	Code: VC38Z	97.84 (19.57)

MUGA, multigated acquisition scan; PET, positron emission tomography; PSSRU, Personal Social Services Research Unit; TOE, transoesophageal echocardiography.

TABLE 24 Resource use across the two treatment arms, without imputation

Primary admission resource	Unit of measurement	Resource use per patient in each treatment arm			
		Maze procedure		Control	
		Observation	Mean (SD)	Observation	Mean (SD)
Theatre	Minutes	176	261.24 (79.68)	176	247.48 (93.27)
Critical care (ICU)	Days	175	3.17 (5.98)	176	2.40 (4.38)
Cardiac ward	Days	174	8.73 (12.08)	176	7.90 (5.19)
Convalescence	Days	167	0.29 (2.07)	169	0.96 (6.96)
Transferred to acute trust[a]	Days	173	2.37 (15.86)	176	0.43 (3.63)
Follow-up admission					
6 weeks in ICU	Days	125	0.07 (0.72)	130	0.04 (0.29)
6 weeks in ward	Days	125	2.49 (9.06)	130	1.25 (4.12)
6 weeks in cardiac ward	Days	125	1.42 (7.29)	130	0.64 (2.37)
6 weeks as a day case	Days	125	0.01 (0.09)	130	0.00 (0.00)
6 months in ICU	Days	115	0.01 (0.09)	122	0.02 (0.20)
6 months in ward	Days	115	0.59 (3.02)	122	0.89 (5.40)
6 months in cardiac ward	Days	115	0.61 (2.75)	122	0.70 (3.16)
6 months as a day case	Days	115	0.04 (0.24)	122	0.01 (0.09)
12 months in ICU	Days	122	0.07 (0.46)	122	0.00 (0.00)
12 months in ward	Days	122	0.02 (0.16)	122	0.11 (0.54)
12 months in cardiac ward	Days	122	1.38 (5.42)	122	0.56 (2.18)
12 months as a day case	Days	122	0.02 (0.20)	122	0.01 (0.09)
24 months in ICU	Days	122	0.02 (0.27)	131	0.00 (0.00)
24 months in ward	Days	122	0.63 (5.31)	131	0.44 (2.06)
24 months in cardiac ward	Days	122	1.28 (5.07)	131	1.30 (5.96)
24 months as a day case	Days	122	0.00 (0.00)	131	0.04 (0.19)
Follow-up tests					
Year 1 angiography	Number of tests	145	0.03 (0.20)	148	0.01 (0.12)
Year 1 MUGA	Number of tests	145	0.04 (0.50)	148	0.00 (0.00)
Year 1 echocardiogram TTE	Number of tests	145	0.46 (0.77)	148	0.53 (0.83)
Year 1 PET scan	Number of tests	145	0.00 (0.00)	148	0.00 (0.00)
Year 1 echocardiogram TOE	Number of tests	145	0.02 (0.14)	148	0.02 (0.14)
Year 1 computerised tomography	Number of tests	145	0.13 (0.40)	148	0.07 (0.26)
Year 1 echocardiogram stress	Number of tests	145	0.00 (0.00)	148	0.00 (0.00)
Year 1 MRI scan	Number of tests	145	0.05 (0.25)	148	0.04 (0.20)
Year 1 exercise test	Number of tests	145	0.04 (0.23)	148	0.01 (0.08)
Year 1 ECG	Number of tests	145	2.42 (2.06)	149	2.16 (2.08)
Year 1 24-hour ECG	Number of tests	145	0.16 (0.44)	148	0.13 (0.36)

continued

TABLE 24 Resource use across the two treatment arms, without imputation (*continued*)

Primary admission resource	Unit of measurement	Resource use per patient in each treatment arm			
		Maze procedure		Control	
		Observation	Mean (SD)	Observation	Mean (SD)
Year 1 > 24-hour ECG	Number of tests	145	0.13 (0.46)	148	0.10 (0.38)
Year 1 24-hour blood pressure monitoring	Number of tests	145	0.02 (0.14)	148	0.02 (0.18)
Year 1 < 24-hour blood pressure monitoring	Number of tests	145	0.01 (0.12)	148	0.09 (0.69)
Year 1 left heart catheter	Number of tests	145	0.00 (0.00)	148	0.00 (0.00)
Year 1 radiography	Number of tests	145	1.61 (2.47)	149	1.12 (1.34)
Year 2 angiography	Number of tests	146	0.01 (0.12)	150	0.01 (0.12)
Year 2 MUGA	Number of tests	146	0.00 (0.00)	150	0.00 (0.00)
Year 2 echocardiogram TTE	Number of tests	146	0.21 (0.53)	150	0.18 (0.42)
Year 2 PET scan	Number of tests	146	0.00 (0.00)	150	0.00 (0.00)
Year 2 echocardiogram TOE	Number of tests	146	0.02 (0.14)	149	0.01 (0.08)
Year 2 computerised tomography	Number of tests	146	0.05 (0.26)	150	0.09 (0.42)
Year 2 echocardiogram stress	Number of tests	146	0.00 (0.00)	150	0.01 (0.08)
Year 2 MRI scan	Number of tests	146	0.05 (0.30)	150	0.03 (0.16)
Year 2 exercise test	Number of tests	146	0.00 (0.00)	150	0.01 (0.12)
Year 2 ECG	Number of tests	146	0.92 (3.67)	150	0.69 (1.13)
Year 2 24-hour ECG	Number of tests	146	0.02 (0.14)	150	0.07 (0.28)
Year 2 > 24-hour ECG	Number of tests	146	0.03 (0.16)	150	0.04 (0.20)
Year 2 24-hour blood pressure monitoring	Number of tests	146	0.01 (0.12)	150	0.01 (0.08)
Year 2 < 24-hour blood pressure monitoring	Number of tests	146	0.01 (0.12)	150	0.03 (0.20)
Year 2 left heart catheter	Number of tests	146	0.01 (0.08)	150	0.00 (0.00)
Year 2 radiography	Number of tests	146	0.19 (0.67)	149	0.18 (0.57)
Follow-up health-care visits					
Year 1 GP visits	Number of visits	145	3.77 (4.59)	148	3.66 (4.28)
Year 1 GP home visits	Number of visits	145	0.17 (0.72)	148	0.24 (1.23)
Year 1 nurse (general practice) visits	Number of visits	145	4.21 (8.74)	148	4.17 (8.84)
Year 1 nurse home visits	Number of visits	145	1.46 (4.90)	148	0.86 (2.39)
Year 1 cardiovascular clinic	Number of visits	145	1.63 (1.34)	148	1.32 (1.20)
Year 1 AF clinic	Number of visits	145	0.03 (0.20)	148	0.01 (0.08)
Year 1 pacemaker	Number of visits	145	0.10 (0.36)	148	0.21 (0.81)
Year 1 physiotherapy	Number of visits	145	0.32 (1.35)	148	0.47 (2.85)
Year 1 occupational therapy	Number of visits	145	0.01 (0.08)	148	0.07 (0.41)
Year 1 A&E visit	Number of visits	145	0.26 (0.76)	148	0.27 (0.75)

TABLE 24 Resource use across the two treatment arms, without imputation (*continued*)

Primary admission resource	Unit of measurement	Resource use per patient in each treatment arm			
		Maze procedure		Control	
		Observation	Mean (SD)	Observation	Mean (SD)
Year 1 wound clinic	Number of visits	145	0.05 (0.27)	148	0.02 (0.18)
Year 1 cardiac rehabilitation	Number of visits	145	3.09 (7.76)	148	3.32 (6.93)
Year 2 GP visits	Number of visits	146	1.55 (3.37)	150	1.54 (2.56)
Year 2 GP home visits	Number of visits	146	0.01 (0.12)	150	0.09 (0.99)
Year 2 nurse (general practice) visits	Number of visits	146	2.63 (6.87)	150	2.40 (4.99)
Year 2 nurse home visits	Number of visits	146	0.21 (1.51)	150	0.15 (1.35)
Year 2 cardiovascular clinic	Number of visits	146	0.54 (0.74)	150	0.59 (0.98)
Year 2 AF clinic	Number of visits	146	0.00 (0.00)	150	0.01 (0.16)
Year 2 pacemaker	Number of visits	146	0.08 (0.30)	150	0.15 (0.50)
Year 2 physiotherapy	Number of visits	146	0.18 (0.99)	150	0.10 (0.82)
Year 2 occupational therapy	Number of visits	146	0.03 (0.33)	150	0.04 (0.42)
Year 2 A&E visit	Number of visits	146	0.10 (0.40)	150	0.17 (0.61)
Year 2 wound clinic	Number of visits	146	0.00 (0.00)	150	0.00 (0.00)
Year 2 cardiac rehabilitation	Number of visits	146	0.36 (3.45)	150	0.33 (3.76)

MUGA, multigated acquisition scan; PET, positron emission tomography; TOE, transoesophageal echocardiography; TTE, transthoracic echocardiography.
a Assumed as transfer to specialised care at an acute trust.

TABLE 25 Resource use across the two treatment arms, with imputation

Primary admission cost	Unit of measurement	Resource use per patient in each treatment arm, mean (SD)	
		Maze procedure ($n = 176$)	Control ($n = 176$)
Theatre use	Minutes	261.24 (79.68)	247.43 (93.13)
Adult critical care	Days	3.16 (5.96)	2.40 (4.38)
Cardiac ward (specialised ward)	Days	8.76 (12.01)	7.90 (5.19)
Rehabilitation	Days	0.30 (2.05)	0.93 (6.83)
Acute (specialised ward)	Days	2.41 (15.74)	0.43 (3.63)
Follow-up admission at week 6			
ICU	Days	0.05 (0.61)	0.03 (0.25)
General ward	Days	1.80 (7.72)	1.03 (3.72)
Cardiac ward	Days	1.47 (6.86)	0.50 (2.08)
Day case	Days	0.01 (0.08)	0.00 (0.00)
Follow-up admission at 6 months			
ICU	Days	0.01 (0.08)	0.02 (0.17)
General ward	Days	0.39 (2.45)	0.66 (4.52)
Cardiac ward	Days	0.40 (2.24)	0.52 (2.67)

continued

TABLE 25 Resource use across the two treatment arms, with imputation (*continued*)

Primary admission cost	Unit of measurement	Resource use per patient in each treatment arm, mean (SD)	
		Maze procedure ($n = 176$)	Control ($n = 176$)
Day case	Days	0.04 (0.23)	0.01 (0.08)
Follow-up admission at 12 months			
ICU	Days	0.05 (0.38)	0.00 (0.00)
General ward	Days	0.02 (0.13)	0.09 (0.48)
Cardiac ward	Days	1.02 (4.58)	0.41 (1.85)
Day case	Days	0.02 (0.17)	0.01 (0.08)
Follow-up admission at 24 months			
ICU	Days	0.02 (0.23)	0.00 (0.00)
General ward	Days	0.48 (4.43)	0.44 (1.91)
Cardiac ward	Days	1.06 (4.38)	1.17 (5.24)
Day case	Days	0.00 (0.00)	0.03 (0.17)
Follow-up tests (year 1)			
Angiography	Number of tests	0.04 (0.22)	0.02 (0.13)
MUGA	Number of tests	0.03 (0.45)	0.00 (0.00)
TTE	Number of tests	0.49 (0.77)	0.55 (0.80)
PET scan	Number of tests	0.00 (0.00)	0.00 (0.00)
TOE	Number of tests	0.03 (0.14)	0.03 (0.14)
Computerised tomography	Number of tests	0.14 (0.38)	0.08 (0.27)
Echocardiogram stress	Number of tests	0.00 (0.00)	0.00 (0.00)
MRI scan	Number of tests	0.05 (0.23)	0.05 (0.19)
Exercise test	Number of tests	0.04 (0.22)	0.01 (0.08)
ECG	Number of tests	2.43 (1.98)	2.34 (2.23)
24-hour ECG	Number of tests	0.15 (0.41)	0.13 (0.34)
> 24-hour ECG	Number of tests	0.16 (0.48)	0.11 (0.39)
Blood pressure monitoring	Number of tests	0.02 (0.13)	0.02 (0.17)
24-hour blood pressure monitoring	Number of tests	0.02 (0.12)	0.09 (0.64)
Left heart catheter	Number of tests	0.00 (0.00)	0.00 (0.00)
Radiography	Number of tests	1.65 (2.38)	1.10 (1.29)
Follow-up tests (year 2)			
Angiography	Number of tests	0.01 (0.11)	0.01 (0.11)
MUGA	Number of tests	0.00 (0.00)	0.00 (0.00)
TTE	Number of tests	0.22 (0.51)	0.19 (0.40)
PET scan	Number of tests	0.00 (0.00)	0.00 (0.00)
TOE	Number of tests	0.02 (0.13)	0.01 (0.08)
Computerised tomography	Number of tests	0.07 (0.26)	0.10 (0.41)
Echocardiogram stress	Number of tests	0.00 (0.00)	0.01 (0.08)
MRI scan	Number of tests	0.05 (0.28)	0.03 (0.18)
Exercise test	Number of tests	0.00 (0.00)	0.02 (0.12)
ECG	Number of tests	0.87 (3.35)	0.71 (1.10)

TABLE 25 Resource use across the two treatment arms, with imputation (*continued*)

Primary admission cost	Unit of measurement	Resource use per patient in each treatment arm, mean (SD)	
		Maze procedure ($n = 176$)	Control ($n = 176$)
24-hour ECG	Number of tests	0.03 (0.15)	0.06 (0.26)
> 24-hour ECG	Number of tests	0.03 (0.16)	0.04 (0.19)
Blood pressure monitoring	Number of tests	0.02 (0.12)	0.01 (0.08)
24-hour blood pressure monitoring	Number of tests	0.01 (0.11)	0.03 (0.20)
Left heart catheter	Number of tests	0.01 (0.08)	0.00 (0.00)
Radiography	Number of tests	0.22 (0.68)	0.20 (0.57)
Follow-up health-care visits (year 1)			
GP visits	Number of visits	3.72 (4.31)	3.81 (4.22)
GP home visits	Number of visits	0.18 (0.69)	0.22 (1.13)
Nurse (general practice) visits	Number of visits	4.16 (8.33)	4.22 (8.58)
Nurse home visits	Number of visits	1.38 (4.55)	0.86 (2.25)
Cardiovascular clinic	Number of visits	1.59 (1.30)	1.30 (1.16)
AF clinic	Number of visits	0.03 (0.19)	0.01 (0.08)
Pacemaker	Number of visits	0.11 (0.34)	0.23 (0.81)
Physiotherapy	Number of visits	0.44 (1.44)	0.64 (3.25)
Occupational therapy	Number of visits	0.01 (0.08)	0.07 (0.38)
A&E visit	Number of visits	0.29 (0.80)	0.29 (0.74)
Wound clinic	Number of visits	0.07 (0.30)	0.02 (0.17)
Cardiac rehabilitation	Number of visits	3.07 (7.34)	3.23 (6.51)
Follow-up health-care visits (year 2)			
GP visits	Number of visits	1.65 (3.25)	1.70 (2.67)
GP home visits	Number of visits	0.01 (0.11)	0.09 (0.91)
Nurse (general practice) visits	Number of visits	2.67 (6.65)	2.53 (4.96)
Nurse home visits	Number of visits	0.22 (1.46)	0.14 (1.25)
Cardiovascular clinic	Number of visits	0.53 (0.72)	0.61 (0.94)
AF clinic	Number of visits	0.00 (0.00)	0.01 (0.15)
Pacemaker	Number of visits	0.11 (0.33)	0.17 (0.49)
Physiotherapy	Number of visits	0.21 (1.05)	0.13 (0.92)
Occupational therapy	Number of visits	0.02 (0.30)	0.05 (0.43)
A&E visit	Number of visits	0.13 (0.43)	0.16 (0.57)
Wound clinic	Number of visits	0.00 (0.00)	0.00 (0.00)
Cardiac rehabilitation	Number of visits	0.36 (3.20)	0.41 (3.87)

MUGA, multigated acquisition scan; PET, positron emission tomography; TOE, transoesophageal echocardiography; TTE, transthoracic echocardiography.

TABLE 26 Total cost across two arms, with imputation

Primary admission cost	Resource use (£) per patient in each treatment arm, mean (SD)	
	Maze procedure ($n = 176$)	Control ($n = 176$)
Theatre use	5224.89 (1593.66)	4948.52 (1862.51)
Ablation device	1211.65 (408.33)	14.20 (132.87)
Adult critical care	4028.75 (7600.08)	3064.51 (5585.89)
Cardiac ward	3396.88 (4661.36)	3064.03 (2013.71)
Rehabilitation	47.75 (324.98)	147.76 (1082.35)
Acute (specialised ward)	936.84 (6105.06)	165.33 (1408.64)
Subtotal	14,846.76 (12,473.61)	11,404.35 (7193.49)
Readmissions at 6 weeks		
ICU	65.19 (774.24)	36.22 (317.57)
General ward	151.58 (706.84)	51.21 (213.78)
Cardiac ward	697.67 (2994.27)	398.99 (1444.75)
Day case	4.10 (54.33)	0.00 (0.00)
Readmissions at 6 months		
ICU	7.24 (96.10)	21.73 (214.40)
General ward	40.97 (230.75)	53.55 (274.82)
Cardiac ward	149.89 (951.09)	254.60 (1753.63)
Day case	28.67 (165.46)	4.10 (54.33)
Readmissions at 12 months		
ICU	61.57 (489.90)	0.00 (0.00)
General ward	105.35 (471.71)	42.72 (190.13)
Cardiac ward	7.72 (52.30)	35.27 (185.59)
Day case	12.29 (121.21)	4.10 (54.33)
Readmissions at 24 months		
ICU	21.73 (288.30)	0.00 (0.00)
General ward	109.45 (451.38)	120.86 (539.50)
Cardiac ward	186.27 (1720.26)	171.94 (741.60)
Day case	0.00 (0.00)	24.57 (125.38)
Subtotal	1649.69 (4192.49)	1219.85 (2994.21)
Follow-up tests (year 1)		
Angiography	9.60 (57.33)	5.17 (35.05)
MUGA	6.56 (87.01)	0.00 (0.00)
TTE	40.78 (64.93)	46.26 (67.46)
PET scan	0.00 (0.00)	0.00 (0.00)
TOE	3.29 (18.52)	3.29 (18.52)
Computerised tomography	17.03 (47.00)	10.08 (32.76)
Echocardiogram stress	0.00 (0.00)	0.00 (0.00)
MRI scan	7.47 (34.12)	6.64 (28.46)
Exercise test	11.42 (64.09)	1.63 (21.64)

TABLE 26 Total cost across two arms, with imputation (*continued*)

Primary admission cost	Resource use (£) per patient in each treatment arm, mean (SD)	
	Maze procedure ($n = 176$)	Control ($n = 176$)
ECG	126.63 (103.36)	122.18 (116.13)
24-hour ECG	25.00 (69.09)	21.64 (57.75)
> 24-hour ECG	26.93 (80.46)	18.75 (65.98)
Blood pressure monitoring	1.05 (7.98)	1.05 (10.34)
24-hour blood pressure monitoring	0.65 (4.54)	3.38 (24.49)
Left heart catheter	0.00 (0.00)	0.00 (0.00)
Radiography	5.71 (8.24)	3.79 (4.45)
Follow-up tests (year 2)		
Angiography	3.69 (29.25)	2.95 (27.64)
MUGA	0.00 (0.00)	0.00 (0.00)
TTE	18.12 (42.47)	15.74 (33.76)
PET scan	0.00 (0.00)	0.00 (0.00)
TOE	2.56 (17.32)	1.10 (10.80)
Computerised tomography	7.99 (32.03)	12.51 (50.35)
Echocardiogram stress	0.00 (0.00)	0.73 (9.69)
MRI scan	7.89 (40.95)	4.98 (26.60)
Exercise test	0.00 (0.00)	4.89 (33.96)
ECG	45.32 (174.84)	37.17 (57.48)
24-hour ECG	5.29 (25.84)	10.58 (43.98)
> 24-hour ECG	5.29 (27.38)	7.21 (32.45)
Blood pressure monitoring	1.22 (7.61)	0.35 (4.63)
24-hour blood pressure monitoring	0.54 (4.31)	1.20 (7.71)
Left heart catheter	7.20 (95.50)	0.00 (0.00)
Radiography	0.75 (2.36)	0.69 (1.96)
Subtotal	387.97 (375.85)	343.96 (283.09)
Follow-up health-care visits (year 1)		
GP visits	241.72 (280.11)	247.63 (274.01)
GP home visits	8.05 (31.01)	9.72 (50.95)
Nurse (general practice) visits	60.17 (120.49)	61.11 (124.06)
Nurse home visits	26.75 (88.09)	16.68 (43.53)
Cardiology clinic	195.36 (160.23)	160.42 (142.71)
AF clinic	3.84 (23.85)	0.70 (9.27)
Pacemaker	8.02 (26.32)	17.78 (61.72)
Physiotherapy	7.10 (23.29)	10.41 (52.43)
Occupational therapy	0.14 (1.40)	1.23 (6.31)
A&E visit	41.14 (113.06)	41.14 (104.50)
Wound clinic	3.90 (16.28)	1.09 (9.45)
Cardiac rehabilitation	300.19 (718.22)	315.76 (637.43)

continued

TABLE 26 Total cost across two arms, with imputation (*continued*)

Primary admission cost	Resource use (£) per patient in each treatment arm, mean (SD)	
	Maze procedure ($n = 176$)	Control ($n = 176$)
Follow-up health-care visits (year 2)		
GP visits	107.29 (211.47)	110.61 (173.24)
GP home visits	0.64 (5.06)	3.96 (41.02)
Nurse (general practice) visits	38.59 (96.18)	36.58 (71.71)
Nurse home visits	4.24 (28.20)	2.64 (24.17)
Cardiology clinic	65.35 (88.02)	75.14 (115.44)
AF clinic	0.00 (0.00)	1.40 (18.55)
Pacemaker	8.02 (25.02)	12.79 (37.30)
Physiotherapy	3.39 (16.89)	2.02 (14.80)
Occupational therapy	0.38 (5.03)	0.81 (7.11)
A&E visit	18.77 (60.69)	23.17 (80.04)
Wound clinic	0.00 (0.00)	0.00 (0.00)
Cardiac rehabilitation	35.58 (313.15)	40.03 (378.61)
Subtotal	1178.66 (1060.51)	1192.79 (1052.40)
Medication		
Year 1	476.94 (1203.98)	510.06 (2092.35)
Year 2	141.20 (456.15)	171.29 (727.75)
Subtotal	618.14 (1583.61)	681.35 (2764.66)
Total	18,681.21 (13,339.82)	14,842.30 (8295.33)

MUGA, multigated acquisition scan; PET, positron emission tomography; TOE, transoesophageal echocardiography; TTE, transthoracic echocardiography.

TABLE 27 Summaries of utility scores by treatment arm, with imputation

Time point of utility score	Utility score by treatment arm, mean score (SD)	
	Maze procedure ($n = 176$)	Control ($n = 176$)
EQ-5D-3L score		
Baseline	0.7417 (0.22)	0.7544 (0.21)
Discharge	0.5927 (0.29)	0.5943 (0.28)
6-week follow-up	0.7187 (0.27)	0.7246 (0.27)
6-month follow-up	0.7312 (0.29)	0.7350 (0.31)
12-month follow-up	0.7167 (0.31)	0.7397 (0.30)
24-month follow-up	0.6719 (0.34)	0.7163 (0.32)
SF-6D score		
Baseline	0.6599 (0.11)	0.6651 (0.11)
6-month follow-up	0.6703 (0.23)	0.6853 (0.23)
12-month follow-up	0.6565 (0.24)	0.6723 (0.24)
24-month follow-up	0.6350 (0.27)	0.6685 (0.25)

TABLE 28 Summaries of utility scores by treatment arm, without imputation

	Treatment arm			
	Maze procedure		Control	
Time point of utility score	Observations	Mean score (SD)	Observations	Mean score (SD)
EQ-5D-3L score				
Baseline	174	0.7417 (0.22)	175	0.7544 (0.21)
Discharge	161	0.6091 (0.28)	165	0.6034 (0.28)
6-week follow-up	164	0.7378 (0.25)	169	0.7339 (0.26)
6-month follow-up	169	0.7311 (0.30)	174	0.7323 (0.31)
12-month follow-up	165	0.7142 (0.32)	167	0.7367 (0.31)
24-month follow-up	158	0.6680 (0.35)	161	0.7098 (0.34)
SF-6D score				
Baseline	172	0.6623 (0.11)	174	0.6644 (0.11)
6-month follow-up	168	0.6694 (0.23)	171	0.6846 (0.24)
12-month follow-up	166	0.6557 (0.24)	166	0.6678 (0.25)
24-month follow-up	158	0.6291 (0.28)	157	0.6597 (0.26)

TABLE 29 Deterministic sensitivity analysis (using the difference between maze procedure and control, adjusted for baseline)

		Difference (maze procedure vs. control), mean (SD)			INMB (£) at a WTP of	
Scenario	Observations	Incremental cost (£) over 24 months	Incremental QALYs over 24 months	ICER (£)	£20,000 per QALY	£30,000 per QALY
Using EQ-5D-3L QALYs (base case)	352	3533 (1129)	–0.0220 (0.0516)	Dominated	–3974	–4194
Using SF-6D QALYs	352	3533 (1129)	–0.0197 (0.0433)	Dominated	–3928	–4125
Including costs and QALYs data only up to discharge	352	3140 (1044)	0.0014 (0.0018)	2,234,462	–3112	–3098
Excluding outliers from the analysis	350	2856 (1013)	–0.0108 (0.0512)	Dominated	–3073	–3181
Alternate imputation model (cost imputation)	352	4103 (1180)	–0.0214 (0.0520)	Dominated	–4530	–4744
Excluding device cost	352	2336 (1128)	–0.0220 (0.0516)	Dominated	–2777	–2997
Complete-case analysis	234	2210 (1108)	0.0264 (0.0571)	83,625	–1682	–1417
Randomised after April 2011	200	3364 (1574)	0.0628 (0.0693)	53,538	–2107	–1479

TABLE 30 Probabilistic sensitivity analysis (using the difference between maze procedure and control, adjusted for baseline)

Scenario	Observations	Difference (maze procedure vs. control), mean (SD)		ICER (£)	INMB (£) at a WTP of	
		Incremental cost (£) over 24 months	Incremental QALYs over 24 months		£20,000 per QALY	£30,000 per QALY
Using EQ-5D-3L QALYs	1000	3656 (1259)	–0.0205 (0.0545)	Dominated	–4066	–4271
Using SF-6D QALYs	1000	3779 (1280)	–0.0204 (0.0443)	Dominated	–4187	–4391
Including costs and QALYs data only up to discharge	1000	3003 (1125)	0.0017 (0.0020)	1,745,221	–2969	–2952
Excluding outliers from the analysis	1000	2936 (1123)	–0.0125 (0.0511)	Dominated	–3186	–3312
Alternate imputation model	1000	3794 (1103)	–0.0144 (0.0515)	Dominated	–4082	–4226
Excluding additional device cost for maze procedure	1000	2580 (1278)	–0.0239 (0.0536)	Dominated	–3058	–3297
Complete case analysis	1000	2256 (1281)	0.0277 (0.0574)	81,516	–1702	–1426
Randomised after April 2011	1000	3942 (1868)	0.0551 (0.0737)	71,505	–2840	–2288

TABLE 31 Coefficients and standard errors from seemingly unrelated regression models for costs and conversions from AF to SR, adjusting for baseline covariates (secondary cost-effectiveness analysis)

Dependent variable	Independent variable	Coefficient	SEM	*p*-value
Conversion of AF to SR	Maze procedure	0.14	0.04	3.37
	Male	–0.05	0.04	–1.12
	Age	0.008	0.002	0.03
	Baseline EQ-5D-3L score	0.23	0.1	2.3
	Paroxysmal AF	0.05	0.04	1.1
	Actual procedure	–0.005	0.01	–0.48
	Constant	–0.25	0.23	–1.07
Total cost (£) per patient	Maze procedure	3533	1129	0.00
	Male	–2131	1205	0.08
	Age	255	75	0.00
	Baseline EQ-5D-3L score	–9367	2645	0.00
	Paroxysmal AF	2693	1300	0.04
	Constant	–1691	6247	0.79

SEM, standard error of the mean.

Appendix 3 Additional tables

TABLE 32 Reasons for exclusion from the trial at the screening at Papworth Hospital ($n = 366$)

Reason for exclusion	Frequency
Patient decision	($n = 107$)
Too much to think about	27
Could not commit to follow-up or too far to travel	18
Patient did not want/was unsure about surgery, or was treated medically	12
Age concerns	4
GP or family concerned about trial participation	3
Patient did not want a 4-day ECG	3
Patient did not want the maze procedure	1
Uncomfortable about blinding of trial allocation	1
No reason given	38
Consultant decision	($n = 115$)
Maze procedure carried out outside the trial	49
Minimally invasive or other procedure required	11
Withdrawn from surgery	13
Too frail/sick/comorbidities	9
Procedure was too high risk/complicated	9
Maze procedure was not desirable because of small left atrium and only paroxysmal AF	1
Not suitable – reason not specified	23
Patient failed eligibility or ethics criteria	($n = 49$)
Patient already in another trial	19
Patient was an in-house urgent patient prior to change in eligibility criteria	6
Patient in SR	5
Patient had previous cardiac surgery	3
Patient unsuitable – reason not recorded	16
Patient died	($n = 8$)
Other	($n = 87$)
Not enough time to consider the trial	25
Unable to contact patient	13
Patient information sheet not given to the patient	6
Language barrier	1
Other administrative reasons – not recorded	42

TABLE 33 Comorbidities at baseline

Comorbidity	Treatment arm, *n* (%)		
	Maze procedure (*n* = 176)	Control (*n* = 176)	Total (*n* = 352), *n* (%)
COPD			
Yes	14 (8.0)	20 (11.4)	34 (9.7)
No	162 (92.0)	155 (88.1)	317 (90.1)
Missing/not known	–	1 (0.6)	1 (0.3)
Pulmonary hypertension			
Yes	25 (14.2)	28 (15.9)	53 (15.1)
No	151 (85.8)	147 (83.5)	298 (84.7)
Missing/not known	–	1 (0.6)	1 (0.3)
Extracardiac ateriopathy			
Yes	9 (5.1)	6 (3.4)	15 (4.3)
No	167 (94.9)	169 (96.0)	336 (95.5)
Missing/not known	–	1 (0.6)	1 (0.3)
Neurological dysfunction			
Yes	5 (2.8)	5 (2.8)	10 (2.8)
No	171 (97.2)	170 (96.6)	341 (96.9)
Missing/not known	–	1 (0.6)	1 (0.3)
Serum creatinine level of > 200 µmol/l			
Yes	2 (1.1)	–	2 (0.6)
No	174 (98.9)	176 (100.0)	350 (99.4)
Rheumatic fever			
Yes	7 (4.0)	6 (3.4)	13 (3.7)
Cardiomyopathy			
Yes	2 (1.1)	4 (2.3)	6 (1.7)
Marfan syndrome			
Yes	1 (0.6)	–	1 (0.3)
Transient ischaemic attack			
Yes	19 (10.8)	12 (6.8)	31 (8.8)
Cerebrovascular accident			
Yes	11 (6.3)	11 (6.3)	22 (6.3)
Other clinical history			
Yes	152 (86.4)	146 (83.0)	298 (84.7)

TABLE 34 Major surgical complications

Complication	Treatment arm, *n* (%)		
	Maze procedure (*n* = 176)	Control (*n* = 176)	Total (*n* = 352), *n* (%)
No complications[a]	34 (19.3)	38 (21.6)	72 (20.5)
AF/arrhythmia	69 (39.2)	80 (45.5)	149 (42.3)
Need for a permanent pacemaker	6 (3.4)	5 (2.8)	11 (3.1)
Bleeding/tamponade	22 (12.5)	17 (9.7)	42 (11.1)
Aortic dissection	–	1 (0.6)	1 (0.3)
Cardiac arrest	2 (1.1)	–	2 (0.6)
Cardiogenic shock	1 (0.6)	1 (0.6)	2 (0.6)
Pericardial effusion	2 (1.1)	–	2 (0.6)
Other cardiac event	4 (2.3)	6 (3.4)	10 (2.8)
Wound infection	3 (1.7)	4 (2.3)	7 (2.0)
Infection/sepsis	1 (0.6)	4 (2.3)	5 (1.4)
Hypotension	5 (2.8)	1 (0.6)	6 (1.7)
Respiratory/respiratory infection	19 (10.8)	21 (11.9)	40 (11.4)
Pleural effusion	14 (8.0)	25 (14.2)	39 (11.1)
Pulmonary embolism	–	1 (0.6)	1 (0.3)
Pneumothorax	–	2 (1.1)	2 (0.6)
Renal failure/dysfunction	13 (7.4)	10 (5.7)	23 (6.5)
Neurological	3 (1.7)	6 (3.4)	9 (2.6)
Delirium	5 (2.8)	1 (0.6)	6 (1.7)
Non-cardiac bleeding complication	1 (0.6)	–	1 (0.3)
Gastrointestinal complication	4 (2.3)	4 (2.3)	8 (2.3)
Metabolic complication	2 (1.1)	4 (2.3)	6 (1.7)
Peripheral oedema	–	1 (0.6)	1 (0.3)
Wound complication	1 (0.6)	1 (0.6)	2 (0.6)
Other complication	5 (2.8)	2 (1.1)	7 (2.0)
Perioperative death	–	1 (0.6)	1 (0.3)
Returned to theatre	20 (11.4)	18 (10.2)	38 (10.8)
Reason for return to theatre			
Bleeding	4 (2.3)	10 (5.7)	14 (4.0)
Tamponade	4 (2.3)	1 (0.6)	5 (1.4)
Gastrointestinal complication	2 (1.1)	2 (1.1)	4 (1.1)
Respiratory complication	2 (1.1)	1 (0.6)	3 (0.9)
Cardiac arrest	2 (1.1)	–	2 (0.6)
Reoperation	–	1 (0.6)	1 (0.3)
Permanent pacemaker	1 (0.6)	–	1 (0.3)
Cardiac reason	–	1 (0.6)	1 (0.3)

continued

TABLE 34 Major surgical complications (*continued*)

Complication	Treatment arm, *n* (%)		Total (*n* = 352), *n* (%)
	Maze procedure (*n* = 176)	Control (*n* = 176)	
Arrhythmia	1 (0.6)	–	1 (0.3)
Bleeding (non-cardiac)	1 (0.6)	–	1 (0.3)
Hypotension	1 (0.6)	–	1 (0.3)
Pericardial effusion	–	1 (0.6)	1 (0.3)
Respiratory complication/infection	–	1 (0.6)	1 (0.3)
Wound	1 (0.6)	–	1 (0.3)
Cardiogenic shock	–	1 (0.6)	1 (0.3)
Pleural effusion	1 (0.6)	–	1 (0.3)
Missing/not known	2 (1.1)	–	2 (0.6)

a No complications: patient did not experience any perioperative complications.

TABLE 35 Number of patients who had transfusions

Type of transfusion	Treatment arm, *n* (%)		Total (*n* = 352), *n* (%)
	Maze procedure (*n* = 176)	Control (*n* = 176)	
Total red blood cell transfusion (units)			
Not known	4 (2.3)	1 (0.6)	5 (1.4)
0	98 (55.7)	106 (60.2)	204 (58.0)
≥ 1	74 (42.0)	69 (39.2)	143 (40.6)
Total platelets transfusion (units)			
Not known	5 (2.8)	3 (1.7)	8 (2.3)
0	128 (72.7)	141 (80.1)	269 (76.4)
≥ 1	43 (24.4)	32 (18.2)	75 (21.3)
Total fresh-frozen plasma transfusion (units)			
Not known	4 (2.3)	3 (1.7)	7 (2.0)
0	136 (77.3)	139 (79.0)	275 (78.1)
≥ 1	36 (20.5)	34 (19.3)	70 (19.9)
Total cryoprecipitate transfusion (units)			
Not known	7 (4.0)	6 (3.4)	13 (3.7)
0	157 (89.2)	162 (92.0)	319 (90.6)
≥ 1	12 (6.8)	8 (4.5)	20 (5.7)
Total human albumin transfusion (units)			
Not known	5 (2.8)	4 (2.3)	9 (2.6)
0	165 (93.8)	163 (92.6)	328 (93.2)
≥ 1	6 (3.4)	9 (5.1)	15 (4.3)

TABLE 36 Results of the mixed-model-adjusted analysis for the primary outcome of return to SR at 12 months

Effect	OR of SR at 12 months (95% CI)	$p > \|t\|$
Intercept	0.69 (0.34 to 1.40)	0.3078
Randomised to the maze procedure arm	2.06 (1.20 to 3.54)	0.0091
In SR at baseline	6.52 (2.83 to 14.98)	< 0.0001
Procedure: CABG vs. MVR	0.84 (0.37 to 1.93)	0.6836
Procedure: AVR vs. MVR	0.42 (0.18 to 0.99)	0.0479
Procedure: CABG and MVR vs. MVR	0.64 (0.21 to 1.93)	0.4288
Procedure: CABG and AVR vs. MVR	0.52 (0.18 to 1.54)	0.2373
Procedure: all others (including none) vs. MVR	0.75 (0.35 to 1.61)	0.4642

TABLE 37 Results of the baseline-adjusted subgroup analyses of SR at 12 months, comparing the maze procedure arm with the control arm

Level	Treatment arm, in SR/total		Adjusted OR (maze procedure/control) (95% CI)	$p > \|t\|$
	Maze procedure	Control		
Paroxysmal AF	32/38	30/42	2.286 (0.679 to 7.693)	0.1815
Non-paroxysmal AF	55/103	38/103	2.036 (1.117 to 3.710)	0.0204
CABG	18/30	15/29	1.625 (0.545 to 4.847)	0.3835
MVR	26/34	24/43	3.375 (1.141 to 9.980)	0.0279
AVR	12/27	7/16	1.511 (0.391 to 5.836)	0.5485
CABG and MVR	9/12	2/10	7.159 (0.907 to 56.496)	0.0618
CABG and AVR	5/10	5/14	1.810 (0.275 to 11.921)	0.5355
All others, including none	17/28	15/33	1.437 (0.497 to 4.154)	0.5032

TABLE 38 Results of the baseline-adjusted mixed model for the effect of the maze procedure on 2-year quality-adjusted survival

Effect	Estimate (95% CI)	$p > \|t\|$
Intercept	0.775 (0.550 to 1.000)	< 0.0001
Baseline EQ-5D-3L utility score	0.962 (0.715 to 1.208)	< 0.0001
Randomised to the maze procedure arm	–0.025 (–0.129 to 0.078)	0.6319
Surgical procedure: CABG	–0.081 (–0.238 to 0.075)	0.3087
Surgical procedure: AVR	–0.103 (–0.271 to 0.065)	0.2280
Surgical procedure: CABG and MVR	–0.043 (–0.256 to 0.170)	0.6925
Surgical procedure: CABG and AVR	–0.282 (–0.471 to –0.092)	0.0037
Surgical procedure: all others, including none	0.010 (–0.142 to 0.161)	0.8986

TABLE 39 Results of the subgroup analysis comparing the maze procedure arm with the control arm for 2-year quality-adjusted survival

Level	Treatment arm, mean (SEM)		Difference (maze procedure vs. control) (95% CI)	*p*-value
	Maze procedure	Control		
Paroxysmal AF	1.430 (0.076)	1.511 (0.075)	–0.081 (–0.283 to 0.122)	0.4356
Non-paroxysmal AF	1.375 (0.045)	1.379 (0.045)	–0.004 (–0.125 to 0.116)	0.9417
CABG	1.403 (0.083)	1.397 (0.085)	0.006 (–0.228 to 0.239)	0.9626
MVR	1.426 (0.078)	1.530 (0.071)	–0.104 (–0.310 to 0.103)	0.3265
AVR	1.451 (0.088)	1.272 (0.104)	0.179 (–0.086 to 0.444)	0.1865
CABG and MVR	1.525 (0.131)	1.345 (0.137)	0.179 (–0.192 to 0.550)	0.3438
CABG and AVR	1.068 (0.122)	1.303 (0.108)	–0.235 (–0.555 to 0.084)	0.1491
All others, including none	1.452 (0.079)	1.533 (0.080)	–0.080 (–0.302 to 0.141)	0.4779

SEM, standard error of the mean.

TABLE 40 Baseline-adjusted mixed model for the effect of the maze procedure on return to SR at 24 months

Effect	OR of SR at 24 months (95% CI)	$p > \|t\|$
Intercept (odds of SR at 24 months for baseline AF, MVR operation and standard care)	0.38 (0.17 to 0.83)	0.0169
In SR at baseline	11.53 (4.48 to 29.70)	< 0.0001
Randomised to the maze procedure arm	3.24 (1.76 to 5.96)	0.0002
Procedure: AVR vs. MVR	0.48 (0.18 to 1.29)	0.1435
Procedure: all others (including none) vs. MVR	0.53 (0.22 to 1.31)	0.1695
Procedure: CABG vs. MVR	0.70 (0.28 to 1.76)	0.4499
Procedure: CABG and AVR vs. MVR	0.56 (0.16 to 1.88)	0.3426
Procedure: CABG and MVR vs. MVR	0.80 (0.24 to 2.63)	0.7144

TABLE 41 Cause of death listed by treatment arm

Cause of death by treatment arm	Months from randomisation to death
Maze procedure	
Haemorrhage	0.00
Pancreatitis	0.13
Unknown	0.16
Multisystem organ failure	0.59
Unknown	0.82
Subdural haematoma	1.28
Infective endocarditis	2.14
Sepsis	2.17

TABLE 41 Cause of death listed by treatment arm (*continued*)

Cause of death by treatment arm	Months from randomisation to death
Cardiac and renal failure	2.46
Bladder cancer	2.50
Cardiac failure and respiratory failure	3.29
Cardiac failure	3.48
Pneumonia and stroke	5.03
Pneumonia	5.75
Sepsis	7.26
Cardiac failure	10.51
Unknown	15.44
Multiorgan failure	15.64
Aspiration pneumonia	19.35
Pneumonia	21.75
Need more information on this, as it is not a cause of death	24.01
Cerebral haemorrhage	24.05
Cardiac failure	24.24
Stroke and cancer	35.94
Unknown	45.27
Unknown	45.83
Haematological cancer	50.66
Pulmonary embolus	54.50
MI	62.88
Unknown	64.62
Control	
MI	0.03
Bowel ischaemia	0.07
Pneumonia	0.10
Cardiac failure	0.13
Air embolus	0.16
Bowel perforation	0.16
Multiorgan failure	0.82
Multiorgan failure	1.05
Gastrointestinal haemorrhage	1.12
Bowel ischaemia	1.28
Cardiac failure	1.77
Sepsis	1.77
Cardiac failure	2.37
Sudden cardiac death	2.79

continued

TABLE 41 Cause of death listed by treatment arm (*continued*)

Cause of death by treatment arm	Months from randomisation to death
Haemorrhage	6.73
Haematological cancer	7.88
MI	21.78
Pneumonia	23.59
Unknown	40.21
Prostate cancer	46.85
Cardiac failure	46.91
Unknown	51.64
Cardiac failure	54.04
Cardiac failure	59.23
Endometrial cancer	63.44

TABLE 42 Summary of NYHA classifications at follow-up

	Treatment arm, *n* (%)		
Classification at each time point	Maze procedure (*n* = 176)	Control (*n* = 176)	Total (*n* = 352), *n* (%)
6 months			
I	66 (37.5)	86 (48.9)	152 (43.2)
II	53 (30.1)	49 (27.8)	102 (29.0)
III	19 (10.8)	15 (8.5)	34 (9.7)
IV	2 (1.1)	–	2 (0.6)
Missing/not known	36 (20.4)	26 (14.8)	62 (17.6)
12 months			
I	73 (41.5)	71 (40.3)	144 (40.9)
II	50 (28.4)	58 (33.0)	108 (30.7)
III	20 (11.4)	15 (8.5)	35 (9.9)
IV	1 (0.6)	–	1 (0.3)
Missing/not known	32 (18.1)	32 (18.2)	64 (18.2)
24 months			
I	68 (38.6)	64 (36.4)	132 (37.5)
II	41 (23.3)	56 (31.8)	97 (27.6)
III	19 (10.8)	16 (9.1)	35 (9.9)
IV	3 (1.7)	2 (1.1)	5 (1.4)
Missing/not known	45 (25.6)	38 (21.6)	83 (23.6)

TABLE 43 Summary of differences in baseline-adjusted SF-36 scores (maze procedure vs. control) at follow-up

SF-36 item score at each time point	Estimate (95% CI)	$p > \|t\|$
Bodily pain		
6 months	–1.010 (–5.539 to 3.520)	0.6619
12 months	–0.353 (–3.585 to 2.878)	0.8302
24 months	0.959 (–4.352 to 6.271)	0.7231
General health		
6 months	–0.292 (–3.860 to 3.275)	0.8723
12 months	–0.769 (–3.314 to 1.775)	0.5531
24 months	–1.723 (–5.907 to 2.461)	0.4191
Mental health		
6 months	–0.183 (–3.022 to 2.655)	0.8991
12 months	–0.374 (–2.399 to 1.651)	0.7171
24 months	–0.755 (–4.082 to 2.572)	0.6561
Physical functioning		
6 months	–0.619 (–4.963 to 3.725)	0.7798
12 months	–0.825 (–3.923 to 2.274)	0.6015
24 months	–1.236 (–6.331 to 3.858)	0.6340
Role emotional		
6 months	–4.050 (–10.837 to 2.736)	0.2418
12 months	–3.013 (–7.860 to 1.834)	0.2228
24 months	–0.939 (–8.899 to 7.021)	0.8170
Role physical		
6 months	–3.761 (–11.461 to 3.939)	0.3380
12 months	–2.776 (–8.271 to 2.720)	0.3218
24 months	–0.806 (–9.837 to 8.224)	0.8609
Social functioning		
6 months	–0.921 (–5.504 to 3.662)	0.6934
12 months	–0.625 (–3.894 to 2.644)	0.7074
24 months	–0.034 (–5.409 to 5.341)	0.9901
Vitality		
6 months	2.641 (–1.051 to 6.333)	0.1606
12 months	1.534 (–1.100 to 4.168)	0.2533
24 months	–0.680 (–5.006 to 3.646)	0.7578

TABLE 44 Summary of SF-36 component summary scores at baseline and follow-up

SF-36 item at each time point	Treatment arm		Total ($n = 352$)
	Maze procedure ($n = 176$)	Control ($n = 176$)	
PCS (UK norm: 50, SD 10)			
Baseline			
Mean score (SD)	30.18 (13.17)	31.00 (13.56)	30.59 (13.36)
Missing	3	2	5
6 months			
Mean score (SD)	38.25 (13.95)	39.23 (13.86)	38.75 (13.89)
Missing	21	17	38
12 months			
Mean score (SD)	38.55 (14.33)	39.99 (13.18)	39.27 (13.76)
Missing	26	25	51
24 months			
Mean score (SD)	38.28 (14.66)	39.54 (12.88)	38.92 (13.78)
Missing	41	37	78
MCS (UK norm: 50, SD 10)			
Baseline			
Mean score (SD)	50.81 (9.92)	49.58 (10.69)	50.19 (10.32)
Missing	3	2	5
6 months			
Mean score (SD)	54.37 (9.47)	54.06 (9.42)	54.21 (9.43)
Missing	21	17	38
12 months			
Mean score (SD)	54.09 (9.64)	54.33 (8.83)	54.21 (9.23)
Missing	26	25	51
24 months			
Mean score (SD)	54.68 (9.14)	54.44 (9.12)	54.56 (9.11)
Missing	41	37	78

TABLE 45 Summary of EQ-5D-3L utility scores in surviving patients at follow-up

EQ-5D-3L utility at each time point	Treatment arm		Total (*n* = 352)
	Maze procedure (*n* = 176)	Control (*n* = 176)	
Baseline			
Mean score (SD)	0.74 (0.22)	0.75 (0.21)	0.75 (0.22)
Missing	2	1	3
Discharge			
Mean score (SD)	0.63 (0.27)	0.64 (0.24)	0.63 (0.25)
Missing/died	20	20	40
6 weeks			
Mean score (SD)	0.77 (0.21)	0.78 (0.19)	0.77 (0.20)
Missing/died	18	17	35
6 months			
Mean score (SD)	0.80 (0.21)	0.80 (0.23)	0.80 (0.22)
Missing/died	21	16	37
12 months			
Mean score (SD)	0.79 (0.22)	0.81 (0.20)	0.80 (0.21)
Missing/died	27	25	52
24 months			
Mean score (SD)	0.78 (0.22)	0.80 (0.24)	0.79 (0.23)
Missing/died	41	33	74

TABLE 46 Summary of AEs by treatment received

AE	Treatment arm, *n* (%)		Total, *n* (%)
	Maze procedure	Control	
Total reported: events (patients)	560 (136)	589 (157)	1149 (293)
Cardiac complications	151 (86)	194 (108)	345 (194)
Arrhythmia	78 (58)	108 (80)	186 (138)
Bleeding	21 (14)	23 (16)	44 (30)
Cardiac failure	12 (6)	26 (20)	38 (26)
Cardiac arrest	1 (1)	6 (6)	7 (7)
MI	2 (2)	1 (1)	3 (3)
Other cardiac event	37 (31)	30 (19)	67 (50)
Respiratory complication	77 (51)	70 (47)	147 (98)
Infection	72 (49)	63 (49)	135 (98)
Prolonged hospitalisation	60 (48)	53 (39)	113 (87)
Neurological complication	25 (20)	22 (20)	47 (40)

continued

TABLE 46 Summary of AEs by treatment received (*continued*)

AE	Treatment arm, *n* (%)		Total, *n* (%)
	Maze procedure	Control	
Death	16 (16)	19 (19)	35 (35)
Renal failure	8 (5)	8 (6)	16 (11)
Thromboembolic event	3 (3)	5 (5)	8 (8)
Vascular complication	2 (2)	6 (5)	8 (7)
Multiple organ failure	2 (2)	3 (3)	5 (5)
Inflammation	3 (2)	–	3 (2)
Other non-cardiac event	141 (70)	146 (86)	287 (156)

TABLE 47 Estimated effect sizes (maze procedure vs. control) and 95% CIs from each of the secondary analyses in the HESTER substudy

Effect	Estimated effect size (95% CI)	*p*-value
Four-chamber echocardiography (primary analysis)		
ALAEF (%)	–8.03 (–12.43 to –3.62)	0.0015
LAV_{min} (ml)[a]	14.53 (–1.75 to 30.81)	0.0765
LAV_{max} (ml)[a]	10.02 (–9.40 to 29.44)	0.2889
LAV_{preA} (ml)[a]	11.30 (–7.71 to 30.30)	0.2244
Passive stroke volume (ml)	0.86 (–2.26 to 3.98)	0.5651
Active stroke volume (ml)	–3.44 (–7.79 to 0.90)	0.1120
LAEF (%)	–6.90 (–11.98 to –1.82)	0.0111
E/A ratio[a]	0.89 (0.16 to 1.62)	0.0205
Two-chamber echocardiography		
ALAEF (%)	–3.48 (–8.45 to 1.49)	0.1545
LAV_{min} (ml)[a]	9.95 (–0.48 to 20.37)	0.0600
LAV_{max} (ml)[a]	9.45 (–7.04 to 25.93)	0.2393
LAV_{preA} (ml)[a]	6.92 (–9.28 to 23.13)	0.3728
Passive stroke volume (ml)	1.49 (–2.32 to 5.30)	0.4138
Active stroke volume (ml)[a]	–1.54 (–9.09 to 6.02)	0.6671
LAEF (%)	–3.36 (–7.86 to 1.13)	0.1310
Multiple-slice MRI		
LAV_{min} (ml)[a]	39.71 (9.05 to 70.37)	0.0146
LAV_{max} (ml)[a]	35.96 (–1.37 to 73.30)	0.0579
LAEF (%)[a]	–13.46 (–19.87 to –7.06)	0.0004

a Analyses performed with pairing as a fixed effect.
All volumes are in ml.

TABLE 48 Mean (SD) of all primary and secondary end points measured using echocardiography (two-chamber view) in the HESTER substudy

	Treatment arm, mean (SD)	
End point	Maze procedure ($n = 22$)	Control ($n = 22$)
ALAEF (%)[a]	19 (8)	24 (8)
LAEF (%)[b]	30 (8)	35 (8)
LAV_{min} (ml)[b]	63 (15)	51 (12)
LAV_{preA} (ml)[a]	77 (17)	68 (17)
LAV_{max} (ml)[b]	90 (19)	79 (17)
Passive stroke volume (ml)[a]	13 (7)	11 (4)
Active stroke volume (ml)[a]	15 (6)	17 (8)

a Two control patients missing.
b One control patient missing.

TABLE 49 Mean (SD) of all primary and secondary end points measured using MRI in the HESTER substudy

	Treatment arm, mean (SD)	
End point	Maze procedure ($n = 22$)	Control ($n = 22$)
LAEF (%)	24 (7)	37 (6)
LAV_{max} (ml)	120 (36)	88 (31)
LAV_{min} (ml)	91 (32)	55.23 (19)

EME
HS&DR
HTA
PGfAR
PHR

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

Published by the NIHR Journals Library