The economic benefit of hip replacement: a 5-year follow-up of costs and outcomes in the Exeter Primary Outcomes Study

Richard Fordham, Jane Skinner, Xia Wang, John Nolan, the Exeter Primary Outcome Study Group

ABSTRACT

Objectives: To assess changes in quality of life and costs of patients undergoing primary total hip replacement using the Exeter prosthesis compared with a hypothetical ‘no surgery’ group.

Design: The incremental quality of life, quality-adjusted life years (QALYs) and cost of Exeter Primary Outcomes Study patients was compared with hypothetical ‘no surgery’ group over 5 years. Scores from annual SF-36 assessments were converted into utility scores using an established algorithm and the QALY gains calculated from pre-operative baseline scores. Costs included implant costs and length of stay.

Setting: Secondary care hospitals.

Participants: Patients receiving a primary Exeter implant enrolled in five of seven Exeter Primary Outcomes Study centres.

Results: On average, patients gained around 0.8 QALYs over 5 years. Younger and male patients or those with lower body mass index and poorer Oxford Hip Scores were significantly associated with increased QALYs. Treatment costs for a primary episode of care were just over £5000 (95% CI £4588 to £5812) per patient. Compared with ‘no surgery’, the cost per QALY was £7182 (95% CI £6470 to £7678), and this remained stable when key cost parameters were varied. The most likely cost per QALY was between £7058 and £7220. Older patients (age 75+) cost more, mainly due to longer average hospital stays and had a higher cost per QALY, although this remained below £10 000.

Conclusions: 85% of cases had a cost of <$10 000 per QALY (with 70% having a cost per QALY under £10 000) compared with no surgery. Cases would be considered cost-effective under currently accepted thresholds (≤£25 000–£30 000) compared with ‘no surgery’. However, depending on age and severity, younger patients and more severe patients had below average cost per QALYs. These results help to confirm the long-term benefits and cost-effectiveness of total hip replacement in a wide variety of patients using well-established implant models such as the Exeter. However, further and ongoing economic appraisal of this and other models is required for comparative purposes.

INTRODUCTION

Hip replacement is one of a few cardinal and successful operations in the NHS and yet it has mainly gone unchallenged from a cost-effective perspective since its inception. However, with the current financial environment in the NHS, it is important to reassess the costs and benefits and to dispel any uncertainties surrounding its cost-effectiveness for the majority of patients.

Long-term studies of hip replacements have not provided good conclusive economic proof, which is surprising, given the longevity of the procedure. However, implant models come and go and many get modified, so the
field is not static for economic assessment. It is therefore important to assess those that are stable and have endured in practice over many years. The orthopaedic outcome literature is mainly based on studies of the implant’s survival and few have incorporated a full economic analysis. Most economic studies nowadays rely on modelling, given the limitations of the randomised controlled trial design and well-designed studies compare alternative models. Good economic studies of total hip replacement (THR) also need an appropriate measure of patient outcome (preferably in terms of changes in health-related utility) as well as a robust costing.

PATIENTS AND METHODS

The Exeter Primary Outcomes Study (EPOS) is one of the largest longitudinal studies of a single prosthesis undertaken in the UK and continues to follow-up recipients for 10 years post-operatively. In this seven-centre study, a case series of 1589 patients underwent hip replacement with the Exeter implant between March 1999 and February 2002. This retrospective economic study was undertaken at the 5-year follow-up stage in patients who had the required outcome data available. These patients were compared with a hypothetical ‘no surgery’ group in terms of their additional costs and outcomes (quality of life (QoL)) from baseline.

The cost of an episode of care (including the operation) was estimated using the average cost of the implant (in the year the operation was performed) and the actual number of bed days used per patient at the Department of Health reference cost per day (for the year in question). The SF-36 patient outcome score was collected annually in the main study. Provided patients did not die or were not lost to follow-up, the change in their QoL between pre-operative assessment and 5-year follow-up was assessed.

Due to a lack of a control group, the quality-adjusted life year (QALY) gain could only be compared hypothetically with the QoL estimates that might have prevailed without surgery. For this, we decided to use a patient’s pre-operative QoL (as measured by the SF-36 score) as the counterfactual scenario. We recognised that in reality, without surgery, QoL might have improved or possibly even deteriorated.

The SF-36 is a multipurpose short-form health survey of 36 questions that yields an eight-scale profile of functional health and well-being as well as psychometrically based physical and mental health summary measures. However, in the EPOS, completion of this was optional and only 938 patients had sufficiently complete scores for the economic study. We calculated the overall SF-36 score following the user’s manual guidance. However, the SF-36 does not directly provide a utility score, so we used the Brazier algorithm to convert SF-36 questions to respective utility scores. This utility score captures a patient’s value for being in different health states on a scale from 0 (death) to 1 (perfect health). Where some patients did not have an SF-36 completed every year over the 5 years, we used their previous year’s SF-36 score to derive the utility score for the missing year. This approach seemed reasonable, as after the first post-operative assessment, most utility scores for the group as a whole did not change greatly over remaining years.

We calculated QALY gains made each year compared with the pre-operative baseline using an ‘area under the curve’ approach (see figure 1). In reality, other treatments might have improved this baseline score and reduced the net potential utility gain found in this study. On the other hand, we might have assumed the condition worsened. The direction of change could not be

| Characteristics | Minimum | Maximum |
|-----------------|---------|---------|
| Age, mean (SD)  | 62.2 (11.5) | 94.0 |
| Body mass index, mean (SD) | 27.2 (4.8) | 53.3 |
| Oxford Hip Score, mean (SD) | 44.2 (8.0) | 60.0 |
| Male, n (%)     | 363 (38.7) |         |

Figure 1 Schematic calculation of quality-adjusted life years (QALYs).
known with any certainty and therefore we assumed that no change occurred in QoL from baseline over the 5-year period. The QALY gain per patient was calculated up to the 5-year follow-up period or until the last annual review before death or revision (whichever occurred first). As the EPOS excluded patients who had revisions, we could only assume a zero QoL gain after revision had taken place (although this would be likely to be higher).

As resource use data were not collected in the main EPOS, we retrospectively constructed a proxy cost per case. This was based on each episode of care and was a composite of the national variable cost per day (for the year in which the surgery was performed) adjusted for the patient’s own length of stay plus the estimated total cost of the Exeter implant and other components added as the fixed cost element. An uplifted value of £7500 was used for the cost of a revision.3

Finally, the average cost per QALY was calculated for patients, assuming a zero cost for no surgery. Again this would have been unlikely but we felt it was a conservative assumption (as other treatment costs would be incurred in reality) reducing the cost difference between the two options.

### Statistical methods

Analyses were carried out in Stata V.10.8. Bootstrapped, bias-corrected methods were used to calculate 95% CIs for costs per QALY.9 Multiple linear regression was used to model QALY gains, with standard t-test and F-test used to evaluate the significance of β coefficients and model fit.

SF-36 dimension scores were calculated using recommended procedures. Missing values were replaced by scale means if valid responses were available for at least half of the scale items. For the items used in the utility scores, we used chained equations (ICE) to estimate the missing values based on the values of all the other variables in the data set.10 This was carried out when fewer than half of the values for the items used in the calculation of a utility score (from the SF-36) were missing in any individual questionnaire. However, where more than half of the values needed were missing, questionnaires were excluded from the analysis.

### RESULTS

The EPOS data set contained complete information on 1589 patients who received the Exeter implant or partial component. However, only five of the seven centres collected QoL data on 1087 patients. Of these, 938 (86%) had sufficient data to be included in the economic study. During the 5-year follow-up, there were 4598 potentially useable questionnaires from the 1087 patients. One thousand and eight of these had completed a baseline score and at least one other follow-up questionnaire completed in sufficient detail for utility...
estimation. Use of the multiple imputation method (see above) was made to adjust 344 of these surveys. Seventy people with important missing prognostic indicators were excluded, leaving 938 subjects for the main economic analysis. There was good SF-36 completion at the 5-year follow-up date (77%) with 720 patients having this maximum follow-up score. There were 69 deaths and only 17 revisions in the study population within the 5-year follow-up period.

Table 1 shows the characteristics of patients included in the economic study.

Nearly two-thirds of patients were women. The mean age of all cases was 62 years, but there was a large SD and upper and lower ranges. The average patient was not obese before surgery, although there was considerable variation in body mass index (BMI). The average patient fell into the third quintile of severity on the Oxford Hip Score (OHS).

Changes in utility scores from baseline to Year 5 are shown in table 2 and figure 2. The change in utility scores varied little after the initial large gain in the first post-operative year. The largest component of the increase in overall utility (around 0.18) was seen in the first year after operation. Much smaller changes were found in subsequent years. Both the overall SF-36 score and the individual dimensions that comprise it showed similar changes during this period. The largest changes on the SF-36 occurred in physical functioning (+47 point increase), physical role functioning (+50 points) and pain (+48 points).

The QALYs gained was calculated over this 5-year period shown in figure 3. The majority of patients (90.7%) gained positive QALYs compared with no surgery. These gains were approximately normal distributed around a mean value of 0.8 QALYs (95% CI 0.76 to 0.84). However, a small group of patients (9.3%, n=87) lost QALYs (in a theoretical sense, they would have been better without surgery).

In terms of estimating cost per episode, average length of stay was 10.8 days (SD 7.3) and the median estimated cost per patient was £5084 (IQR: £4588—£5812). The distribution of costs is shown in figure 4.

Table 3 shows the average QALYs gained and combined with the average cost to derive a cost per QALY. In order to take account of variation and uncertainty in these estimates, we calculated the associated CI using bootstrapping simulation methods. The average cost per QALY for all 938 subjects was £7182 (95% CI £6740 to £7678).

Table 4 shows the length of stay, estimated cost, QALY gain and cost/QALY by age group.

Table 4 Length of stay, estimated cost, QALY gain and cost/QALY by age group

- Bootstrapped, bias-corrected CI, based on 10000 bootstrap replications. QALY, quality-adjusted life year.
Table 5  Length of stay, estimated cost, QALY gain and cost/QALY by Oxford Hip Score (OHS) quintile

| OHS (score band) | N  | Length stay/day | Estimated cost/£ | QALY gain | Cost/QALY |
|------------------|----|-----------------|------------------|-----------|-----------|
| Quintile 1 (13–37) | 191 | 10.2 | 5084 | 0.61 | 9188 |
| Quintile 2 (38–43) | 206 | 11.3 | 4981 | 0.83 | 7102 |
| Quintile 3 (44–47) | 174 | 10.4 | 4981 | 0.70 | 7907 |
| Quintile 4 (48–51) | 191 | 11.0 | 5233 | 0.98 | 6628 |
| Quintile 5 (52–60) | 176 | 11.2 | 5084 | 0.98 | 5924 |

*Bootstrapped, bias-corrected CI, based on 10 000 bootstrap replications.
QALY, quality-adjusted life year.

Table 6  Linear regression modelling of quality-adjusted life years (n=938)

| β Coefficient | SE (β) | p Value |
|---------------|--------|---------|
| Age (years)   |        |         |
| Centred age   | −0.009 | 0.002   | <0.001 |
| Centred age²/100 | −0.027 | 0.011   | 0.012  |
| Sex           | Female | −0.086 | 0.045   | 0.053  |
| Body mass index | −0.015 | 0.005   | 0.001  |
| Oxford Hip Score | 0.017 | 0.003   | <0.001 |

Model F-statistic (5 df) = 12.74, p<0.0001, R²=0.084.

We also analysed the QALYs and cost per QALY by age group as shown in Table 4. As might be expected, QALY gains were significantly lower in older patients, with the largest gains found in younger patients. Conversely, cost per QALY increased in older age groups because of the increased length of stay combined with a lower QALY gain.

With regard to the baseline OHS, we found that pre-operative severity was a good predictor of cost-effectiveness. The poorer the initial score on the OHS, the greater the QALY gain found and similarly the lower the cost per QALY. There were significant differences in the cost per QALY between quintile 1 (least severe) and quintiles 3, 4 and 5 (most severe) and also between quintiles 3 and 5 (see Table 5).

The QALY gain was approximately normally distributed and therefore linear regression could be carried out to determine patient and treatment characteristics associated with total QALYs. Age, BMI and OHS were significantly associated with QALYs gained (see Table 6).

Given the limitations of the cost data on which the study was based, a sensitivity analysis was undertaken to determine robustness of the cost and cost per QALY results (see Table 7).

Using various cost assumptions, mean estimated cost per case varied from £4950 to £5156. Given a small variation in cost, the cost per QALY remained relatively stable in the range of £7058–7220. This confirmed that the cost-effectiveness results were robust and insensitive to some relatively large changes in cost assumptions. This is also reassuring in terms of potential variability in costs between treatment centres and/or surgical practice that occurs in practice.

A cost per QALY threshold analysis is shown in Figure 5. Over 85% of cases had a cost per QALY of £20 000 or less with 70% of these having a cost per QALY under £10 000 thus making it very cost-effective when compared hypothetically with no surgery. However, 40 cases had a cost per QALY over £50 000. These patients were largely those where the QoL gain was very small rather than due to their cost being above average.

DISCUSSION

We have shown that from the perspective of the absence of surgery, the majority of EPOS subjects were treated cost-effectively. The value to patients in terms of their health utility and QALY gains has been demonstrated.
Figure 5  Probabilistic sensitivity analysis.

Based on reasonably conservative assumptions, the mean QALYs gained was 0.8 QALYs (95% CI 0.76 to 0.84), while the mean cost per hospital stay was just over £5000 per patient. Although these costs would have been more accurate if the study had been undertaken prospectively, comprehensive data allowed us to build a reasonably accurate cost profile for each patient. Most of this cost could be attributed to length of stay, although the study could not directly account for variability in the price of implant costs. Uncertainty surrounding the fixed cost data was examined using sensitivity analysis and its cost was shown to possibly increase to £5500 per case. Bootstrapping techniques increased the robustness of these findings by reducing bias by multiple replications of the primary study results.

In terms of cost per QALY, we have shown that THR may be more sensitive to optimal treatment and care in the most appropriate patient groups than to local variations in cost. But such results should be treated with caution. An actual alternative comparator implant rather than no surgery would have been more realistic, but no such data existed in this study. However, we deliberately made highly conservative assumptions both about cost and any likely net QALY gain. Furthermore, a probabilistic sensitivity analysis demonstrated that THR is likely to be good value for money even when willingness-to-pay thresholds are set quite low.

This study confirms what is perhaps implicitly assumed in every day orthopaedic practice that hip replacement (using a reliable implant) is worth doing for the majority of patients. The EPOS patients had their pain, function and ultimately their QoL improved by the Exeter hip, even those with above average age, disability and BMI.

RECOMMENDATIONS

Further modelling studies are still needed to establish the longer-term cost-effectiveness of THR.9 The most cost-effective implants will be those with the best survival rates (and hence the fewest revisions), with the best patient outcomes and the least cost. More studies of a comparative nature incorporating economic evaluation would immensely improve the still imperfect knowledge of the cost-effectiveness of different THR implants in today’s NHS.

Acknowledgements We are grateful to Stryker UK Ltd. and in particular, Mr David Forsythe for sponsoring this study and for his comments on earlier drafts of the paper and his support throughout. We are also grateful to Professor David Murray, Oxford University, Mr John Timperley, Exeter and the members of the EPOS Group for allowing us access to the data. The following are principal investigators of the EPOS group: Prof DW Murray, Mr G Andrew, Mr J Nolae, Prof DJ Beard, Mr P Gibson, Mr A Hamer, Mr M Fordyce and Mr K Tuson. The following are or have been study coordinators for the EPOS group: A Potter, A McGovern, K Reilly, C Jenkins, K Barker, A. Cooper, C. Darrah, L. Cawton, P. Inaparthy and C. Pitchfork. We would also like to thank Professor Alastair Gray, Director of the Health Economics Research Centre, Division of Public Health and Primary Care, University of Oxford and Helen Dakin, Researcher, Health Economics Research Centre, University of Oxford for their critical advice on later drafts.

Contributors RF designed the conception, economic evaluation, led the analysis and wrote the first and subsequent drafts. JS undertook all the statistical processing and analysis and worked with RF and JN on the interpretation of results. XW helped revise the paper for publication. JN provided clinical input to the study throughout. JN as representative of Exeter Primary Outcomes Study (EPOS) provided approval of this paper along with other members of the EPOS Team.

Funding Stryker UK Ltd.

Competing interests RF and JS received consultancy payments for the original economic analysis. RF is currently receiving a 2-year study grant from Stryker to undertake further work on the Outcomes and Costs of Hip Replacement evaluation (the ‘OCHRE’) project looking at long-term cost-effectiveness of the Exeter prosthesis.

Ethics approval Ethics approval was provided by EPOS.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Requests for data sharing need to be submitted to the Exeter Primary Outcomes Study Group coordinator initially.

REFERENCES

1. Briggs A, Sculpher M, Dawson J, et al. The use of probabilistic decision models in technology assessment: the case of total hip replacement. *Appl Health Econ Health Policy* 2004;3:79–89.
2. Nuijten MJ, Pronk MH, Brorens MJ, et al. Reporting format for economic evaluation: focus on modelling studies. *PharmacoEconomics* 1998;14:259–68.
3. Fitzpatrick R, Shortall E, Sculpher M, et al. Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technol Assess* 1998;2:1–64.
4. Wylde V, Blom A, Dieppe P, et al. Prevalence of Poor Patient-Reported Outcomes After Lower Limb Joint Replacement. Paper F185. 10th EFORT Congress. Vienna: European Federation of National Associations of Orthopaedics and Traumatology (EFORT), 2009.

5. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press, 2005.

6. Ware JE, Kosinski M, Keller SD. SF-36 Physical And Mental Health Summary Scales: a User's Manual. Boston, MA: Health Inst, 1994.

7. Brazier J, Roberts J, Deverill M, et al. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.

8. Stata Statistical Software [Program]. Release 10. College Station: StataCorp, 2007.

9. Campbell MK, Torgerson DJ. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. QJM 1999;92:177–82.

10. Royston P. Multiple imputation of missing values Stata J 2004;4:227–41.
## STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of cohort studies

| Section/Topic | Item # | Recommendation | Reported on page # |
|---------------|--------|----------------|--------------------|
| **Title and abstract** | 1 | (a) Indicate the study’s design with a commonly used term in the title or the abstract | p.1 |
|  |  | (b) Provide in the abstract an informative and balanced summary of what was done and what was found | p.2 |
| **Introduction** | 2 | Explain the scientific background and rationale for the investigation being reported | p.5 |
| **Objectives** | 3 | State specific objectives, including any prespecified hypotheses | p.5 |
| **Methods** | 4 | Present key elements of study design early in the paper | p.5-7 |
| **Setting** | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | p.5-7 |
| **Participants** | 6 | (a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up | p.5-7 |
|  |  | (b) For matched studies, give matching criteria and number of exposed and unexposed | n/a |
| **Variables** | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | p.5-7 |
| **Data sources/measurement** | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | p.5-7 |
| **Bias** | 9 | Describe any efforts to address potential sources of bias | p.5-7 |
| **Study size** | 10 | Explain how the study size was arrived at | p.5-7 |
| **Quantitative variables** | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | p.5-7 |
| **Statistical methods** | 12 | (a) Describe all statistical methods, including those used to control for confounding | p.7-8 |
|  |  | (b) Describe any methods used to examine subgroups and interactions | p.7-8 |
|  |  | (c) Explain how missing data were addressed | p.7-8 |
|  |  | (d) If applicable, explain how loss to follow-up was addressed | p.7-8 |
|  |  | (e) Describe any sensitivity analyses | p.7-8 |
| **Results** |  |  |  |
| Component            | Checklist Item | Description                                                                                     | Page |
|----------------------|----------------|-----------------------------------------------------------------------------------------------|------|
| Participants         | 13*            | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | p.8  |
|                      |                | (b) Give reasons for non-participation at each stage                                           | p.12 |
|                      |                | (c) Consider use of a flow diagram                                                             | n/a  |
| Descriptive data     | 14*            | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | p.12-13 |
|                      |                | (b) Indicate number of participants with missing data for each variable of interest             | p.8  |
|                      |                | (c) Summarise follow-up time (eg, average and total amount)                                   |      |
| Outcome data         | 15*            | Report numbers of outcome events or summary measures over time                                 | p.10 |
| Main results         | 16             | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | p.14 |
|                      |                | (b) Report category boundaries when continuous variables were categorized                      | p.12-13 |
|                      |                | (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | n/a  |
| Other analyses       | 17             | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses  | p.15 |
| Discussion           |                | Key results                                                                                    | p.15-16 |
|                      |                | Summarise key results with reference to study objectives                                        |      |
| Limitations          | 20             | Interpretation                                                                                 | p.2  |
|                      |                | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence |      |
| Generalisability     | 21             | Generalisability (external validity) of the study results                                       | 17   |
| Other information    |                | Funding                                                                                        | p.18 |
|                      |                | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based |      |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.
# STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of cohort studies

| Section/Topic         | Item # | Recommendation                                                                                                                                                                                                 | Reported on page # |
|-----------------------|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| **Title and abstract** | 1      | (a) Indicate the study’s design with a commonly used term in the title or the abstract                                                                                                                       | p.1                |
|                       |        | (b) Provide in the abstract an informative and balanced summary of what was done and what was found                                                                                                       | p.2                |
| **Introduction**      | 2      | Explain the scientific background and rationale for the investigation being reported                                                                                                                        | p.5                |
| **Objectives**        | 3      | State specific objectives, including any prespecified hypotheses                                                                                                                                             | p.5                |
| **Methods**           | 4      | Present key elements of study design early in the paper                                                                                                                                                      | p.5-7              |
|                       | 5      | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection                                                                               | p.5-7              |
|                       | 6      | (a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up                                                                              | p.5-7              |
|                       |        | (b) For matched studies, give matching criteria and number of exposed and unexposed                                                                                                                      | n/a                |
|                       | 7      | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable                                                                         | p.5-7              |
|                       | 8*     | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group                                | p.5-7              |
| **Bias**              | 9      | Describe any efforts to address potential sources of bias                                                                                                                                                    | p.5-7              |
| **Study size**        | 10     | Explain how the study size was arrived at                                                                                                                                                                    | p.5-7              |
| **Quantitative variables** | 11   | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why                                                                                 | p.5-7              |
| **Statistical methods** | 12   | (a) Describe all statistical methods, including those used to control for confounding                                                                                                                        | p.7-8              |
|                       |        | (b) Describe any methods used to examine subgroups and interactions                                                                                                                                       | p.7-8              |
|                       |        | (c) Explain how missing data were addressed                                                                                                                                                               | p.7-8              |
|                       |        | (d) If applicable, explain how loss to follow-up was addressed                                                                                                                                         | p.7-8              |
|                       |        | (e) Describe any sensitivity analyses                                                                                                                                                                      | p.7-8              |
| **Results**           |        |                                                                                                                                                                                                            |                    |
|Participants| 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed p.8 |
|---|---|---|
| | | (b) Give reasons for non-participation at each stage p.12 |
| | | (c) Consider use of a flow diagram n/a |
|Descriptive data| 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders p.12-13 |
| | | (b) Indicate number of participants with missing data for each variable of interest p.8 |
| | | (c) Summarise follow-up time (eg, average and total amount) |
|Outcome data| 15* | Report numbers of outcome events or summary measures over time p.10 |
|Main results| 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included p.14 |
| | | (b) Report category boundaries when continuous variables were categorized p.12-13 |
| | | (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period n/a |
|Other analyses| 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses p.15 |
|Discussion| | |
|Key results| 18 | Summarise key results with reference to study objectives p.15-16 |
|Limitations| | |
|Interpretation| 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence p.2 |
|Generalisability| 21 | Discuss the generalisability (external validity) of the study results 17 |
|Other information| | |
|Funding| 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based p.18 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.