PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Rationing of total knee replacement: A cost-effectiveness analysis on a large trial dataset
AUTHORS	Helen Dakin, Alastair Gray, Ray Fitzpatrick, Graeme MacLennan and David Murray

VERSION 1 - REVIEW

REVIEWER	Mark Pennington Lecturer in Health Economics London School of Hygiene & Tropical Medicine UK No competing interests
REVIEW RETURNED	31/08/2011

RESULTS & CONCLUSIONS	The authors cite previous evaluations of the cost-effectiveness of knee replacement but do not discuss. It seems likely that no previous analysis has undertaken similar subgroup analysis.
GENERAL COMMENTS	This is a timely piece of work. I found the script very clear and the analysis appeared to be thorough. Discussion of the regression analysis is minimal. For instance did the authors consider and reject higher order powers of age in the regression model? I suspect the relationship between EQ5D scores and age is not linear over the range 55-85. Were other specifications considered in a sensitivity analysis? I was also surprised that revisions were not mentioned. Presumably some of the KAT trial participants underwent revision of the knee in the subsequent five years? My guess is that revisions are lower in this trial population than those observed in practice and pulished in the National Joint Register reports. Is there a clinical argument for delaying knee surgery in younger patients with higher OKS scores (and hence less incapacitating disease) in order to reduce the risk of a revision? Overall I think consideration of revision risk (probably necessitating some sort of extrapolation with a Markov model) would be very unlikely to change the conclusions. However, I did wonder if it was something that might be addressed in the discussion. I noted two minor errors. There is a typo in the footnote to table 2 where the explanataion of the ICER erroneously refers to ASAgrade3or4. Also the baseline EQ5D data in Table 1 is wrong for all groups except the summary column at the far right. I would consider the above points minor and for the authors consideration at their discretion. There is already evidence that the PROMs data on joint replacement is being used inappropriately to justify restricting access to highly effective surgery. This is an important piece of work and I hope the editors can expedite publication.

REVIEWER	M. Nuñez, PhD Investigator IDIBAPS Hospital Clinic of Barcelona, Spain I DECLARE THAT I HAVE NO CONFLICTS OF INTEREST WITH RESPECT TO THIS MANUSCRIPT
REVIEW RETURNED	13/12/2011

GENERAL COMMENTS	This is an interesting, well-conducted, well-written cost-effectiveness analysis of TKR from the NHS perspective. The analysis is sound and the conservative assumptions made enhance the validity of the results. The Discussion is relevant and focused on the results. The authors suggest that decisions about suitability for TKR can be based on OKS alone. This interesting suggestion could be discussed further. The finding that obesity had no effect on the cost-effectiveness is interesting in light of the controversy that exists with respect to this type of patient and could be discussed further. For example, Michelle M. Dowsey et al in Arthritis Care & Research ,2011; Vol 63, No 10, pp 1375-1381. The authors suggest that for most patients, TKR is cost-effective and that, using less-conservative assumptions, TKR is cost-effective in nearly all patients. This may well be so, but, in the current economic climate it might be useful for the authors to provide some policy guidance as a conclusion of the study in order to stimulate debate. To this end, it might be useful to include in the text some examples of procedures where high costs per QALY gained are allowed through, and not restrict themselves to providing only references, although perhaps discretion is required on this point. The authors could also discuss whether, in spite of acceptable cost-effectiveness figures, TKR is actually indicated in all these patients, especially those with high OKS. The last sentences on unicompartmental knee replacement seem to be irrelevant to the study and could be replaced by conclusions. The authors state that 2532 patients were originally studied and that 121 were excluded (108+13). Later, they state that 2131 patients were finally analysed. What happened to the remaining 100 patients? In the Methods section, the authors should explain more clearly the scoring methods for the OKS and EQ-5D (ie. higher is better or worse?) and also explain clearly what the EQ-5d actually represents and what the scores mean. This is important as the tables and figures may be confusing for non-statisticians who may not appreciate what they mean. Minor Points Page 7, line 59: " 108 rheumatoid patients and 13 patients of ASA grade 4" would be better as "108 rheumatoid arthritis patients and 13

	patients with ASA grade 4". (same applies to Abstract, line 31). Page 8, line 23: "We also calculated number" would be better as "We also calculated the number" On page 8, line 21, the authors state that 45% of patients were female. This is curious, as in most series of this type, females predominate. Could the authors clarify and comment on this. Table 1. 1) Baseline EQ-5D for all patients is given as 0.39. However, baseline decile scores are all much higher - could the authors clarify this. 2) Gains in EQ-5D in the second and successive years are only seen in deciles 8 and 9 (20% of patients). In year 1, gains are only seen in deciles 3,4,6,8,9 and 10 (62% ofpatients) Could the authors comment on this. 3. It would be useful to have the OKS scores for all measures at 3 months and successive years in order that the reader could see the gains after TKR in terms of the health status. Table 2. The footnote states ASA grade3 or 4, while on page 7 it is stated that patients with grade 4 were excluded.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Mark Pennington: Lecturer in Health Economics, London School of Hygiene & Tropical Medicine UK
The authors cite previous evaluations of the cost-effectiveness of knee replacement but do not discuss. It seems likely that no previous analysis has undertaken similar subgroup analysis.
In response to this comment, we have expanded the discussion of previous research in the discussion to state "Our study also confirms previous findings[7, 9, 16] that costs and benefits vary with age and comorbidity, but evaluates such characteristics, and pre-operative knee function, in multivariate analyses, which has not (to our knowledge) been done previously"

This is a timely piece of work. I found the script very clear and the analysis appeared to be thorough. Discussion of the regression analysis is minimal. For instance did the authors consider and reject higher order powers of age in the regression model? I suspect the relationship between EQ5D scores and age is not linear over the range 55-85. Were other specifications considered in a sensitivity analysis?
We conducted several sensitivity analyses on the regression model, including higher order powers of age. These suggested that the effect of age may not be linear, as an age-squared variable was statistically significant. However, including an age-squared variable had no effect on the threshold OKS level at which TKR ceased to be cost-effective, so the age-squared term was omitted from the final model for parsimony. In response to this helpful comment, we have mentioned these sensitivity analyses in the results (page 12, paragraph 2).

I was also surprised that revisions were not mentioned. Presumably some of the KAT trial participants underwent revision of the knee in the subsequent five years? My guess is that revisions are lower in this trial population than those observed in practice and pulished in the National Joint Register reports. Is there a clinical argument for delaying knee surgery in younger patients with higher OKS scores (and hence less incapacitating disease) in order to reduce the risk of a revision? Overall I think consideration of revision risk (probably necessitating some sort of extrapolation with a Markov model) would be very unlikely to change the conclusions. However, I did wonder if it was something that might be addressed in the discussion.
Data on the cost of readmissions, surgery and knee components for all revisions occurring within five-

years of primary TKR were collected in the KAT trial and included in the costing analysis. In response to this comment, we have stated this explicitly on page 6 (paragraph 2), as well as on page 9. KAT used a highly pragmatic design, with no restrictions on patient entry criteria and leaving all aspects of care apart from the randomised comparisons (including length of hospital stay, post-operative rehabilitation and revision) at the surgeons' discretion. Since the design mirrors routine clinical practice and no significant differences in revision rates have (to date) been observed between randomised comparisons,1 there is no reason to suppose that the rate or type of revisions would differ from that in registries. We have added this strength of the study on page 13 (paragraph 1).

As the costs and quality of life changes associated with revisions were included in the analysis, any effect of age on revision rates is already taken into account in our analysis, so we do not feel that it is appropriate to discuss delays to surgery in younger patients: particularly since patients' quality of life and functional status tends to decrease over time in the absence of joint replacement, giving worsening final post-operative function.

I noted two minor errors. There is a typo in the footnote to table 2 where the explanation of the ICER erroneously refers to ASAgrade3or4. Also the baseline EQ5D data in Table 1 is wrong for all groups except the summary column at the far right.
Thank you very much for spotting these errors, which have now been corrected. We have checked the paper again to ensure that all tables, figures and results are correct.

I would consider the above points minor and for the authors consideration at their discretion. There is already evidence that the PROMs data on joint replacement is being used inappropriately to justify restricting access to highly effective surgery. This is an important piece of work and I hope the editors can expedite publication.

Reviewer: M. Nuñez, PhD; Investigator IDIBAPS; Hospital Clinic of Barcelona, Spain
This is an interesting, well-conducted, well-written cost-effectiveness analysis of TKR from the NHS perspective. The analysis is sound and the conservative assumptions made enhance the validity of the results. The Discussion is relevant and focused on the results.

The authors suggest that decisions about suitability for TKR can be based on OKS alone. This interesting suggestion could be discussed further.
The main concern of this paper is on rationing and on evaluating the impact of the restrictions to TKR access based on OKS that some PCTs have recently introduced. The paper has therefore focused on the OKS. In addition to the variables considered in this analysis, the indications and contraindications for TKR include a variety of other individual considerations (including radiographic findings and patient choice) that are harder to quantify and are beyond the scope of our paper. Although our research suggests that age, sex, BMI and number of joints affected have little effect on cost-effectiveness, it would not be appropriate to assess suitability for TKR based on OKS alone. We have added a comment to this effect on page 14 (paragraph 1)

The finding that obesity had no effect on the cost-effectiveness is interesting in light of the controversy that exists with respect to this type of patient and could be discussed further. For example, Michelle M. Dowsey et al in Arthritis Care & Research ,2011; Vol 63, No 10, pp 1375-1381.
We thank the reviewer to bringing this paper to our attention. We have cited this paper in the introduction (page 4, paragraph 1) and in the discussion (page 14, paragraph 2).

The authors suggest that for most patients, TKR is cost-effective and that, using less-conservative assumptions, TKR is cost-effective in nearly all patients. This may well be so, but, in the current economic climate it might be useful for the authors to provide some policy guidance as a conclusion

of the study in order to stimulate debate. To this end, it might be useful to include in the text some examples of procedures where high costs per QALY gained are allowed through, and not restrict themselves to providing only references, although perhaps discretion is required on this point. The authors could also discuss whether, in spite of acceptable cost-effectiveness figures, TKR is actually indicated in all these patients, especially those with high OKS.
The main policy implication of our paper is that if access to TKR is to be rationed, any restrictions should be based on the best available evidence on outcomes and cost-effectiveness, and we find no evidence to support the rationing already being proposed or implemented by UK PCTs. Given the small number of patients in KAT with very high OKS, and that our assumptions have been very conservative we feel that it would be premature to specify a maximum OKS beyond which TKR is not cost-effective. We also feel that it is not necessary to explicitly name examples of treatments recommended with higher ICERs, since the treatments affected by NICE's end of life guidance (such as sunitinib) are already well-known and the number of treatments recommended with ICERs above £10,697/QALY gained are too numerous to list. Furthermore, we feel that the paper is already likely to elicit extensive debate on this topic.

In response to this helpful comment, we have added a short conclusions paragraph at the end of the discussion and a sentence on page 14 (paragraph 1).

The last sentences on unicompartmental knee replacement seem to be irrelevant to the study and could be replaced by conclusions.
In response to this comment, we have condensed the discussion on unicompartmental knee replacement and added a short paragraph summarising the conclusions at the end of the discussion.

The authors state that 2532 patients were originally studied and that 121 were excluded (108+13). Later, they state that 2131 patients were finally analysed. What happened to the remaining 100 patients?
The 100 patients were omitted from all economic analyses as they either died or withdrew from the trial before surgery, or were randomised in the total vs unicompartmental comparison, for which follow-up was discontinued. We have indicated this on page 8, paragraph 1.

In the Methods section, the authors should explain more clearly the scoring methods for the OKS and EQ-5D (ie. higher is better or worse?) and also explain clearly what the EQ-5D actually represents and what the scores mean. This is important as the tables and figures may be confusing for non-statisticians who may not appreciate what they mean.
In response to this comment, we expanded the sentences describing the two scores (page 6, paragraph 1) and additional footnotes beneath Table 1 and Figure 1.

Minor Points

Page 7, line 59: "108 rheumatoid patients and 13 patients of ASA grade 4" would be better as "108 rheumatoid arthritis patients and 13 patients with ASA grade 4". (same applies to Abstract, line 31). These changes have been made.

Page 8, line 23: "We also calculated number" would be better as "We also calculated the number"
We thank the reviewer for alerting us to this typo, which has been corrected.

On page 8, line 21, the authors state that 45% of patients were female. This is curious, as in most series of this type, females predominate. Could the authors clarify and comment on this.
The majority of the cohort is indeed female, rather than male. Unfortunately the sex variable was mislabelled, which has now been corrected throughout. Thank you very much for alerting us to this error.

Table 1.
1) Baseline EQ-5D for all patients is given as 0.39. However, baseline decile scores are all much higher - could the authors clarify this.
Unfortunately there was an error in the original version of Table 1 submitted in that 3-month EQ-5D results were erroneously inserted in the baseline row. Thank you very much for alerting us to this. The error has now been corrected and we have checked the paper again to ensure that all tables, figures and results are correct.

2) Gains in EQ-5D in the second and successive years are only seen in deciles 8 and 9 (20% of patients). In year 1, gains are only seen in deciles 3,4,6,8,9 and 10 (62% ofpatients) Could the authors comment on this.
This was again due to the error in Table 1, which has now been corrected. Mean EQ-5D utility increases in all 10 deciles, as shown in Figure 1 and the revised Table 1.

3. It would be useful to have the OKS scores for all measures at 3 months and successive years in order that the reader could see the gains after TKR in terms of the health status.
OKS scores at different time points after TKR have been published previously for the whole cohort.[1 2] Furthermore, since post-baseline OKS was not included in multiple imputation, it would not be possible to present OKS data at each timepoint for the whole cohort without rerunning the multiple imputation and all results presented in the paper. We have therefore not added OKS into the paper.

Table 2. The footnote states ASA grade3 or 4, while on page 7 it is stated that patients with grade 4 were excluded.
We thank the reviewer for alerting us to this error, which has now been corrected.

References
1. Breeman S, Campbell M, Dakin H, Fiddian N, Fitzpatrick R, Grant A, et al. Patellar Resurfacing in Total Knee Replacement: Five-Year Clinical and Economic Results of a Large Randomized Controlled Trial. J Bone Joint Surg Am 2011;93:1473-81.
2. Johnston L, MacLennan G, McCormack K, Ramsay C, Walker A. The Knee Arthroplasty Trial (KAT) design features, baseline characteristics, and two-year functional outcomes after alternative approaches to knee replacement. J Bone Joint Surg Am 2009;91(1):134-41.