Author's response to reviews

Title: The impact of preoperative patient characteristics on the cost-effectiveness of total hip replacement: a cohort study

Authors:

Matthias Vogl (matthias.vogl@helmholtz-muenchen.de)
Rainer Wilkesmann (rainer.wilkesmann@barmherzige-muenchen.de)
Christian Lausmann (christian.lausmann@barmherzige-muenchen.de)
Werner Plötz (werner.ploetz@barmherzige-muenchen.de)

Version: 3
Date: 29 May 2014

Author's response to reviews: see over
RESPONSE TO REVIEWER’S COMMENTS:

MANY THANKS FOR BOTH REVIEWERS. THEIR COMMENTS ARE NOW ADDRESSED.

Reviewer #1: Page 1-3
Reviewer #2: Page 4-6

MODIFICATIONS IN THE TEXT ARE HIGHLIGHTED IN YELLOW.

Reviewer #1:

Major Compulsory Revisions

I do not see how replacing preoperative EQ-5D with the WOMAC removes the regression to the mean problem. Both are measuring health state, so it would be as true for the general measure as for the specific one that patients at the extremes of the distribution are more likely to move towards its mean in future observations. I can understand using a disease specific measure to divide the sample more finely, but that is a distinct justification. It would be better to accept regression to the mean as a limitation arising inevitably from comparing sub-populations by health state. At present I get the impression from the manuscript the problem ahb been solved.

Response: Indeed we could not remove the regression to the mean problem. It is a general limitation of cohort studies when analyzing subgroups. By using the WOMAC score to distinguish between subgroups, we wanted to provide more meaningful subgroups for orthopaedic surgeons. For total hip replacement patients, the WOMAC is one of the standard patient measure and more often used than the EQ-5D questionnaire. Finally, with using the WOMAC instead of the EQ-5D in grouping patients, we also wanted to slightly reduce the regression to the mean problem for QALY calculation. However, as WOMAC and EQ-5D are correlated, we have to accept it as a limitation of the study. To not mislead in the methods section and discussion section we changed the reporting of the regression to the mean problem. Thank you for this hint.

CHANGES: For patient differentiation we used the WOMAC as a standardized and specific, patient-based reporting measure. As WOMAC and EQ-5D are correlated we could not eliminate the “regression to the mean” issue [43], a general limitation of cohort studies when analyzing subgroups. [page 7]

However, as WOMAC and EQ-5D utility scores are correlated, we could not eliminate the general limitation of regression toward the mean, potentially leading to inaccurate results when comparing preoperative patient groups on costs and QALYs. [page 11]

Despite the reasonable concern about regression to the mean, the ratio results for different subgroups don’t seem to imply this is dominant. Instead, there is a trend of more healthy patients receiving more QALYs and being cheaper than less healthy ones. This could be more clearly drawn out, as noted below, and is also quite surprising. While it is intuitive to me that healthier patients are cheaper to treat, I would expect them to receive less benefit from that treatment, since there is less scope for benefit. Some discussion of this outcome would therefore add value to this paper, particularly as one of its advances is this subgroup analysis.
Response: There is a trend that healthier patients cost less. We mention this now in the results and discussion section. We have also the trend that patients with lower (poorer) preoperative WOMAC scores receive more QALYs. This was probably misleading and we changed the terms lower/higher scores into poorer/adequate and good where applicable. Indeed QALY gains are higher when patients are preoperatively less healthy. We now clearly indicate which health state is poor and which is good. The changes were made through the whole paper. E.g., on page 9:

CHANGES: The grouping by preoperative WOMAC score shows a trend that healthier patients cost less (Table 4). [page 9]

All three value sets for QALY and QALY-like calculation showed a large increase in QALY gains with a decrease in preoperative WOMAC scores, indicating that patients with poor WOMAC scores benefit most from THR (Table 4). QALY gains are higher when patients are less healthy. [page 9]

In modelling QALYs in Table 3, EQ-5D pre-op is included but not WOMAC. The model replacing pre-op EQ-5D with WOMAC is needed to support the assertion in the Results section that this is the most important factor, along with pre-op EQ-5D. Similarly, I do not see the correlation between WOMAC and cost referred to. This also needs to be included to justify creating sub groups with this variable.

Response: When modelling QALYs, WOMAC is dominated by EQ-5D. Both are correlated and when using backward regression methods WOMAC is inevitably removed from the regression when EQ-5D is included. We did not force variables to stay in the regression any more now and because of possible multi-collinearity issues with WOMAC and EQ-5D used two separate regressions now. We have long thought about using EQ-5D or WOMAC for the grouping and calculated the whole Table 4 with all QALYs for both groupings with WOMAC and EQ-5D. However, because of limiting the regression to the mean issue and because WOMAC is much more meaningful in an orthopaedic environment we chose to use WOMAC as the grouping variable and not EQ-5D tariff. In many orthopaedic departments WOMAC is already included in standardized surveys but EQ-5D isn’t. In day-to-day decision making, WOMAC is better accepted among orthopaedic surgeons and for comparison to other studies the WOMAC grouping is more meaningful. As we have high correlation between WOMAC and EQ-5D we then chose WOMAC for the grouping.

We see your argument that in multivariate regression the use of WOMAC as the grouping variable for costs or for QALYs has to be justified. Concerning costs, WOMAC and EQ-5D are not significant (see Table 3). Procedures are very similar among patients with different pre-op scores, therefore costs do not differ much in pre-op WOMAC or in pre-op EQ-5D groups.

This is completely different for QALY gains. There WOMAC as well as EQ-5D have of course high impact. Thus, we add multivariate regression for QALY gains with including WOMAC. Both WOMAC and EQ-5D have a similar impact on QALY gains. In the regression EQ-5D (-14.817) has at first view only higher impact than WOMAC (-0.151) because of the scaling of EQ-5D (0-1) and WOMAC (0-100). Beta is then much closer and one can see that both preoperative scores and age have the highest impact. Therefore we provide also QALY gains by age in Table 5.

CHANGES: The impact of WOMAC and EQ-5D was provided separately to avoid multi-collinearity. Both preoperative measures are correlated. For the grouping of QALY gains we
used the WOMAC as it is most accepted among orthopedics and mostly used in day-to-day routines. Both preoperative scores have similar impact on QALY gains but low impact on inpatient costs, as patients were treated very similar. Generally, there was low variance in inpatient THR costs but the number of operation and procedure codes, the number of secondary diagnoses, health insurance, and revision THR had an impact on costs (Table 1).

Bivariate analysis showed high correlation of preoperative WOMAC scores with QALYs gained. In a multivariate OLS analysis, the preoperative EQ-5D and WOMAC scores had the highest impact on the effectiveness of THR and therefore QALYs (Table 3). Thus, we provided QALY calculation for preoperative WOMAC groups. Among all preoperative characteristics, the preoperative WOMAC score suits best for payment by results approaches, modeling approaches, and management interventions to provide cost-effective THR. [page 9]

See the new Table 3 with regressions for both WOMAC and EQ-5D on QALYs.

Minor Essential Changes

The authors give a thorough explanation of the EQ-5D and a simpler explanation for WOMAC. The one thing lacking from both of these is interpretation. Specifically, the meaning of 1,1,1,1,1 in EQ-5D and 0 in WOMAC.

Response: We did not provide explanation for results of the questionnaires and included the following sentences. Thank you for this hint.

Changes: Thereby the best health state is 1,1,1,1,1 (no problems in each dimension) and the worst health state is 3,3,3,3,3 (severe problems in each dimension). [page 6]

To measure hip-specific outcome from a patients’ perspective, we use the WOMAC with its three subscales on pain (5 sub-questions), stiffness (2 sub-questions), mobility (17 sub-questions), and an overall score based on the three subscales. To make the overall score comparable to EQ-5D results, we transformed WOMAC results to a 0-100 scale where 0 is the poorest measure and 100 is the best measure. Each WOMAC question has a Likert scale from 0 to 10. [page 6]

Discretionary Changes

Having explained the direction of the health measures, see above, it would still be clearest to indicate the interpretation of results in the results section. So instead or in addition to making the point that QALY gains are higher when WOMAC is low, to say QALY gains are higher when patients are more healthy.

Response: Thank you for this hint. We see that the wording high/low is misleading and rephrased this and similar statements in relation to WOMAC, EQ-5D and QALYs. We now use the wording poor health state and acceptable/good health state. Indeed QALY gains are higher when patients are preoperatively less healthy.

Changes: All three value sets for QALY and QALY-like calculation showed a large increase in QALY gains with a decrease in preoperative WOMAC scores, indicating that patients with poor WOMAC scores benefit most from THR (Table 4). QALY gains are higher when patients are less healthy. [page 9]
Reviewer #2:

Many thanks for having the opportunity to review this manuscript. The authors estimate the impact of pre-operative patient conditions on the cost effectiveness of hip arthroplasty. The manuscript is well written and I recommend accepting it after some minor revisions.

Results section:
I’m missing some comparisons of the population in your analysis (292 patients) with other studies (e.g. EuroDRG work group: Geissler et al. 2012 http://onlinelibrary.wiley.com/doi/10.1002/hec.2848/pdf). It seems to me that your population is relatively less severe because of the low age and the small number of secondary diagnoses. This validation is important especially because of the small number of cases in your study. Maybe it is also worth to state some DRG statistics in order.

Response: To allow for a comparison with other studies, we use the suggested literature as it gives an overview across Europe and seems to be representative for many studies, especially for Germany.

CHANGES: Compared to the population in a large European Study that includes also a representative German sample [44], our population is slightly younger, has less secondary diagnoses (lower average Charlson Comorbidity Index), and patients have shorter length of stay and lower costs. 89% of patients were in the major DRGs for THR (I47A and I47B). [page 8]

Discussion section:
Third sentence# Isn’t that the case in Germany that revisions are paid and group differently than primary replacements?

Response: You are right, in Germany and in our case revision cases have usually the DRGs I46A and I46B. As this is no major finding in this paper we have removed this sentence as it was confusing. Thank you for this hint.

Limitations:
Please incorporate a comment on the selection bias. Data comes from a single but large university hospital. I wonder if you integrated all THR cases from January to June 2012? How different are these cases compared to Germany as a whole?

Response: We now included a better sample description in the section Methods/study population. Our data come from a university teaching hospital (BBM Clinic) and generally all cases were included. 393 patients were eligible of which 387 decided to participate (98%), 321 (82%) participated in the evaluation after 6 month. And 292 patients had full health outcome data so that we could include them in the QALY calculation. We have no data on the average German cases. Therefore we use your recommended literature to compare it to compare it to an approximated average German population.

CHANGES: The study was performed at an orthopedic hospital in Munich (Hospital Barmherzige Brüder München, university teaching hospital) and had approval from the ethics committee of Klinikum rechts der Isar, Technical University Munich. [page 4]

From January to June 2012, 393 patients were eligible to participate, 387 THR patients participated in the baseline health state evaluation and 321 (82%) patients participated in the follow-up. For 292 patients (74%) we received full HRQoL information including all
WOMAC and EQ-5D health measures and attributable, patient-based costing measures. [page 4]
Compared to the population in a large European Study that includes also a representative German sample [44], our population is slightly younger, has less secondary diagnoses (lower average Charlson Comorbidity Index), and patients have shorter length of stay and lower costs. [page 8]
As we have a monocentric study, representativeness of patients might not be given. [page 12]

Conclusions:
It is quite obvious and that the healthier patients the better the outcomes should be. But it might be more interesting to conclude more detailed how your measures can ideally be used in order to improve health care systems.

Response: We extended the section “In a future perspective...” just before the Conclusions section and gave more detail in the Conclusions section how to improve health care systems now.

CHANGES: Thus, WOMAC and health outcomes should be analyzed for patients who decided against THR compared with THR patients. A separation into preoperative WOMAC groups, as in this study, can then show which patients benefit most and for which patients cost-utility is highest with conservative therapy. A matching of pre- and postoperative EQ-5D values with population normative values can quantify the utility of total hip replacement for preoperative health state groups. Finally, a comparison of patient reported outcome measures and related micro-costing with other countries allows a detailed comparison of cost-effectiveness and possibly conclusions on weakness in the health care system. [page 12-13] CER results might be used in pay for performance approaches for an efficient and effective health care system [52]. Health outcome and micro-costing measures in combination are useful performance measures to compare hospitals on a patient-level basis. CER results support patients in a shared decision making situation before THR based on a personalized risk assessment approach. They support prioritization decisions in health policy concerning preoperative health states. And they support hospital management through differentiated costing and health outcome measures on a patient basis. [page 13]

Minor comments:
Abstract: Results#it should be written primary “replacement” and not revision
Response: Thank you for this hint. We changed this.

Page 3: second paragraph: Abbreviation HRQol should be explained (not just in the abstract)
Response: We explained the abbreviation now.

Tables: please consider to use capital letter in order to facilitate the readability
Response: Done.

Table 1:
• “side diagnosis” should be replaced with “secondary diagnosis” (also on other places e.g. page 7)
Response: Done.
• top 10 or top 5 diagnoses and procedures should be enough
  
  **Response:** We understand that the table might be a bit long. However, we wanted a detailed analysis of all possible impacts. We could reduce the table to a top 5 table or keep only the significant parts of correlation. But then one can’t see anymore which parts have no impact. Therefore we understand your argument. We also thought to put the table in the appendix, but it includes the complete descriptive part of our sample which is not in the text, so we decided finally to keep it.

• please consider the headline: Relevant pre-operative patient characteristics
  
  **Response:** Changed.

• Please note that these are ICD-10 diagnoses and German OPS procedure codes What means TVT? I can’t find it on the abbreviations list.
  
  **Response:** TVT was the German abbreviation and we changed it now. (deep venous thrombosis (DVT)). Thank you for this hint.

Table 2:
• Please think about to use the common matrix format of the InEK cost calculation scheme
  
  **Response:** We chose to use a simplified version. However, we understand that it is better to make the costing completely comparable to the German standard and we now use the common matrix.

  **CHANGES:** See the new Table 2