Author's response to reviews

Title: Different patient subgroup, different ranking? Which quality indicators do patients find important when choosing a hospital for hip- or knee arthroplasty?

Authors:

Nicolien C Zwijnenberg (n.zwijnenberg@nivel.nl)
Olga C Damman (o.damman@vumc.nl)
Peter Spreeuwenberg (p.spreeuwenberg@nivel.nl)
Michelle Hendriks (m.hendriks@nivel.nl)
Jany JDJM Rademakers (j.rademakers@nivel.nl)

Version: 2 Date: 21 June 2011

Author's response to reviews: see over
Dear editor,

Please find enclosed our resubmission of “Different patient subgroup, different ranking? Which quality indicators do patients find important when choosing a hospital for hip- or knee arthroplasty?”, by Nicolien Zwijnenberg, Olga Damman, Peter Spreeuwenberg, Michelle Hendriks, and Jany Rademakers. Note that we have changed the original title into a new title.

We thank you for considering our paper and for giving us the opportunity to submit a revised version of the manuscript. We were able to follow the reviewer’s suggestions and we think that the paper has improved significantly as a result.

We are looking forward to hearing from you.

On behalf of all authors,
Yours sincerely,

Nicolien Zwijnenberg, MSc

Correspondence concerning this article should be addressed to:
Nicolien Zwijnenberg
NIVEL
PO. Box 1568
3500 BN Utrecht
n.zwijnenberg@nivel.nl
+31 30 2729 856 (voice)
+31 30 2729 729 (fax)
Editor:

1) The editor states that i) experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee and ii) research carried out on humans must be in compliance with the Helsinki Declaration. If approval was not required, a statement regarding why this is the case should appear in the Methods section.

**Answer i and ii:** Approval by an ethics committee is not necessary under Dutch law, since the survey used in this study did not concern medical research involving a high burden for respondents. We have added a statement to the methods section.

"Ethical approval of the study was not necessary as research by means of surveys that are not taxing and or hazardous for patients (i.e. the once-only completion of a questionnaire containing questions that do not constitute a serious encroachment on the person completing it) is not subject to the Dutch Medical Research Involving Human Subjects Act (WMO). Subjects were free to respond to the questionnaire and they were informed about the aim of the survey."

2) The editor requested i) clarification on our response rate and ii) information on what response rate cut off was used.

**Answer:** i) The authors believe that the response rate has been clearly described in the manuscript. In the participants and data collection section the following passage has been described:

“In total, we distributed 265 questionnaires. Each of the three general hospitals received 50 questionnaires. Doctors or nurses handed out the questionnaire when patients were discharged from the orthopaedic departments. The researchers sent 115 questionnaires by mail to participants who enrolled themselves in this study through the different websites.”

In the result section the following is mentioned:

One hundred and ten THA/TKA patients (response rate of 41.5%) participated

ii) We did not use an explicit response rate cut off. We were aiming for at least 50 respondents and the inclusion period lasted for 3 months. This information has been added:

“Data were collected through a survey in the period from May 2009 till August 2009. In total, we distributed 265 questionnaires. Each of the three general hospitals received 50 questionnaires. Doctors or nurses handed out the questionnaire when patients were discharged from the orthopaedic departments. The researchers sent 115 questionnaires by mail to participants who enrolled themselves in this study through the different websites. All participants had to return the questionnaire by mail. We aimed for at least 50 respondents (expected response rate: 20-50%)”.
Reviewer 1

Major compulsory revisions

1) The reviewer requested (i) a table comparing the sample with the theoretical population, (ii) revisions in the manuscript by which we reframe the study and modifying some of the conclusions based on our sample and (iii) more information about possible sources of both systematic and random errors.

Answer:

i) Unfortunately, we do not have access to information about background characteristics of the total THA/TKA population in the Netherlands. However, as an alternative theoretical population, we used data from the 2009 nation-wide survey on patient experiences with the Consumer Quality Index Hip Knee Questionnaire among 8,675 THA/TKA patients (response rate 67%). In table 2 of the manuscript, the results of this measurement have been added.

In the result section, we now provide more information about how the demographic characteristics of our participants compare to the characteristics of this larger population (derived from the CQI survey):

"To get some idea of the representativeness of our study population, we compared the characteristics of our participants to the characteristics of a larger population of THA/TKA patients. These data were derived from the 2009 nation-wide measurement with the Consumer Quality Index Hip Knee Questionnaire (CQI Hip Knee) among 8,675 patients who underwent THA/TKA. The response rate in this measurement was 67% (N=5,163) [35]. Gender and perceived health status of both participants groups were comparable. However, participants in the current study were younger (M=64 years vs. M=68 years) and had a higher education level than participants of the measurement with the CQI Hip Knee-survey."

In the limitation section these differences are discussed. By providing this information, more insight into the representativeness of the study sample and its consequences has been provided.

"Second, there are some limitations related to the representativeness of the sample. The majority of the participants had enrolled themselves in this study. When background characteristics of our participants were compared to participants of a nation-wide study using the CQI Hip Knee Questionnaire, it appeared that participants of this study were younger and had a higher educational level than the participants of the CQI Hip Knee-survey. Additionally, given the self-selection, the participants of this study may be more interested in the topic of comparative healthcare information than the average THA/TKA population and this can slightly bias the results. Finally, the majority of the participants consisted of patients who already underwent surgery. Patients who still have to undergo surgery may evaluate other quality indicators as important, because the phase or severity of the disease can determine patients’ preferences [7]. It is, therefore, unclear to what extent the results can be generalized to a larger group of THA/TKA patients."

ii) We agree with the reviewer that the conclusions and implications of our study can be formulated more conservative in the manuscript. In fact, as mentioned by the reviewer, the purpose of our study was mainly to explore the preference patterns in different subgroups of patients. We made several revisions throughout the manuscript to put our results in the right perspective. The conclusion, for example, in the abstract is now as follows:

"This study provides a first insight into which quality indicators patients find important when choosing a hospital for THA/TKA, and how subgroups of patients differ in the value they attach to these indicators. More extended research is needed to establish the indicators that should at least be presented in succinct overviews of comparative healthcare information for patients choosing a hospital for THA/TKA."

In the introduction, we introduce our research questions now as follows:

"The aim of this study was to examine which quality indicators patients find most important when choosing a hospital for THA/TKA, in order to get some first suggestions for essential indicators to be presented as comparative healthcare information. This study is the first step in studying the preferences of this specific patient group."

3
iii) The reviewer suggests that our effects may have been capitalized on chance only, as a result of the small sample size and the multiple tests we have used. As also recommended by another reviewer, we have chosen to set the alpha at a more stringent level (that is to 0.01 instead of 0.05). In this way, we can be more confident that the remaining significant results are not the result of chance. In the strengths and limitation section we have expressed our concern that a small sample size can lead to possible sources of both systematic and random errors:

“Although this explorative study adds to our understanding of what patients find important when choosing a hospital for THA/TKA, it also has several important limitations. First, the number of participants is not very high, which will influence the statistical power negatively. Additionally, the small sample size combined with the multiple tests performed may lead to possible sources of both systematic and random errors. By using an $\alpha$ of 0.01, however, the chance that results are capitalizing on chance only is smaller and therefore the results are more reliable.”

2) The reviewer questions the development of dynamic web pages tailoring different patient groups based on background characteristics of patients. According to the reviewer, a system based on person characteristics could lead to relativism and inequities in the system. The reviewer requested to at least modify the text so that it expresses this concern and possibilities that such a system may lead to inequities.

**Answer:** The authors agree with this argument and, because we also had considered this once, we have decided not to support a webpage tailoring information based on background characteristics of the visitors. Only the option that patients themselves can select additional information on websites has now been described. Moreover, that this kind of system could lead to inequities in the system is now added as a major concern for web designs in the discussion section:

“Given these results, an interesting thought is to what extent the perceived relevance of indicators by (future) patients should drive the selection of quality indicators for comparative health care information. One could argue that crucial information, like the number of performed surgeries, should be available for every patient even if patients do not perceive it as relatively important. When such crucial information is only provided to patients who attach value to it, this may possibly result in inequities in the healthcare system. Ideally, comparative healthcare information consists of a mixture between indicators perceived as important by patients, and crucial indicators known to be related to treatment outcomes perceived as important by experts.”

**Minor essential revisions**

3) The reviewer stated that the sentence “First, patients….study” was awkward to read.

**Answer:** We have rewritten this sentence:

“First, patients were approached at orthopaedic departments of three general hospitals. These hospitals were selected via personal networks”.

4) It was not clear to the reviewer that we claim that “an important strength of this study is that participants ranked different quality indicators instead of rating the importance of these indicators”, especially because we also identify the limitations of this methodology. Furthermore, the reviewer mentioned that we do not address the fact that we can not compare across the three dimensions.

**Answer:** The authors are of the opinion that ranking quality of care indicators is a better method than rating quality indicators. The arguments for this statement are explained in the text:

“An important feature of this study is that participants ranked different quality indicators instead of rating the importance of these indicators. By ranking indicators, patients must explicitly weigh what they find more and less important. By rating indicators, in contrast, patients can rate every indicator as equally important. So, by using the ranking method, the results provide more insight into what patients find most important.”

We also do acknowledge that the ranking method has some limitations in general and therefore these should be described as well. However, the original position of the description of the limitations in the manuscript has been misplaced and can lead to confusion to the readers. Therefore we have removed this part to the limitations of the study.
Regarding the inability to compare the three different quality of care indicators, we agree with the reviewer that his should be added as a limitation:

"by using separate ranking assignments for the three types of quality indicators, it was not possible to compare across the three dimensions of the quality indicators. The use of different ranking assignments, however, was a deliberate choice to decrease the cognitive burden for participants to fulfill the assignments."

5) The reviewer requested to delete the sentence in paragraph 2 in the section “discussion of the results”.
Answer: the sentence has been deleted.

Discretionary revisions

6) The reviewer requested more information about the properties of the scale (Cronbach’s alpha, mean and SD) developed by Groenewoud.
Answer: The psychometric properties of the scale have been tested again in our study and these results are now provided in the manuscript:
“The scale showed high reliability (α=.77) and the mean score was 18.5 (95% CI 17.9-19.2), suggesting a relatively active search and selection behaviour of the participants.”

7) The reviewer requested an actual example from the material in paragraph 1 in the section “data analyses”.
Answer: We now provide another example in the text:
“For example, when a participant had to rank four indicators (e.g. the patient experience indicators), the indicator ranked most important received four points, the second indicator received three points, the third indicator received two points and the indicator ranked least important received one point.”
Reviewer 2:

Minor essential revisions

1) The reviewer would like to see i) more information about the representativeness of the sample and ii) would like to see more elaboration on the third limitation (the fact that we did not use a random sample).

Answer: i) The response rate of the patients who have enrolled themselves in this study has been added:

“One hundred and ten THA/TKA patients (response rate of 41.5%) participated (see Table 2). Of the participants, 83.6% have enrolled themselves in this study by reacting on calls on websites.”

As the majority of the participants are patients who have enrolled themselves in this study, we found a comparison between this group and the patients who are recruited at the hospital departments concerning their background characteristics not fruitful. It does not provide valuable information to the reader (because of a too small number of participants). Instead, a comparison between our study sample and a larger sample of THA/TKA patients (N=5.163) has been provided now in Table 2 and in the result section similarities and differences in background characteristics have been described:

“To get some idea of the representativeness of our study population, we compared the characteristics of our participants to the characteristics of a larger population of THA/TKA patients. These data were derived from the 2009 nation-wide measurement with the Consumer Quality Index Hip Knee Questionnaire (CQI Hip Knee) among 8,675 patients who underwent THA/TKA. The response rate in this measurement was 67% (N=5,163) [35]. Gender and perceived health status of both participants groups were comparable. However, participants in the current study were younger (M=64 years vs. M=68 years) and had a higher education level than participants of the measurement with the CQI Hip Knee-survey.”

ii) The limitation section of the manuscript has been extended:

“Although this explorative study adds to our understanding of what patients find important when choosing a hospital for THA/TKA, it also has several important limitations. First, the number of participants is not very high, which will influence the statistical power negatively. Additionally, the small sample size combined with the multiple tests performed may lead to possible sources of both systematic and random errors. By using an α of 0.01, however, the chance that results are capitalizing on chance only is smaller and therefore the results are more reliable. Second, there are some limitations related to the representativeness of the sample. The majority of the participants had enrolled themselves in this study. When background characteristics of our participants were compared to participants of a nation-wide study using the CQI Hip Knee Questionnaire, it appeared that participants of this study were younger and had a higher educational level than the participants of the CQI Hip Knee-survey. Additionally, given the self-selection, the participants of this study may be more interested in the topic of comparative healthcare information than the average THA/TKA population and this can slightly bias the results. Finally, the majority of the participants consisted of patients who already underwent surgery. Patients who still have to undergo surgery may evaluate other quality indicators as important, because the phase or severity of the disease can determine patients’ preferences [7]. It is, therefore, unclear to what extent the results can be generalized to a larger group of THA/TKA patients.

Besides the usual limitations inherent to an explorative study, there are also some other methodological issues. The use of a ranking method forces people to make choices and this creates a hypothetical situation, since people can decide in real life to make no choice at all. In addition, by using separate ranking assignments for the three types of quality indicators, it was not possible to compare across the three dimensions of the quality indicators. The use of different ranking assignments, however, was a deliberate choice to decrease the cognitive burden for participants to fulfil the assignments. Finally, this study was limited to three types of quality indicators. Other indicators also can be important for patients when choosing a hospital, as was illustrated by the answers of participants to our open ended question. Taking...
these limitations into account, some caution is warranted by the interpretation and generalization of the results of this study.”

2) The reviewer requested for harmonization of the table numbers and place of tables in text. **Answer:** The numbers of the tables into in the additional files have been adjusted into characters; the table character is corresponding with the additional file character (e.g. Additional file A: table A). The authors believe that the place of the tables in the text is correct and are not in need for any adjustment.
Reviewer 3:

Major compulsory revisions

6a) The reviewer requested further consideration of the methodological limitations of this study in the discussion, more specifically i) the usual limitations for exploratory analysis are acknowledged in terms of potential for spurious relationships, ii) some statement urging caution in interpretation and generalisation of the results is made and iii) the requirement for future replication/extended study is specified.

Answer: i) The chance of potential spurious relationships will be diminished now by using an alpha level of 0.01. In the strengths and limitations section, however, we have acknowledged the methodological limitations of explorative research. The following sentence has been added:

“Besides the usual limitations inherent to an explorative study (e.g. potential for spurious relationships), there are some other methodological issues.”

ii) In the strengths and limitations section the following sentence has been added:

“Taking these limitations into account, some caution is warranted by the interpretation and generalizations of the results of this study.”

iii) In the first section of the discussion the following part has been added:

“However, due to the explorative nature of this study, definite conclusions as to which indicators to present cannot be drawn. More extended research based on the current results is needed to select the indicators to be presented. Subsequent studies should preferably use experimental methods and include a larger group of patients to further investigate the perceived importance of the various quality indicators.”

Minor essential revisions

1) The reviewer would like to see that words like ‘important’ and ‘relevant’ are replaced by “perceived importance”/“perceived relevance”. Also in the title and research questions the reviewer would like to see these terms, because one interpretation of ‘importance’ or ‘relevance’ is validity of the metric as a quality indicator for care at a specific centre, while the focus of this study is on the perceptions of salience or priority by a patients.

Answer: The authors are of the opinion that it has to be clear to the readers that in this study the patient perspective has a central role, that is, what do patients themselves perceive as important quality indicators. Therefore, we have made several changes. For one, the title of the manuscript has been changed:

“Different patient subgroup, different ranking? Which quality indicators do patients find important when choosing a hospital for hip- or knee arthroplasty?”

As for the research questions, the authors believe that the questions could be correctly interpreted. Therefore no adjustments have been made.

The first sentence of the procedures and measures section has been adjusted:

“The questionnaire contained three assignments in which participants had to rank the quality indicators from most important to least important when choosing a hospital for a total hip- or knee replacement.”

Also in the discussion section the first sentence has been changed:

“This study explored which quality indicators patients find most important when choosing a hospital for THA/TKA and whether patient subgroups differed in which indicators they find important”

Although we understand the concern of the reviewer, by reformulating the title and the first sentence of the discussion section and procedures and measures section, the authors believe that the study can be correctly interpreted and that no further adjustments in the text are necessary.

2b). The reviewer requested some adjustment to the Alpha level, because multiple tests of significance may increase the family-wise error rate

Answer: The authors agree with this argument. The chance that these results are capitalizing on chance only is smaller and therefore the results are more reliable. The Alpha level has been set to 0.01 instead of 0.05:

“A reliability level of 99% (p<0.01) was used for all tests.”
As a consequence the results are changed and less significant differences are found between patient subgroups. The result section has been rewritten.

2c) The reviewer would like to see further details about the search and selection behaviour scale. **Answer:** we have provided a table in which the six items of the scale are presented in Table 1. Additionally, the alpha of this scale has now been provided also in the method section.

**Table 1- Items of the search and selection behaviour scale**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>It doesn't matter too much to me where and by whom I am treated.</td>
<td></td>
</tr>
<tr>
<td>I don't want to invest too much time and energy in the choice process.</td>
<td></td>
</tr>
<tr>
<td>If I need care, I usually go the therapist/care facility to which my GP or specialist has referred me.</td>
<td></td>
</tr>
<tr>
<td>If I need care, I usually investigate thoroughly how, where and from whom I will receive the best treatment.</td>
<td></td>
</tr>
<tr>
<td>I have experience with the health care system and therefore know which therapist or care facility is best for me.</td>
<td></td>
</tr>
<tr>
<td>I think it's important to weigh possible treatments, therapists and care facilities against each other properly.</td>
<td></td>
</tr>
</tbody>
</table>

5) The reviewer suggests that it would be useful to acknowledge that there is an extent to which the “one size fits all” approach should be challenged and the degree to which subjective perception of relative importance of different facets of health care quality by different patient sub-groups should drive selection of indicators for inclusion. The reviewer mentioned that we should be cautious in concluding that significant inter-group variation in preferences should lead to unconstrained tailoring of available categories of information. **Answer:** The authors agree that inter-group variation in preferences should not lead to unconstrained tailoring of information. However, we do also believe that, by integrating indicators which are perceived as important by patients, the information becomes more relevant for them and, consequently, could stimulate the use of the information. In the ideal situation, a mixture should be created between indicators perceived by patients as important/relevant and more crucial indicators known to be related to treatment outcomes. In the manuscript the following passage has been added in the discussion section:

“As given these results, an interesting thought is to what extent the perceived relevance of indicators by (future) patients should drive the selection of quality indicators for comparative health care information. One could argue that crucial information, like the number of performed surgeries, should be available for every patient even if patients do not perceive it as relatively important. When such crucial information is only provided to patients who attach value to it, this may possibly result in inequities in the healthcare system. Ideally, comparative healthcare information consists of a mixture between indicators perceived as important by patients, and crucial indicators known to be related to treatment outcomes perceived as important by experts.”

**Discretionary Revisions**

2a). The reviewer states that we report only a minimum of information about the parameters of the models fitted, effect sizes an details of the hierarchical linear model (e.g. intra-class variance). **Answer:** In this research no traditional linear regression models were used. Instead, we used random effects regression analyses. As a result of this, no intra-class variances were calculated and could therefore not be provided. We have rewritten the last part of the data analyses section, to make more explicit how the ranking data have been analysed:

“To assess the differences in rankings between patient subgroups, we analysed the mean scores and the between-participants variances for every patient subgroup, using random effects regression analysis [33]. This type of linear regression allowed for the simultaneous estimation of the mean score and a between-participants variance for every indicator in the assignment. Traditional regression analysis would be inadequate, because it would assume that for every indicator the variance was equal. The difference in mean score was tested by adding a subgroup-variable (e.g. gender) to every indicator in the assignment (interaction term). The regression coefficient showed the difference in ranking between the two patient
subgroups (e.g. men vs. women). For the between-participants variance, for every subgroup a separate variance was estimated and these were compared using a contrast (Wald) test [33,34]. A reliability level of 99% (p<0.01) was used for all tests.

Concerning the parameters of the models fitted; we have conducted 204 tests to test differences in mean scores or variances. The authors believe it would be beyond the scope of this manuscript to provide more information about the parameters used for all 204 models. We have chosen to only provide the most relevant results of the tests (mean scores, variances and significant results) in the tables in order to keep the results understandable and clear.

4a) The reviewer would like to see more details concerning the parameters of the fitted statistical model.
Answer: see our answer to 2a.

4b) The reviewer suggested to discuss the strengths and limitations of the study after the discussion of the results.
Answer: We have removed the strengths and limitations section and now discuss the strengths and limitation after a discussion of the findings.

6b) The reviewer suggests that the description of the indicators offered to the respondents would be more clear if further indication was provided to respondents of why each indicator is relevant.
Answer: The description offered to the respondents is a global, neutral description of the meaning of every indicator. In the light of this study, patients themselves decide if they perceive an indicator as important/relevant or not. For this reason, information concerning the relevance of an indicator from the perspective of experts was not provided.

6c) The reviewer requested some further suggestions for productive future work.
Answer: We thank the reviewer for his valuable suggestions, which correspond to our own thoughts about future work. We have explicitly included several of his suggestions in the implication section. In addition, suggestions for extended research based on the current study are now specified (see our answer to 6a, point iii).

“For future research we would recommend to examine which quality indicators are important for other patient groups when choosing a healthcare provider, in order to develop relevant comparative healthcare information for them as well. In our opinion, using different methods to elicit patient preferences is preferable. The use of the ranking method to explore patients’ preferences is a good starting point, complemented with other research (e.g. discrete choice experiment or qualitative research) to profoundly explore patients’ preferences. Although the central focus of this study concerned which aspects have to be presented for consumers, we acknowledge that the presentation approach is just as important [8,16] for making comparative healthcare information successful. So, for future research and the development of comparative healthcare information, focusing on the presentation of comprehensible information and the visual display of information is just as important as presenting information that is relevant for consumers”

9) The type error has been adjusted.