Author's response to reviews


Authors:

Lars P Hölzel (lars.hoelzel@uniklinik-freiburg.de)
Levente Kriston (l.kriston@uke.de)
Christina Reese (christina.reese@uniklinik-freiburg.de)
Martin Härter (m.haerter@uke.de)

Version: 2 Date: 19 February 2013

Author's response to reviews: see over
Reviewer’s report

Title: How does Shared Decision-Making influence patient satisfaction? -
Empirical testing of a heuristic model

Version: 1 Date: 11 October 2012
Reviewer: Randy Wexler

Reviewer’s report:

Major Compulsory Revisions: None
Minor Essential Revisions: None

• Thank you for reviewing our manuscript.

Discretionary Revisions:

Page 4. I would simply state that primary care is often the source of first contact
as some physicians find the term “gatekeeper” to be pejorative.

• The text was corrected. It now reads:
  “Patient involvement plays a central role in primary care because primary care is
  often the source of first contact, and general practitioners have to collaborate with
  patients to identify their healthcare needs and choose corresponding services”

Page 6. Please describe how the patients taking part in the healthcare project
were so selected. Were they recruited? If so how? Were they assigned? If so
how?

• The text now reads:
  “All insurants of the Health Insurance Fund AOK Baden-Württemberg and the
  Health Insurance Fund LKK Baden-Württemberg residing in the area of the
  Kinzigtal were suitable for participation in the study as members of the
  intervention or the first control group. The intervention group consisted of all
  insurants that were taking part in the “Kinzigtal” project at the reference date of
  31.03.2007. The first control group consisted of insurants living in the same area
  but not taking part in the project. The second control group consisted of insurants
living in an area that is comparable to the Kinzigtal with respect to healthcare infrastructure, population density, and local economy. To ensure that the control groups are comparable to the intervention group regarding central characteristics, insurants in these groups were selected by stratification of sex, age, kind of insurance, and healthcare costs they induced in the previous 12 months. All selected insurants were invited to participate in the study by a written invitation sent by their Health Insurance Fund (a more detailed description available in (1)).”

Page 7. Although general information on the postal questionnaire is included, more detail on the contents would be helpful perhaps in table or chart form.

• We added an additional table (Table 2) of the contents of the postal questionnaire.

Page 7. Why was the last clinical appointment chosen as the reference point as opposed to care over the previous 6 months or some other broadened time frame? A single visit, especially the last one can impact patient response.

• Our intention was to collect valid data (i.e., that reflects usual care) by not allowing participants to choose an arbitrary reference consultation. By doing so we hoped to reduce idealization and social desirability. A clear disadvantage of our approach is that not every consultation is equally appropriate to judge patient involvement. However, as we did not want to make judgments on single cases but were exclusively interested in group differences, the appropriateness of single consultations seemed less important to us as long as they remain comparable between groups. As the appropriateness of the consultations should be independent of the assignment of the patients, the appropriateness of the consultations should have no systematic bias on our results.
Reviewer's report


Version: 1 Date: 29 October 2012
Reviewer: Oliver Groene

Reviewer's report:
This is a well written manuscript addressing the link between preferences for involvement, shared decision making (SDM), decision conflict and satisfaction. The focus to link SDM to intermediate and long-term endpoints is original and a relevant contribution to the literature.

The study proposes a heuristic model and then appropriately uses a structural equation model to test it. Positive to note is the split sample approach (using a development and validation sample).

The paper has some other limitations, but they are all transparently discussed in the discussion section and mostly quite typical for this type of research. Therefore, the limitations put forward by the authors (e.g. recall bias, temporal relationships) should not prevent this manuscript from being published.

• Thank you for reviewing our manuscript.

Major Compulsory Revision
My major concern is that the study uses a pooled set of data from “Gesundes Kinzigtal” which might attenuate the results, given intervention and control groups were pooled while “increasing patient involvement in clinical decisions” was a special focus of the overall project. I do not think that the use of a pooled dataset
• *The main advantage of using the pooled set of data is that the pooled data provide a sufficiently large sample to use one half for the development of the model and the other half for testing it. However, we absolutely agree with the concern with regard to using pooled data from the intervention and the control groups. Group differences, if not accounted for, may seriously affect the findings. However, our analyses were based exclusively on baseline data that were collected before the intervention, and participants for the control group were selected with special care regarding the comparability of the groups (stratified sampling). Therefore, it seems unlikely that the groups differed with regard to patient involvement at this time point. However, as we did not use a randomized design there may still be some group differences we are not aware of (e.g., in unobserved variables) and accordingly have not controlled for (e.g. relationship with physician, health literacy etc.). As we cannot rule out such group differences with certainty, we decided to conduct a sensitivity analysis by fitting the statistical model to each of the three groups separately and investigate group differences with regard to the standardized regression weights. Regression weight differences among these group-specific models were all below .20. Standardized regression weight differences above .15 were exclusively found for confounders. Accordingly, our analysis revealed no meaningful difference between the intervention and the control groups in any of the standardized regression weights. We added two tables for a web appendix and a short comment to the text:*

**METHODS**

*Data Analysis:* “An additional sensitivity analysis was conducted to test for possible model differences among the investigated groups. In this analysis, we fitted the statistical model to each of the three groups separately and investigated group differences with regard to the standardized regression weights. Thus, the consistency of our results (that were based on the pooled sample of the three groups) was tested additionally in the three original subgroups.”

**RESULTS**
Model development and testing:
“The results of our sensitivity analyses are displayed in table 6 and 7. Group differences with regard to the standardized regression weights were all below 0.20.”

DISCUSSION
“The sensitivity analysis revealed no meaningful difference regarding the model parameters of central interest (standardized regression weights) among the groups of which data were pooled for the primary analysis.“

Other issues that I would like to ask the authors to comment on:
- Table 2: how does the sample differ from the population it was drawn from
  (potential selection bias)

  • We found significant deviations between responder and original population with regard to group status, age, costs and kind of insurance. However theses variables could explain only 2.8% of the variance in the response behavior (responder vs. non-responder) [19]. The low explained variance suggests a largely representative sample.

- Table 3: goodness-of-fit appropriately reported, it would be useful though to report on the descriptive of the constructs in order to allow for comparison to other studies that used the same measure.

  • Thank you for this idea. We agree that the descriptives of the constructs may be of interest for comparison with other studies. However, as we used latent variable modeling (with restricting the mean of latent constructs to be zero), we cannot give this information on the basis of the results in this article. Thus, we decided to add a reference were the scale descriptives can be found:
    “Further information on descriptive results is given elsewhere [19].”

- All sources of data are derived from the same questionnaire: might this lead to
common methods bias

- Thank you for this remark. We added this limitation to our discussion: “Furthermore, all sources of data are derived from the same questionnaire, what might lead to common methods bias.”

- Inconsistent results regarding the association between patients’ preference for involvement and current involvement. In line with my major criticism above, this should be re-assessed taking separately into consideration the data from intervention and control group.

- We conducted a sensitivity analysis. The results were robust and we can rule out meaningful differences of the associations among different subsamples (see above).

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests: I declare that I have no competing interests
Reviewer’s report
Title: How does Shared Decision-Making influence patient satisfaction? -
Empirical testing of a heuristic model
Version: 1 Date: 4 November 2012
Reviewer: Margaret Holmes-Rovner
Reviewer’s report:
This paper investigates a well-studied question, “Does SDM influence patient satisfaction with care received from a physician?” The authors have a new data set, which confirms earlier findings. They appear to elevate these findings to the testing of a model. However, the approach and results are unconvincing. While hypotheses are stated about expected correlations, it is not so clear what the question is. Is the question how do we predict patient satisfaction? Is the question about the effectiveness of shared decision making? If that is the case, do we have evidence that shared decision making occurred? Page 9 suggests that the SDM-Q-9 measures patient involvement in the decision making process from the patient’s perspective, attributed to the last remembered encounter. Is the study then fundamentally about patient’s perception of involvement?

Thank you for reviewing our paper and providing essential and helpful comments. We agree that the question “Does SDM influence patient satisfaction with care received from a physician?” is already well-studied. We also agree that another investigation of the correlation between SDM and patient satisfaction would be not very innovative. However, the aim of our study was to integrate existing knowledge (pairwise associations between constructs) into a larger picture. We think that investigations of SDM address several different constructs but a comprehensive model of how these constructs are related to each other is largely missing. We consider a model of the associations of SDM related constructs important to gain a better understanding of how SDM works and may be helpful, e.g. for choosing outcomes for the evaluation of SDM interventions. We think that such a model should incorporate theory and this theory should be tested empirically. Accordingly, we developed a model of SDM related constructs that is in accordance with theoretical assumptions in the literature. This theory
was empirically tested using structural equation modeling (SEM). Aim of this method is to test theoretical assumptions (i.e., multiple associations between constructs) empirically. The theoretical assumptions (or theories) that are tested with SEM are usually termed “models”. We understand that the text can be read as if it would “elevate these findings to the testing of a model”. However, “model” is a common term in SEM. You are right that we did not state clear questions. The primary question of our study was the global testing of the whole theoretical model. Additional questions, whether the empirical data support our hypotheses, are stated in the text. Title and text were changed accordingly. We hope the aim of our study is now clear.

The statistical methods deserve a review by a statistician. In addition, the rationale for the particular method is not included. The survey instruments are previously developed. However, their use in this model building from cross-sectional survey data should be considered from a theoretical as well as correlational view.

- The statistical methods have been reviewed by another reviewer (Oliver Groene). We chose not to explain the rationale of the method we used in depth. We agree that SEM is not a typical method for health services research and thus some readers may be unfamiliar with it. However, the explanation of the method would go far beyond the scope of the article. We added a reference with a comprehensive but still fairly brief explanation of the method used. Several SEM-textbooks exist, and it is increasingly taught in graduate courses of quantitative research methods and statistics in health, behavioral, and social sciences.

You comment that the use of the survey instruments “in this model building from cross-sectional survey data should be considered from a theoretical as well as correlational view”. We tried to include a “theoretical view” by developing our heuristic model on basis of the current literature. The analysis method (SEM) relies basically on associations between variables, thus, it is in perfect accordance with the proposed “correlational view”. However, we do not think that
pairwise correlations are suitable to test the model as a whole. We think that regression analysis of latent variables (SEM) is more convincing in this case.

The results consistently over-state the positive results of their study. The abstract presents only positive findings and (unstated) “counter-intuitive” findings.

Thank you for highlighting this important issue. We changed the abstract and refocused it on the primary question of the article.

The authors acknowledge the cross-sectional nature of the data and suggest that future work should be longitudinal. Should the present study be considered exploratory? What modifications of the approach would the authors suggest for future studies? Is the “confirmed” model ready to be tested longitudinally or in an RCT?

- Although longitudinal data would enhance the value of our analyses, we think the label “exploratory” would be misleading. In our analyses we tested a theoretical based model with empirical data. In our point of view exploratory would imply a model development without predefined theoretical assumptions. We would suggest modifying the instrument measuring “patient preferences” (API). We think that it may be helpful to test whether the counter-intuitive findings are specific for the API. Furthermore we think that other SDM-related constructs could be included in the model.

We think that the model is ready to be tested longitudinally to strengthen causal conclusions. It may also be useful for developing and evaluating interventions. The model provides insights on how possible outcomes of a randomized controlled trial may be associated, so choosing between them for evaluating a certain intervention can now be made on a more informed basis. The findings can also be useful for developing interventions aiming at different points of the proposed model. However, we do not think that the model itself should be tested in an RCT in any ways. For this a longitudinal design may be more appropriate.
- Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. Please correct the grammatical use of “between”. It means between two things. When used to describe relationships (plural) AMONG variables, between is not grammatically correct. This occurs repeatedly.

   - We corrected the use of “between” throughout our text.

- Major Compulsory Revisions

1. This report requires a more balanced statement of results, both in the text and in the abstract. The authors show known relationships between SDM and satisfaction. Their other assumptions were largely not confirmed, and yet the text repeated says the model is a good fit and confirms their hypotheses.

   - We tried to state our results in a more balanced way (see above). However, the fit of a model refers to a statistic method. It simply informs on how well a model describes the data. It does not provide information on the fit between a hypothesis regarding the strength of association between two variables and the empirically derived association. Using the term “fit” is well-established in the SEM-methodology. We think that the results of our analyses should be clearly stated in the text and in the abstract. As our model generally showed an acceptable fit to the empirical data and as most of our hypotheses could be confirmed we think that our model was largely confirmed.

2. There is an absence of theoretical grounding for what is set up as an important model building exercise. Without theory, the empirical work becomes data-dredging. Why should these constructs be related in these ways?

   - We tried to improve the description of our theoretical assumptions in the background of our article.

Do the survey measures each simply re-state the same ideas? Is that why the
correlations are good in some areas?

- We do not think that all survey measures re-state the same ideas. We used established and psychometrically tested instruments that were developed to measure different constructs. The content of the items varies broadly among the scales used. We think that the preference for involvement, perceived involvement, decisional conflict and satisfaction are distinguishable constructs. The results of our analyses support this view as the interrelations between the constructs are not strong enough to indicate redundancy. Of course, we agree that the constructs themselves are possibly somewhat overlapping. It is a very interesting research question currently being intensively investigated. We hope that our study adds a valuable piece of information to this literature.

What is modifiable and what is not? What conclusion should we draw if this model is not adequate? How does this model suggest the field move ahead?

- Every part of the model has to be confirmed empirically. Otherwise the model has to be adapted. However the fit of our model was acceptable. Thus we do not think that the whole model has to be rejected. We think that the model could help to understand how shared decision making works. Other constructs of SDM could be incorporated in future investigations. Based on the model, nomological networks for the validation of psychometric instruments could be developed and tested. The model could also help to choose adequate outcome parameter for studies of shared decision making (e.g. a study focusing on patient preferences would need other instruments than a study with focus on the process; see also comments above). The text was adapted accordingly.

3. The assumptions behind the model and the analysis are not clearly described. Why structural equation modeling? The authors indicate that with cross-sectional data, they cannot draw causal inferences. But then, Table 4 makes causal arguments, as one might expect from an SEM analysis.
• We chose SEM as this method was developed for the empirical testing of complex theoretical models. We absolutely agree that we cannot draw causal inferences from our data and stated this position clearly in the text. We are aware that the models investigated with SEM analysis sometimes look like they would be causal. In our opinion, causality is more likely to be established by design rather than statistical analysis. To avoid misunderstandings we revised our text accordingly.

Why is a stepwise regression approach used here?

• The stepwise approach was used as we wanted to include some variables as confounders that are not part of the theoretical model of interest. As we did not want to test any hypotheses about these confounders we adapted our model to fit the data with regard to the relationship between the confounders and the model. The text was revised accordingly:

“A path analysis using structural equation modelling [28, 29] was employed to explore multiple associations. Data were analysed with AMOS 5 (SPSS Inc., Chicago, Illinois). The elements of the heuristic model were included in the model as defined a priori based on our theoretical assumptions. To control for possible confounding effects, the influence of the following known measures on the central constructs was also modelled: demographic characteristics (age, sex, education), clinical characteristics (cardiovascular, musculoskeletal, or endocrinological disease), quality of life (mental and physical), and type of the clinical decision to be made (diagnostics, therapy, or referral). As these elements are not part of the heuristic model, their effects were freely estimated to fit the empirical data of these elements. A satisfactory model was developed in a development subsample and cross-validated in a confirmatory sample (split-half method).”

4. The relationship of the Ende API measure to the other constructs was disconfirmed. The authors acknowledge this, but then claim their model is confirmed. They should be more forthcoming about the fact that the API measure
fails to perform as expected in their model. There is a passing statement that they are not interested in validating new measures. However, that begs the question. Is the API an in-valid, though reliable measure of patient preference for involvement? Is the model conceptually flawed or is the measure flawed? The article is unclear about the authors' thinking on this result. There is sound theoretical work on patient preference in the psychological literature showing that preferences are not elicited, but rather are formed in the process of decision making. (See B. Fischhoff. A thorough discussion can be found in the IPDAS background Chapter on preferences at www.ipdas.ohri.ca.) If this is true, what does it suggest for the model tested here? If the authors think the earlier studies are wrong, what is the evidence? Is Ende measuring something other than preference? An alternative explanation for the findings is that patient preferences are formed in the decision making process, and that the Ende measure is perhaps reliable, but not valid as a preference measure. Whatever, the approach, it deserves theoretical consideration of the relationships, as well as empirical correlating of scales.

- **We agree that the results on patient preferences for involvement are inconsistent with our model. However the model includes more variables. The other hypotheses stated could be confirmed. That does not mean that we want to ignore the negative result on patient preferences. However we think that not the whole model has to be rejected.**

We absolutely agree that our works beg the question on the validity of the API. We feel that we are unable to answer that question on the basis of the present study. To answer it, it would be necessary to reinvestigate the model with different operationalizations of “preference for involvement”. By that we could investigate whether the inconsistent finding is a problem of the API or of the model. We agree that preferences are formed by the process of decision making. We included this idea into the revised manuscript. However we think that in addition to a preference for involvement in the specific medical encounter there is also a generic preference for involvement in decision-making. We think that this generic preference for involvement is relatively stable. We revised the text accordingly:
“Another explanation could be that preference for involvement is highly subjective depending on the context and circumstances [40]. Thus, involvement preference in the specific medical encounter might be very different from the generic preference for involvement in decision-making.”

Level of interest: An article of limited interest

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests
In addition to addressing the referees's comments we require the following editorial points be addressed:

1. Ethical Approval - Research involving human subjects (including human material or human data) that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the Helsinki Declaration (http://www.wma.net/en/30publications/10policies/b3/index.html). A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate.

- The reference number and statement were added.