Author's response to reviews

Title: Development and testing of culturally sensitive patient information material for Turkish, Polish, Russian and Italian migrants with depression or chronic low back pain in a double-blind randomised-controlled trial (KULTINFO): A study protocol

Authors:

Lars P Hölzel (lars.hoelzel@uniklinik-freiburg.de)
Zivile Ries (zivile.ries@uniklinik-freiburg.de)
Jördis M Zill (j.zill@uke.de)
Levente Kriston (l.kriston@uke.de)
Jörg Dirmaier (j.dirmaier@uke.de)
Martin Härter (m.haerter@uke.de)
Isaac Bermejo (i.bermejo@celenus-kliniken.de)

Version: 2
Date: 2 May 2014

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

Reviewer: Maura Marcucci

Reviewer's report:

Comments to the authors

Hölzel and colleagues present the protocol of a randomized controlled trial on the usefulness of culture-sensitive health information material compared to standard translated information on depression or chronic low back pain for patients with a migration background living in Germany. The topic is surely interesting, and the background information describes a condition of equipoise and supports the rationale of the study. The protocol is well written and easily readable. The research question is quite well translated into appropriate methods. However, some relevant details to judge the internal and external validity of the study appear under-developed/reported throughout the protocol. See comments below.

Thank you for carefully reviewing our manuscript.

Major Compulsory Revisions

#1 Setting. The information on the study setting (primary care) is missing in the abstract and in the formulation of the study question. In general, the importance of the setting is overlooked throughout the protocol, whereas it is fundamental for the interpretation and the generalizability of the future study results. The same interventions could be tested in a different healthcare setting, for example in specialist clinics or in health services specialized in the assistance of migrants. The findings of the current study will not be necessarily interchangeable with the findings of these hypothetical studies conducted in different settings, with different patient populations. The rationale of the choice of the primary care as setting for the study should be added to the “METHODS/DESIGN” (perhaps after the last paragraph on page 4).

We added the study setting to the ABSTRACT, METHODS/DESIGN and DISCUSSION.

#2 General practitioner recruitment. How the general practitioners participating to the study will be recruited by the study centers should be specified. What is the
existing relationship between the study centers and the primary cares?

We added an additional paragraph in which we described the recruitment of general practitioners and the relationship between study center and the primary cares:

“Recruitment of general practitioners
General practitioners will be recruited by inviting academic training practices, practices of the cooperating networks (see above) and general practitioners (using public registers) to participate in the study. General practitioners received an allowance €40.00 per patient recruited.”

#3 Patient inclusion criteria and recruitment. The protocol should state more clearly if it is the participating general practitioner to assess patient eligibility and which data source for patient socio-demographic and clinical information he/she will rely on. It should also specify which sampling method will be used to enroll patients, whether a consecutive approach can be accomplished. In addition it is not clear if any patient with a migration background affected by depression/chronic low back pain will be considered for inclusion regardless of the reason for the consultation, or if the reason for the index consultation needs to be related to one of the two targeted diseases. With regards to the inclusion criteria as currently specified, the investigators are invited to reflect on the risk of including typologies of patients whose response might lead to a dilution of the effects, due to: 1) no specification of the timing in the disease course when the patients are eligible (with the possible inclusion of patients with already a long experience with their disease); 2) the adoption as criteria for migration background of “being born in another country” or “having a parent who was born in a country other than Germany” (with the possible inclusion of people grown up in Germany, completely integrated in the German culture, even if able to read their parents’ tongue).

We adapted the text accordingly:

“Potential limitations of the study are that the inclusion criteria (especially the
diagnosis) are exclusively based on the clinical judgment of the general practitioner. For feasibility reasons we chose to include patients regardless of the reason for consultation and we did not specify the timing in the disease course when the patients are eligible. This may dilute the effects of the intervention.”

“Inclusion criteria

General practitioners will consecutively assess patients for eligibility. Adult primary care patients with a Turkish, Polish, Russian or Italian migration background and with unipolar depressive disorder or non-specific chronic low back pain will be included. Having a migration background is defined as being of non-German nationality, having a first language other than German, being born in another country, or having a parent who was born in a country other than Germany. To be included, patients have to report that they perceive their non-German origin as part of their identity. Unipolar depressive disorder is defined according to ICD-10 (F32.xx, F33.xx, F34.1 ICD-10). In accordance with the German national guideline, chronic low back pain is defined as pain in the area beneath the costal arch, above the inferior gluteal folds, with or without referred leg pain for more than 12 weeks and without signs of a specific cause (ICD-10 codes: M54.5, M54.8, M54.9). Potentially eligible patients will be identified using routine health records. Socio demographic information will be confirmed within the consultation. Diagnosis will be based on clinical judgment.”

#4 Blinding. Please, add something on the blinding status of outcome assessors and statisticians.

The outcome will be assessed by a self-rating instrument. Patients will be blinded to their group assignment. Our statistician will be blinded for the analysis of the primary outcome. We added the information to the text:

“Primary Outcome: The primary outcome will be the perceived usefulness of the written patient information material assessed by patient with the Usefulness Scale for Patient Information Material (USE) following the consultation (T1). Patients will be blinded to their group assignment.”

AND

“Statistical Analyses
The primary outcome (usefulness of written patient information material; USE) will be analyzed by a blinded statistician with a one-sided t-test comparing the scores on the USE between the intervention and the control group.”

#5 Study schedule. In the text and in table 1, please add the actual timing corresponding to –T1 and T1 (e.g. how many days, with T0 as reference).

_We added the information in the text and in table 1. However, when designing the study we did not define the exact time between –T1 and T1. Patients will be reminded after two weeks per phone and after 3 weeks per post. We revised the text accordingly._

In addition, it sounds strange that demographic information (“additional parameters”), including information on migration background and mother tongue, is not collected until T1, i.e. after the enrolment and allocation. Please, add also how the patients have to return the prepaid envelope at T1.

_We decided to collect demographic information via self-rating questionnaires. The information the physician has to collect was kept to a minimum to enhance feasibility of the study. The information how the patients have to return the envelopes was added to the text._

#6 Sample size calculation. Considering a Cohen’s d of 0.3, a Type I error of 0.05 and a power of 80%, how could the investigator get the sample size of 280 (140 per group)? Based on a simple t-test, one would get about 90 per group. Did the investigators account/adjust for something not declared?

_Two of the authors recalculated the necessary sample size independently and reached the conclusion that the calculation is correct. The number of individuals per group can be calculated as (2(Zα+Zβ)^2)/(d^2), where α and β are the type I error rate (note that we use a one-sided test) and the type II error rate (1-power), respectively. Z is the value of the standardized normal distribution and d is Cohen's d. with α=0.05 and β=0.80 the formula gives (2^*(1.65-0.84)^2)/(0.3^2)≈138._
The calculations were replicated with the software G*Power 3.1.3. Thus, no adjustment was performed.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

#7 Unit of recruitment. Please explicitly specify, in the abstract and early in the main text, that the patient is the unit of randomization.

We added the information to the abstract.

#8 Rationale. In the rationale paragraph of the “METHODS/DESIGN” (page 4, last paragraph), the investigators justify the choice of depression and chronic low back pain as the two index diseases saying that the former represents “a predominantly mental disorder” and the latter “a somatic disorder”; with this choice, the generalizability of their findings would be enhanced. In fact, this justification sounds not completely valid. As a clinician, it is common to experience the “psychic” component of a chronic low back pain (in fact it is not uncommon to see the association of a chronic low back pain with a depressed mood). Therefore, the choice of those two index disorders appears appropriate since they are chronic, generally highly prevalent (and in particular among people with a migration background), they “require an active involvement of patients in their own health care”, and are expected to be sensitive to the cultural background; but if the investigators were looking for a chronic somatic disorder with these characteristics, there are other conditions, with these characteristics, that would fit more the definition of somatic disease than the low back pain (to make an example, diabetes mellitus and management of insulin therapy).

We adapted the text in accordance to your suggestion:

“We chose depressive disorders and low back pain as medical indications as both conditions are among the leading causes of burden of disease worldwide [11]. Both conditions are often chronic, require an active involvement of patients in their own health care, and are expected to be sensitive to the cultural background. In addition, they are highly prevalent in German primary care [12, 13].”
Randomization. The randomization process, based on a computer-based algorithm, numbered sealed envelopes assigned consecutively to the enrolled patients, and 3 stratifying factors and blocks of variable size, seems quite complex. I would suggest describing it more clearly, specifying if it will be central or locally managed (e.g. each physician, provided in advance with his/her own envelopes, according to the disease and the migration background of the eligible patient, will simply hand to the patient the next envelope from the appropriate group of envelopes).

*We adapted the description in accordance to your suggestion (see page 7).*

Intervention. I would suggest the inclusion of some details on how the brochure looks like (length, inclusion of pictures, etc.).

*Information on length, inclusion of figures, pictures and case examples was added (see page 6).*

Outcomes. A) Some details on the content of the self-administered scale to assess “usefulness” should be provided.

*We have currently finished the development of the score and the article is under preparation. The USE turned out to be a reliable (Cronbachs α >.90) and valid scale (including structural, convergent and criterion validity). The SD was ~ 20 (see page 7).*

B) Among the feasibility issues of the study (Discussion), the high number of scales the patient needs to self-administer (with the risk of hindering patient response and increasing losses of follow up and missing data) might be recognized. C) Among the methodological issues of the study (Discussion), the use of several self-constructed scales, with unknown validity and reliability, might be recognized.

*We added these limitations to the discussion (see page 10).*
Additional parameters on participating practitioners. The rationale and the objectives for collecting data on the participating practitioners should be added.

*We collected data on the participating practitioners to gain information on the generalizability of our results. The text was adapted accordingly (see page 8).*

**Level of interest:** An article of importance in its field  
**Quality of written English:** Acceptable  
**Statistical review:** No, the manuscript does not need to be seen by a statistician.  
**Declaration of competing interests:**  
I declare that I have no competing interests

**Reviewer's report**  
**Reviewer:** Reitze Rodseth  
**Reviewer's report:**  
Development and testing of culture-sensitive patient information material for Turkish, Polish, Russian and Italian migrants with depression or chronic low back pain in a double-blind randomised-controlled trial (KULTINFO): A study protocol  
Thank you for the opportunity to review this manuscript. In this protocol the authors have designed a trial that attempts to determine the impact of culturally-sensitive information material provided to primary care patients with a migration background suffering from depression or chronic lower back pain in comparison to the administration of non-culturally adapted information material.

*Thank you for carefully reviewing our manuscript.*

**Major compulsory**  
The design chosen for this study is a double-blind, randomized controlled multicentre trial. This design is appropriate to address the question being posed in the trial objectives. The broad inclusion criteria used make it very likely that the results seen in this trial will be generalisable to the broader medical community.  
What I am not clear on is what exactly is meant by culturally-sensitive information. The target groups in this study are Turkish, Polish, Russian, and Italian. I can understand that these groups have different cultural norms to the German population. However, the degree of cultural variance with regard to the...
German population will obviously not be the same between the four groups. Reading the process by which the interventions will be developed I surmise that the interventions for each of the groups will differ from each other. One could further deduct that in populations where the material undergoes the greatest degree of modification you would then expect the greatest impact. All this would suggest that the intervention being studied would not have the same impact across the entire group.

*It is possible that the degree of modification varies between the different cultures. This variation may cause differences across the entire group. However, due to the design of the study (randomization on patient level) the primary effect estimate will not be biased by this difference. Although the idea of differences between the groups is plausible, we have no data for this assumption. We therefore decided to investigate the question of differences between the groups of migrants in an exploratory analysis.*

The randomization and allocation process seem robust and the blinding method seems appropriate for the situation. I am not convinced that patients would not be able to identify culturally-sensitive material, however, if the data collection is blinded, which it is, this should be adequate.

*We agree.*

The primary outcome of the trial is the perceived usefulness of the material. What would constitute a clinically significant change in this usefulness score?

*As patient information materials can be used with very little resources and therefore can reach large groups of patients we think that even a small between-group difference in the primary outcome can represent a meaningful effect on the population level. Our study was powered for a small effect (d=.3) which corresponds to a difference of 6 points (assuming a SD of 20) on a scale from 0-90. The text was adapted accordingly (see page 9).*

This would be of vital importance when determining the sample size required for a
successful trial. It is a pity that the trial has not been designed to address a more clinically robust outcome.

We decided to use the subjective rating of the usefulness of the information material as an outcome as it is based on current theoretical evaluation models of patient information brochures (Garner et al. 2012) and effects in the scale are a direct result of the kind of material that has been handed out to the patient. Other more clinically robust outcomes, e.g., change of symptoms/course of the disease, are more distant outcomes. We therefore expect them to be less responsive to our adaptations.


Further, the score is being developed on an inpatient population and has not been validated outside the hospital. Why would the authors believe this score to perform as well in an outpatient population without having tested it?

Based on the nature of the construct (subjectively experienced usefulness of written medical/health information materials) we found no reason to believe that setting would have any substantial impact on the psychometric properties of the scale. Of course, it can be tested a posteriori in the present and future studies.

Why will a one-sided t-test be used for scoring comparison? Surely it is possible that this material could result in patients perceiving it as less useful?

We absolutely agree that it is possible that the culturally adapted materials will be perceived as less useful and it would be desirable to be able to test this hypothesis as well (two-sided testing). Our rationale for using a one-sided test is a pragmatic one: limited resources. Trading off between power, size of the detectable effect, and type I error rate, we decided that we prefer having the possibility to detect small effects (d=0.3) with a sufficient power (>80%). However, the necessary sample size resulting from these parameters combined with a two-sided test was judged unfeasible by recruiting centers
during planning of the study. Thus, by switching to a one-sided test we have kept the detectable effect size and the power, but lost the possibility of testing the hypothesis that the new materials are less useful. In case of a statistically non-significant difference we will not know if usefulness of the adapted materials is similar or worse than that of the conventional ones. This trade-off was both subjective and unavoidable for a feasible study.

Please specify the direction that you expect to see in the sub-group analysis.

Sub-group analysis will be tested in a pure exploratory manner. We therefore do not have any expectations on the direction of the differences between subgroups.

I am concerned about the method by which the sample size has been calculated. First, it is not clear what a clinically significant improvement usefulness would be. Second, as the score has not yet been developed I am not clear what the normal variance of such a score would be. Developing the score in an inpatient setting, and then using this score to design an outpatient study, without validating the score, is concerning. Third, there is a high likelihood that there will be a different treatment effect from the intervention across the four groups. When taking these factors into account, as well as the multi-factorial nature of this intervention I believe this to be an over optimistically small sample size. The dropout rate is probably realistic though.

As patient information materials can be used with very little resources and therefore can reach large groups of patients we think that even a small between-group difference in the primary outcome can represent a meaningful effect on the population level.

We have currently finished the development of the score and the article is under preparation. The USE (range 0-90) turned out to be a reliable (Cronbachs α >.90) and valid scale (including, structural, convergent and criterion validity). The SD was ~ .20. The information was added to the text (see page 7).
Based on the nature of the construct (subjectively experienced usefulness of written medical/health information materials) we found no reason to believe that setting would have any substantial impact on the psychometric properties of the scale (see above).

There may be different treatment effects from the intervention between the groups, we think that our design (patient level randomization) is robust and our results will not be biased by these differences. We are going to test differences between the groups in exploratory analyses. Our study is powered to show a small, but in our opinion still meaningful difference. We have adapted the text accordingly.

Minor essential
The manuscript could be improved by minor language and grammar changes. For example “low back pain” should rather be written as “lower back pain”; “culture-sensitive” should read “culturally sensitive”. Words such as “operationalised” (pg 5) are not commonly used and alternative synonyms should be chose. Similarly the abstract states “…are only insufficiently reached by existing…” should read “… inadequately reached by existing…”

Thank you for these suggestions. We decided to use “low back pain” as this term is used in the NICE-Guideline, although “lower back pain” is also very often used. We changed the term “culture-sensitive” to “culturally sensitive”. We replaced the term “operationalized” with “defined”. The abstract was changed according to your suggestion.

Summary
The trial seems to be robust, and sufficient details have been provided to allow replication of the work. However, the sample size would seem to be inadequate and I have serious reservation about the scale that is to be used to evaluate the primary outcome.

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

**Quality of written English:** Needs some language corrections before being published
Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
No conflicts

Reviewer's report

Reviewer: James Paul

Reviewer's report:
Title: Development and testing of culture-sensitive patient information material for Turkish, Polish, Russian and Italian migrants with depression or chronic low back pain in a double-blind randomised-controlled trial (KULTINFO): A study protocol
Methodology: multicenter (4 institutions in Germany), double-blind randomised-controlled parallel-group study
Population: 480 patients with a Turkish, Polish, Russian or Italian migration background with a diagnosis of depressive disorder or chronic low back pain will be included. The sample size was rationalized with an anticipated Cohen’s d of 0.3 which is a small treatment effect as it assumes the difference in the primary outcome will amount to one third of a standard deviation. In addition the authors allowed for a 40% loss to follow up which is reasonable for a questionnaire study.
Intervention: Culture-sensitive patient information was handed to the patient at the end of the physician consultation.
Control: Standard translated patient information material will be administered.
Outcome(s) & Timeframe: Questionnaire after the physician consultation, and at 8 weeks and 6 months following the consultation. Primary outcome was subjective usefulness.
Analysis: The primary outcome (usefulness of written patient information material; USE) will be analyzed with a one-sided t-test comparing the scores on the USE between the intervention and the control group.

Instrument development:
Methodological Issues

Thank you for carefully reviewing our manuscript.

Major Compulsory Revisions
1. It is possible that the intervention could reduce the perceived usefulness of the patient information, hence the primary analysis should use a two-sided t-test.

   *We absolutely agree that it is possible that the culturally adapted materials will be perceived as less useful and it would be desirable to be able to test this hypothesis as well (two-sided testing). Our rationale for using a one-sided test is a pragmatic one: limited resources. Trading off between power, size of the detectable effect, and type I error rate, we decided that we prefer having the possibility to detect small effects (d=0.3) with a sufficient power (>80%). However, the necessary sample size resulting from these parameters combined with a two-sided test was judged unfeasible by recruiting centers during planning of the study. Thus, by switching to a one-sided test we have kept the detectable effect size and the power, but lost the possibility of testing the hypothesis that the new materials are less useful. In case of a statistically non-significant difference we will not know if usefulness of the adapted materials is similar or worse than that of the conventional ones. This trade-off was both subjective and unavoidable for a feasible study.*

2. Reliability and validity of the Usefulness Scale on Patient Information (USE) needs to be evaluated, either during its current development or in the design of this trial. Test retest reliability should be assessed and so should construct validity. If the patient information package is deemed as useful by the patient then presumably the patient understands their disease better. Maybe the patients can be tested on their disease, either by another questionnaire, and see if scoring well on the USE correlates with increased knowledge about their diagnosis.

   *We tested the psychometric properties of the Usefulness Scale on Patient Information (USE) in an independent study. We found the USE turned out to be a reliable and valid scale. The structural validity was confirmed in a confirmative factor analysis. Reliability analysis of the total scale showed a Cronbach’s α > .90. Convergent, divergent and construct validity could additionally be confirmed. The information was added to the text (see page 7).*
3. The impact of the information package might differ between patients with depression and low back pain. Particularly, patients with depression might be less receptive to an information package, regardless of its quality, given the nature of their condition. Also, the impact of the information package might differ depending on their course with their condition. Patients doing well with treatment might rate the usefulness of the package better for example. It would be worth quantifying the results overall and by disease type to explore this. The sample size might have to be adjusted accordingly.

   *We have added these additional analyses to our secondary analyses section. However, as these analyses are of exploratory nature we decided not to adjust the sample size calculation of our study, because the sample size was calculated with respect to our primary endpoint.*

4. The intervention might have a different impact on different cultures, with some finding in more or less useful. This variability in results might also increase the required sample size.

   *We added the idea for this analysis to our secondary analyses section (see page 8). As we have no specific hypotheses regarding differences between different cultures this analyses will be of purely explorative nature. As patients will randomly be allocated to the intervention and the control group, the results of the primary outcome will not be biased.*

5. As acknowledged by the study team, the planned intervention is quite complex, especially the cultural adaptation aspect. In order to aim for a consistent impact of the tool for all populations the cultural adaptation aspect should focus on presentation and not the content of the material.

   *We agree that the cultural adaptation has to focus on the presentation and not the content of the material. We revised the text accordingly to clarify this point (see page 6).*
Level of interest: An article of importance in its field
Quality of written English: Needs some language corrections before being published
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: None.

Reviewer's report
Reviewer: Ashley Bonner
Reviewer's report:
This was a very well written protocol.

Thank you for carefully reviewing our manuscript.

Major Compulsory Revisions:
1. Issue: The statistical analysis and sample size calculations for any trial fundamentally depends on the primary outcome. The manuscript in its current state does not provide sufficient details about the primary outcome to determine if the statistical analysis or sample size calculations are appropriate. The primary outcome for this trial is a measurement of perceived usefulness of the written patient information material. It is measured by the patient based on the Usefulness Scale for Patient Information Material (USE), that the authors mention is currently being developed. Requested resolution: Adding a description of the expected numerical range of this scale and nature of how the patients will respond to this scale would allow the analysis and sample size calculation portion of the manuscript to be interpreted and critiqued. Alternatively, if the development of the scale is at such a premature stage where this information is not known, adjustment of the statistical analysis section to include options, such as non-parametric tests if suitable for the outcome (e.g., Wilcoxon test), would be appropriate.

We already finalized the development of the USE and are currently preparing an article for publication. We are glad to inform you that the USE owns very
good psychometrical properties. The scale has 9 Likert items (from 0-10) and the range of the total scale is between 0 and 90 points. Reliability analysis of the total scale revealed a Cronbach’s $\alpha > .90$. Moreover structural, construct and criterion validity could be confirmed in our study. The text was revised accordingly:

“To be able to identify a small intervention effect (Cohens’ $d$ of .3, which corresponds to a difference of 6 points in the USE [assuming a standard deviation of 20]) with a Type I error of .05 and a power of 80%, a total sample size of 280 patients (140 per group) is needed for statistical analysis.”

Minor Essential Revisions:
None

Discretionary Revisions:
None

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

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests