Author’s response to reviews

Title: Feasibility and long-term efficacy of a proactive health program in the treatment of chronic back pain: a randomized controlled trial

Authors:

Angelika Hüppe (angelika.hueppe@uksh.de)

Christel Zeuner (christel.zeuner@uksh.de)

Sven Karstens (karstens@hochschule-trier.de)

Martin Hochheim (martin.hochheim@generali.com)

Max Wunderlich (max.wunderlich@central.com)

Heiner Raspe (hraspe@uni-muenster.de)

Version: 2 Date: 09 Jul 2019

Author’s response to reviews:

Point-by-point response to reviewer’s comments on our manuscript BHSR-D-18-02271 “Feasibility and long-term efficacy of a proactive health program in the treatment of persistent back pain”, now being submitted in its revised form

We thank Cary Carrington Reid for helpful comments. Please find below our point-by-point responses to issues raised:

1. How do you know that the back experts delivered the interventions (recommendations of the National Disease Management Guidelines) as intended?

All clinicians, doctors, physiotherapists etc. of the “FPZ Network” have signed a contract to strictly adhere to the recommendations of our national practice guidelines. We have, however, no empirical data on their adherence.

2. You report that participants received on average 191 minutes of coaching. Not clear whether this represents an amount near the maximum amount desired. Please elaborate.

Coaching had to respect patients’ individual needs, preferences and values. This alone must have led to some inter-patient variation. A maximum of 222 minutes spread over 16 contacts with each participant was planned. Our mean of 191 minutes is not far from the upper limit. The
actual range (see p. 11) of contact times reported/observed is from about 60 to more than 420 min.

Amendment to Methods Section (page 6, line 8)

A maximum of 222 minutes spread over 16 contacts with each participant was planned.

3. I believe readers will want to know whether number of therapeutic exercises attended and amount of coaching received over the phone were (or were not) associated with treatment outcomes. Other interactions that should be explored include participant age, gender and level of education.

The prescribed frequency of exercises followed the severity and prognosis of back pain of participants; 91% of all male participants received (as reported) a maximum of 24 hours. Coaching was not prescribed but requested and negotiated. Since the association between amount of coaching and outcome is most likely confounded by the severity of back pain and we see no convincing solution to this problem, we refrained from further (and in any case only descriptive-exploratory) analyses.

Referring to the interesting question about interactions between sociodemographic variables (age, gender, level of education) and treatment effects we follow the recommendation of Altman & Matthews (1996): “Results of tests for interactions are likely to be convincing only if they were specified at the start of the study”. Our statistical analysis plan only predefined subgroup analysis to explore interactions with different risks of persistent back pain measured by STarT-Back Tool. Further subgroup analyses - not “ex ante” specified - will result in less reliable data. However, to satisfy the reviewer’s (as also our) thirst for knowledge, we are able to say that none of the three demographic variables mentioned showed any interaction effect.

Amendment in the Methods Section (page 9, line 6-7)

… subgroup analyses were done to explore the heterogeneity of treatment effects in participants with different risks of persistent BP.

Amendment in the Discussion Section under Limitations (page 15, line 1-3)

The interesting question of whether sociodemographic variables (as age, gender, formal education) were (or were not) associated with treatment outcomes remains unanswered, being outside the scope of the study.

4. Concerns about your primary exposure variable. Table 1 provides the criteria used to determine eligibility, but reviewing these criteria it is not clear to me that your sample would be composed of individuals with persistent back pain. For example, to have 2 cases assigned to an ICD code for dorsopathy over what period of time? A patient could receive two or more opioid prescriptions for an acute pain problem. The temporary work disability could have happened
prior to recruitment and not be active. These criteria seem to lack sensitivity for identifying individuals with persistent back pain. Did you consider to do a medical record review to determine how participants had a diagnosis of persistent back pain their chart?

Recruitment was based on administrative data alone. We agree that our search keys (Table 1), neither singly nor in combination, guarantee the inclusion of all subjects with persistent back pain (due to lack of sensitivity) and only those with this condition (lack of specificity). In our view, incomplete sensitivity seems to be the minor problem. However, all study participants gave consent after being informed elaborately (in writing) that we will include only subjects with persistent back pain. Figure 1 shows that only a minority (< 20 %) of insured persons initially contacted actually participated in the study. We assume that those without long-lasting back pain problems did not respond to the study invitation (self-selection).

5. What was the rationale for using data collected at 2 years for your between-group comparisons (as opposed to looking at change scores for individuals who provided data at baseline and follow-up)?

We decided against analyzing change scores because we found imbalanced baseline scores for primary and secondary outcomes (IG had worse scores than CG). In this case, using change scores from the baseline would produce biased effect estimates of mean difference (regression to the mean). Using ANCOVA provides the least biased estimates (see Vickers & Altman: Statistics Notes - Analysing controlled trials with baseline and follow up Measurements in BMJ 2001;323:1123–4).

Other issues

- Title, would recommend removing 'using Zelen's design' from the title, this is not necessary.

We followed this recommendation, see (page 1, line 2)

- Abstract, line 11. What are settlement data?

We have some difficulties to find the correct translation: settlement data, billing data, payroll data, medical claims data? It is about administrative data of the health insurance that comprise information on type and cost of treatment - outpatient treatment (e.g. drugs) and/or inpatient treatment (e.g. surgeries) and daily sickness allowance of the insured persons.

We now consistently use the wording “billing data” in the manuscript.

Correction in Abstract: page 2, line 6 and Methods: page 4, line 25

- Methods section, would provide some description of Zelen's design in your method's section.

Amendment to Methods: page 4, line 20-21
The specific characteristic of Zelen’s design (also called randomized consent design) is that consent to participate is sought only after randomization [18, 19].

- Methods section: How did you gain access to payroll data. That would not be allowed in the US.

Probably we used an incorrect translation (see point above). We did not have access to “payroll” data of the study participants in terms of their “earnings”. Regarding the data used as selection criteria: only the employees of the private health insurance had access to them, the researchers had no access.

- Results, did you compare attributes of those who responded to the recruitment letter to those who did not (in terms of the criteria used to determine eligibility for study entry)?

This would be highly interesting. But unfortunately we had no access to these data of the health insurance company.