

BMJ Open Clinical trial participants' experiences of completing questionnaires: a qualitative study

Christine Holmberg,^{1,2} Julia J Karner,¹ Julia Rappenecker,¹ Claudia M Witt^{1,3}

To cite: Holmberg C, Karner JJ, Rappenecker J, et al. Clinical trial participants' experiences of completing questionnaires: a qualitative study. *BMJ Open* 2014;**4**:e004363. doi:10.1136/bmjopen-2013-004363

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-004363>).

Received 30 October 2013
Revised 10 January 2014
Accepted 20 February 2014



CrossMark

¹Institute for Social Medicine, Epidemiology, and Health Economics, Charité Universitätsmedizin Berlin, Berlin, Germany

²Berlin School of Public Health, Charité Universitätsmedizin Berlin, Berlin, Germany

³Institute of Complementary and Integrative Medicine, University Hospital Zürich, Zürich, Switzerland

Correspondence to

Dr Christine Holmberg;
christine.holmberg@charite.de

ABSTRACT

Objectives: To improve clinical study developments for elderly populations, we aim to understand how they transfer their experiences into validated, standardised self-completed study measurement instruments. We analysed how women (mean 78±8 years of age) participating in a randomised controlled trial (RCT) cognised study instruments used to evaluate outcomes of the intervention.

Setting: The interview study was nested in an RCT on chronic neck pain using common measurement instruments situated in an elderly community in Berlin, Germany, which comprised of units for independent and assisted-living options.

Participants: The sample (n=20 women) was selected from the RCT sample (n=117, 95% women, mean age 76 (SD±8) years). Interview participants were selected using a purposive sampling list based on the RCT outcomes.

Outcomes: We asked participants about their experiences completing the RCT questionnaires. Interviews were analysed thematically, then compared with the questionnaires.

Results: Interviewees had difficulties in translating complex experiences into a single value on a scale and understanding the relationship of the questionnaires to study aims. Interviewees considered important for the trial that their actual experiences were understood by trial organisers. This information was not transferrable by means of the questionnaires. To rectify these difficulties, interviewees used strategies such as adding notes, adding response categories or skipping an item.

Conclusions: Elderly interview participants understood the importance of completing questionnaires for trial success. This led to strategies of completing the questionnaires that resulted in 'missing' or ambiguous data. To improve data collection in elderly populations, educational materials addressing the differential logics should be developed and tested. Pilot testing validated instruments using cognitive interviews may be particularly important in such populations. Finally, when the target of an intervention is a subjective experience, it seems important to create a method by which participants can convey their personal experiences. These could be nested qualitative studies.

Trial registration number: ISRCTN77108101807.

Strengths and limitations of this study

- Validated study instruments insufficiently capture elderly participants' experiences, qualitative methods can facilitate a better understanding of the experiences with RCT interventions and outcomes.
- The use of qualitative methods helped to understand what clinical trial participants understand as 'good data'.
- Clinical trial participants knew the importance of complete data collection for the success of the trial, but they had a different understanding of 'good data' than the one underlying quantitative study instruments.
- The study sample had experienced pain for a very long time. This may make this group of women especially thankful for providing options to treat their long-lasting pain and therefore more eager to comply with study requirements. Thus, these findings may be particular to an elderly and female population.

INTRODUCTION

There are many factors that are crucial to the success of clinical trials, including validated study instruments. An adequate assessment of the study endpoint is a crucial aspect of clinical trials; for this, validated questionnaires are considered as one of the assessment tools for this purpose. The utilised instruments should be able to measure the same constructs consistently and accurately across individuals. There are some well-known questionnaire completion strategies such as marking the midpoint of a scale that prevent an accurate assessment of outcomes. Much effort has been devoted to the design of study instruments to discourage such behaviour.

The gold standard to assess subjective study endpoints is valid and reliable instruments. Validity corresponds to the question of how well an instrument measures what it intends to measure, such as pain and intensity.¹ Reliability is established through tests

and retests and validity through the comparability of a scale with other scales.^{2,3} This means that if an instrument is used repeatedly and achieves the same results throughout or gives similar results to an instrument that has already been validated then its results are considered valid and reliable. For fluctuating, subjective experiences, such as pain, reliability and validity of scales only depicts part of the picture.⁴⁻⁷ The experience of pain is influenced by context, meaning, emotional aspects, expectations, attitudes and beliefs associated with pain.⁵ These aspects make it difficult to know what dimensions pain scales capture. Indeed, while commonly used one-dimensional pain rating scales, such as the reliable visual analogue scale (VAS)⁸ are considered the gold standard for pain assessment,⁹ have been validated in various populations, including elderly populations,¹⁰⁻¹⁶ and are more often used in clinical practice and research¹⁷; it remains unclear what the meaning of the information that such one-dimensional pain scales deliver represents.^{18,19} Thus, diagnosing chronic pain poses problems to researchers and clinicians, despite existing validated instruments.²⁰⁻²²

As the example of one-dimensional pain scales show, adequate results for commonly used performance criteria such as validity and reliability do not necessarily suggest that such results suffice to depict complex subjective experiences.

In a randomised controlled trial (RCT) that compared the effects of Qigong and exercise therapy on neck pain in the elderly, no effect on pain intensity could be detected.²³ Three groups were compared: a Qigong group, an exercise therapy group and a waiting list group. No difference between groups was found for the primary (VAS) and the secondary endpoints (Neck, Pain and Disability Scale based on a common depression, health-related quality of life, sleep quality and satisfaction with the therapies). However, almost all study participants indicated that they would recommend the therapy to others and some even chose to continue the interventions at their own expenditure.²³ Thus, we were interested to understand how participants transferred their observations and experiences into the study measurement instruments.

The analysis aimed to understand how women (mean 78±8 years of age) who participated in an RCT cognised the study instruments that were used to evaluate the primary and secondary endpoint outcomes of the intervention.

METHODS

Study design

We conducted a qualitative study nested within an RCT to better understand the RCT results.²³ The trial was conducted by the Institute for Social Medicine, Epidemiology and Health Economics at the Charité Universitätsmedizin Berlin. Participants gave written and oral consent to participate in the RCT. The selected participants were invited

by phone to participate in an interview on their experiences with the RCT. The interviews took place at the participants' homes and they were asked to provide additional oral consent for a home visit. The consent process was documented in the case report forms. The RCT included 117 patients with chronic neck pain who were randomised to a Qigong group, an exercise therapy group or to a waiting list group. At three different time points, all three groups completed four validated questionnaires: the VAS, the Neck Pain and Disability Scale (NPDS),²⁴ the Short-Form-36-Questionnaire (SF-36)^{25,26} and a common depression scale (ADS).²⁷ The NPDS is a specific evaluation instrument for neck pain that has been shown to be valid and reliable to measure neck pain²⁸⁻³⁰ and to detect clinically relevant changes in neck pain.³¹ It consists of 20 items that assess intensity of pain using neck problems as well as emotional and cognitive influences on work and everyday life.³² The ADS, which is the German version of the Center for Epidemiological Studies Depression Scale (CES-D),³³ assesses the length and adverse effects of depressive symptoms, bodily problems and negative thought patterns. This instrument is recommended for use with chronic pain patients.³⁴ These instruments are the standard tools for these diagnoses. However, they are not satisfactorily validated for the age group under study.²³

We developed a semistructured interview guide that included questions related to the intervention and study instruments, more specifically asking about difficulties the patients may have had in completing the questionnaires and what was important for them in their experiences related to the study interventions. Prior to conducting the interviews, the interview guide was piloted in mock interviews with older patients with neck pain to ensure that the questions functioned well and the information was received as intended by the study aims.

Recruitment

In order to achieve a diverse selection of interview participants from the quantitative study (QIBANE) for the interview study, sampling was based on the results of the primary endpoint of the study. We wanted to ensure that the interview sample reflected the entire range of responses to the primary endpoint, which was decreased in neck pain as measured by the VAS.³⁵ In addition, secondary endpoints such as the NPDS³⁶ and the quality of life questionnaire SF-36³⁷ were considered as secondary criteria for sample diversity. Thus, we created different groups of QIBANE participants: one group comprised QIBANE participants who had indicated an improvement of symptoms between baseline and follow-up assessments, one group who had shown worsening symptoms and the other group comprised those who had no change between baseline assessment and 3-month follow-up. In each group, a ranking was established that started with the individuals with the largest differences between both assessment points. Once the rankings were established, participants were called until 10 participants from the Qigong group and 10 participants from

the exercise therapy group agreed to a qualitative interview. Interviewee recruiters who called participants had previously conducted the RCT and were known to participants. Participants in the RCT were mostly women (95%), which led to a list of potential interview participants which comprised predominantly women. Recruitment ended after the first 10 RCT participants from the Qigong and another 10 from the exercise therapy group had agreed to participate. A sample size of 20 participants was chosen based on the experiences of other qualitative studies that were nested in RCTs.^{38 39}

Data collection

Interviews were conducted in the homes of the participants to ensure that participants felt comfortable and were willing to speak openly.⁴⁰ Interviewers had previously organised the RCT and were well known to the interviewees. Interviews were conducted at the home of the interviewees to accommodate study participants and to create a relaxing atmosphere for the interviewee.⁴⁰ To help their memory, interview participants received blank sample questionnaires similar to the ones they had filled out during their RCT participation. While an interview guide was prepared for the interview, it was used in a flexible manner to allow for discussion that was important to the interviewees.^{41 42} After each interview, the interviewer completed a standard protocol developed by Miles and Huberman⁴³ to capture the atmosphere, setting and main themes of the interview. Interviews were digitally recorded and transcribed. The text documents were then entered into software programme ATLAS.ti for coding and analysis.

Data analysis

As interviewers were not involved in data analysis, the interview protocols provided the contextual information for the research team to situate the interview, its dynamics and content. Analysis of the study was multilayered. As a first step, the qualitative interview materials were read by all researchers and analysed independently by JK and JR using content analysis according to Mayring.^{41 42 44} This allowed focusing the analysis on the interview passages in which the questionnaires were discussed. The coding scheme was developed based on the emerging themes from the interview material by two of the authors

(JK and JR) and then refined by the research team (all authors). In addition, coding and results were regularly presented and discussed in a qualitative working group. The goal of the presentation to the working group was to ensure that materials and results were consistent with each other and to broaden the perspectives on the materials and ensure intersubjectivity of results. After analysis of interviews, we compared the quantitative questionnaires that had been completed by the interviewees in the RCT with interview results to identify strategies of how study instruments were used.

RESULTS

Sample description

Of those who were called and invited to participate in the interview study, six declined a home visit due to fear of fraud. A short time prior to the recruitment for the interview study there had been some robberies in the senior residency and there was heightened awareness with regard to possible scam calls. The remaining 20 people agreed to participate in the interviews. Table 1 shows the changes the interviewees had indicated on the validated scales during the RCT. Eleven of the interviewees indicated a wish to continue the therapy even though they had not experienced an improvement of pain according to the validated instruments.

All interviewees were women with an average age of 76 years. They had an age range of 67–85 years. On average, they had experienced pain for 15 years. All interviewees lived in residencies for seniors in Berlin.

Experiences completing the questionnaires

Many of the interviewees were dissatisfied either with the questionnaires and scales that they had to complete or the strategies they used to complete them. They complained about the difficulties of expressing complex experiences in the standardised terms the questionnaire asked of them.

Questionnaires are always terrible because you never can express by checking a box what one wants to say. [QG2/241]

If I make this movement, it hurts here. If I make that movement, it hurts there. Now the pain is gone. Now I

Table 1 Changes in measurements between baseline and primary endpoint of the interviewees

Questionnaire*	Improvement (number of patients)	Worsening (number of patients)	Missing data (number of patients)
VAS	9	11	0
NPDS	13	6	1
SF-36	11	8	1
ADS	5	8	7

*One participant had no change in the NPDS.

ADS, Common Depression Scale; NPDS, Neck Pain and Disability Scale; SF-36 (mcs=mental component score: mental component summary scale of the Short-Form-36-Questionnaire); VAS, visual analogue scale.

look at you and I don't experience any pain. Now you tell me, do I have pain or do I not have pain? You tell me! [QG2/318]

Some women were also concerned about the type of questions that were asked of them; questions related to their mental state as asked on the ADS and partially in the NPDS were especially disconcerting to some interviewees. Some interviewees were concerned that study staff may not adequately interpret their answers in the questionnaires because they were not able to precisely express on them how they felt.

In these questions one often has potential answers that partially fit and partially do not fit, so that one would say, 'yes, that is how it is, but....' (...) and since there is no possibility for the opposite, the whole answer isn't right. [QG10/011]

None of the interviewees felt that their experiences with pain or with living as an elderly person could be adequately described based on responses to the questionnaires that were administered to them. Particularly, translating complex experiences into a single response on a scale was a challenge for the women. Participants used different strategies to deal with these problems when they completed the scales. These were mainly additional notes, placing the mark in the middle of a scale, adding answer categories or skipping an item. The women used these strategies because they felt that the scales could not capture their individual experiences. At the same time at least some felt indebted to the study since it gave them free exercise classes and they wanted to attend to the questionnaires in the best possible manner. Thus, they added to the questionnaires the information they found pertinent.

Specifying standardised answers

Adding notes

Adding notes was a common strategy among the interviewees. Of the interviewees, 15 added information on an item to clarify what the value on the scale they marked signified. For example, one participant added to her answer for the item, 'frequency of physical activity', the time frame, '30 or 60 minutes!' and the circumstances of the exercise, 'with partner or by myself' [PN7]. The same participant added to her answer to the item, 'frequency of falls', 'in snow'. She had indicated that she had fallen once. Finally, on the NPDS, the patient wanted to specify her pain and added "in the lumbar spine and in the knees." Another participant added a note to the value she selected on the VAS, "I exercise daily. This is the only way I can remain relatively painless" [PN6].

Others added handwritten notes to the response options instead of selecting a response on the scale. For example, one participant [QG5] added verbal signifiers to the scale on the NPDS such as *seldom*, *satisfied* or *little*. In the interviews, this particular participant reported about the questionnaires. Another participant [QG6]

specified one question in the NPDS in the interview. Instead of putting a mark next to the question "does the pain hinder you with activities such as eating, dressing or hygiene?" the participant responded by writing *dressing*.

Similarly, where items asked for specific time frames, participants sometimes chose to change the time frame in order to meaningfully answer the questions. They noted on the side the time frame they referred to in the answer. For example, for a question that asked for a judgement of the past 3 months, one respondent wrote, "[t]his has been in the last six months" [QG10]. The theme addressed in the question seemed more important to the interviewees than the requested time frame.

Selecting parts of an item

Another strategy to respond to the questionnaire and specify general questions was to underline parts of a question to highlight what exactly the answer referred to. For example, one participant [PN7] underlined 'kneeling' in an item of the SF-36 that stated "to bend forward, kneeling." Another such example comes from an either/or question in the NPDS. Two of the participants [PN1] [QG4] marked one of the two given possibilities in the item, "How difficult is it for you to look up or down?" Underlining was also used in questions that required a response along a scale. Several of the interviewees simply underlined one of the top or bottom values on the scale instead of marking a point along the scale.

Being a study participant

Some of the women had a clear understanding of the reciprocal relationship with the staff of the RCT; the women received interventions in exchange for completing the questionnaires. However, this required that they seriously considered their responsibility and wanted to complete the questionnaire adequately and accurately. Furthermore, the questionnaires used in the study left some women feeling uneasy and unhappy with their contribution.

I was actually glad when I was done, just like school work that I had to do and I did very thoroughly. But I was not satisfied with my work and also not with the questions! So, I wasn't— but I have experienced such feelings with other questionnaires before. [QG2/265]

I hope I have filled out everything correctly. I do not know if I filled them out correctly. [PN9/163]

One woman called a family member and her family physician to assist her in completing the questionnaire in order to ensure the correctness of the questionnaires.

I don't remember for which question that was. I really did not know what to do with that question. I did not want to do anything wrong, so I called my daughter. She is a teacher and she also really had to think about it. But I cannot tell you which question that was at the time. I

don't know. But the question was phrased very strange. [QG8/207]

Seriously considering their role as study participant was a common theme in the interviews and was the main reason why interviewees were dissatisfied with the assessment tools used in the RCT. The interviewed women assumed that their precise and exact experiences were of importance to the clinical trial staff and they were very concerned that the staff could not interpret their marks correctly on the assessments. The women made clear in the qualitative interviews that they preferred such an assessment much more than the questionnaires because the interviews enabled them to correctly state their experiences.

One just could not answer that question clearly. I don't know. I basically followed my feelings—but did you understand my answer? What you really get out of my answer is the question [...]. So I had the feeling after I filled out the questionnaire that you cannot learn anything from those answers. I guess I would have to say: I would not trust those questionnaires. But those are your main interest, aren't they?! [QG6/028]

The questions [in the questionnaire] do not make sense. I would have thought it better if you would have, just like you are doing now, asked the people directly. [QG4/250]

DISCUSSION

The interviewees in this study considered their role as a study participant important and perceived it as their responsibility to answer the questionnaires as accurately as they could to depict their experiences with chronic pain. However, they reported that this task was not easy for them. They had difficulties making non-specific statements about specific experiences and many thought that their experiences could not be depicted in the questionnaires; many also feared that their answers could be misunderstood. Several strategies were used by respondents to deal with the problem, such as adding notes, marking particular parts of a question or leaving an item open. Some also asked others to help them complete the questionnaires correctly. Strategies such as adding notes have been called 'optimising' strategies.⁴⁵ In addition, leaving an item blank or putting the mark in the middle of a scale is called 'satisficing' strategy, suggesting that questions are answered cursorily.

Satisficing strategies are more common when study participants do not understand why certain questions are asked.⁴⁶ This is what the interviewees described especially in relation to the ADS. In the rationality of the researchers, it was necessary to use the ADS for the RCT because an association between depression and chronic pain has been found before, and therefore needed to be controlled for in the RCT. However, examining the association between mental state and pain was one that alienated research participants from the study. They did not consider pain and

mental state as related to each other. The age group of the interviewed women may be one in which depression and other psychiatric diseases have a strong stigma associated with them. In addition, women in the study population may belong to a generation that had learned that one had to be strong and go about one's business without any problems. Such attitudes may make it difficult to admit psychological problems as well as difficulties with chronic pain more generally. The conflict between the logic of quantitative research and that of the study participants was obvious throughout the interview results. Medical research needs standardised questionnaires of intraindividual and interindividual comparisons and a particular kind of objectivity.⁴⁷ It depends on decontextualising personal experience in order to make the experience comparable and transferrable independent of time and place. This contrasts with the participants' sense of personal experience. Participants aimed to describe a precise and specific personal experience that aimed at being as accurate as possible. Questionnaires are developed to deduce complex experiences for statistical analysis. For our interviewees, this reduction, in fact, meant that it was more difficult to answer the questionnaires and some of the interviewees felt frustrated by their inability to give an exact depiction of their experience through their answers to the questionnaires.

The extra effort interviewees went through to document their particular experiences contradicted the researchers' efforts to obtain quantitative data that are comparable across time and place. To adequately present their experiences, interviewees manipulated the questionnaires where they found it necessary for a more accurate description of their experiences. In addition to adding notes, women in our sample marked different points of a scale to describe their experiences. While this was an optimising strategy for the women, researchers consider such items as 'missing data' or 'unscorable data'. Thus, the effect was the opposite of what the women had intended and, in fact, these strategies could undermine the validity of study results. The interviewed women aimed at optimising their data to give a full picture of their experiences and, in some instances, produced data that were then not interpretable anymore from a statistical point of view, for example, two marks on one scale. However, the range of using these strategies and the amount of missing data in the overall study (5% across all measurements and time points) are comparable to other RCTs. To minimise such faulty data, it is important to know how the elderly population may understand the significance of their study participation in order to intervene and improve data collection in this age group.

The conflict lies in a classic problem: questionnaires by default oversimplify complex experiences. The way these are reduced reflect interests of the researchers more than the patients.⁴⁸ In the process of such reduction, research subjects, in fact, become objects that produce data that are acceptable to the researchers.⁴⁹

What are the implications to potentially untangle these two different logics that clash in clinical trial participation, specifically in completing questionnaires?

Warms *et al*⁵⁰ analysed the strategy of adding notes more closely and found that questionnaires were seen as a means that study participants communicated with study researchers. This corresponds to our findings of the importance interviewees assigned to trial participation. As such they assumed their individual experiences were of importance. When the communication tool is not perceived as a good one, study participants may react with frustration.⁵¹ Again, this may have direct consequences on study results as it may lead to satisficing strategies in completing the questionnaires.⁵² However, these strategies are not a sign that study participants do not want to comply with study requirements. On the contrary, our study participants developed these strategies precisely because they knew how important accurate data are for an RCT to be successful. Thus, understanding adding notes as a participant's wish to directly communicate with the researchers of the study and as participants' correct understanding of the importance of completing these questionnaires helps develop solutions that may not undermine the efforts of the research. Nesting qualitative components such as interviews into clinical trials may facilitate such communications and help to respect participants' perspectives and give them voice to communicate with researchers.^{53 54} In addition, measurements have been developed that assess participants' experiences of study participation.^{55–57} If the study endpoint consists of a subjective experience that needs to be assessed in a standardised manner, it may be necessary to address that 'accurate' has a particular meaning in research that may differ from how study participants consider 'accurate' and explain the importance of sticking to provided instructions. It may be useful to develop such a standard leaflet explaining the need of standardisation.

In our interview study, only women participated. The RCT in which this study was nested had mostly female participants (95%). Thus, the female sample in the interview study is a reflection of the RCT population, which, in turn, is a reflection of the larger proportion of women in this age group overall. Regardless, it is likely that men may not have been as eager to adequately depict their personal experiences in the questionnaires or would not have taken their responsibility as study participant as important as have the interviewees in the study. Similarly, since we only interviewed 20 of the 117 RCT participants, it is conceivable that those who agreed to participate in the interview study took their role as study participant seriously. However, since we had created a ranking list with which we began recruitment and only six refused because they feared fraud, it would be surprising that we found those that were extraordinarily eager. Their seriousness about study participation could be a reflection of the values of a particular generation and age group. Finally, this group of women had

experienced pain for a very long time and was open enough to try treatments that they had not tried before, which may make this group of women especially thankful for providing options to treat their long-lasting pain. It is, therefore, possible that these findings are particular to an elderly population. Considering that there is a need for more medical research in elderly populations, it seems important to carefully evaluate the types of questionnaires used in such populations and to consider ways to explain the importance of standardised answers for clinical trials research. The need to use cognitive interviewing to improve questionnaires has been voiced before⁵⁸ and the findings of this study underline the importance of such pilot testing before instruments are used in specific populations.

While the strategies that were used by the women in this study in completing their questionnaires have been described in the literature,^{45 46 52} no study has yet described the relationship between perceptions research participants have about their role and the ways they complete their questionnaires. Overall, participants were frustrated with the questionnaires used, all of which are standards for diagnosis that are commonly used in research. To improve knowledge production in medicine, it may be important to address these differential understandings of the ways in which clinical trial participants are of importance.

In this study, we showed that a clear discrepancy existed between the logic of quantitative research and the logic of RCT participants. Interviewees thought it was important for the trial that their actual experiences were understood by trial organisers. These were not transferrable by means of the provided questionnaires, so they added their experiences by handwritten notes on the questionnaires. However, the statistical analysis of RCT data needs this reduction of experience in order to produce results.⁵⁹ Study participants are a crucial component of clinical trials research as they are necessary for data production, but these data necessarily are reductionist and aim to generate data that are comparable and quantitative in nature. Individual experiences need to be reworked to fit such criteria as comparability and objectivity. Interviewees who had participated in QIBANE knew of their importance for the trial. Consequently, they seriously considered their task of filling out questionnaires and tried to provide the best possible information. However, it was exactly this effort that, in some cases, led to strategies to convey their personal experience as best as possible, that undermined the aims of the study to get complete data. To improve data collection, increased effort may have to be invested in educating about the ways 'experiences' need to be translated into comparative, standardised information to be able to use them for clinical trials research and what 'accurate' filling out of questionnaires means from a research perspective. Similarly, additional venues to the regularly used validated instruments that measure subjective and fluctuating experiences should be implemented to enable research participants to voice their experiences. These

could include group discussions or interviews. Integrating qualitative and quantitative components such as implementation and process evaluation in addition to interviews can provide essential information that can improve research with this unique and growing population.

Contributors CH and CMW have developed the study design and have supervised the analysis of the study. JR and JK have analysed the materials. CH has written the manuscript and CMW, JR and JK have given substantial input throughout the development and writing of the paper.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None.

Ethics approval Charite Ethics Committee and ethics clearance by the appropriate ethics review board (EA1/265/05).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The interview transcripts are available to show proof of the paper. However, they would only be available for legal purposes. They are confidential and can only be given access to in case of legal requirements.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Kollbaum A. *Die Gütekriterien der quantitativen Forschung*. Flensburg: Universität Flensburg, 2009.
- Ferrell BA. Pain management. *Clin Geriatr Med* 2000;16:853–74.
- Kaasalainen S, Crook J. A comparison of pain-assessment tools for use with elderly long-term-care residents. *Can J Nurs Res* 2003;35:58–71.
- Rosier EM, Iadarola MJ, Coghill RC. Reproducibility of pain measurement and pain perception. *Pain* 2002;98:205–16.
- Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J Clin Nurs* 2005;14:798–804.
- Wood BM, Nicholas MK, Blyth F, et al. Assessing pain in older people with persistent pain: the NRS is valid but only provides part of the picture. *J Pain* 2010;11:1259–66.
- Katz J, Melzack R. Measurement of pain. *Surg Clin North Am* 1999;79:231–52.
- Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. *Acad Emerg Med* 2001;8:1153–7.
- Miaskowski C, Bair M. *Principles of analgesic use in the treatment of acute pain and cancer pain*. American Pain Society, 2008.
- Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117–26.
- Flaherty E. Pain assessment for older adults. In: Greenberg SA, ed. *Try this: best practices in nursing care to older adults*. Hartford Institute for Geriatric Nursing, 2000.
- Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 2011;41:1073–93.
- Peters ML, Patijn J, Lame I. Pain assessment in younger and older pain patients: psychometric properties and patient preference of five commonly used measures of pain intensity. *Pain Med* 2007;8:601–10.
- Summers S. Evidence-based practice part 2: reliability and validity of selected acute pain instruments. *J Perianesth Nurs* 2001;16:35–40.
- Herr KA, Spratt K, Mobily PR, et al. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin J Pain* 2004;20:207–19.
- Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Validity of four pain intensity rating scales. *Pain* 2011;152:2399–404.
- Eccleston C, Crombez G, Aldrich S, et al. Attention and somatic awareness in chronic pain. *Pain* 1997;72:209–15.
- Broderick JE, Stone AA, Calvanese P, et al. Recalled pain ratings: a complex and poorly defined task. *J Pain* 2006;7:142–9.
- de CWAC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. *Pain* 2000;85:457–63.
- Breivik H, Collett B, Ventafridda V, et al. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain* 2006;10:287–333.
- Gagliese L, Melzack R. Chronic pain in elderly people. *Pain* 1997;70:3–14.
- Miller LL, Talerico KA. Pain in older adults. *Annu Rev Nurs Res* 2002;20:63–88.
- Von Trott P, Wiedemann AM, Ludtke R, et al. Qigong and exercise therapy for elderly patients with chronic neck pain (QIBANE): a randomized controlled study. *J Pain* 2009;10:501–8.
- Wheeler AH, Goolkasian P, Baird AC, et al. Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. *Spine* 1999;24:1290–4.
- Bullinger M. [Assessment of health related quality of life with the SF-36 Health Survey]. *Rehabilitation (Stuttg)* 1996;35:XVII–XXVII.
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- Hautzinger M. *Allgemeine depressions skala*. Weinheim: Beltz, 1993.
- Goolkasian P. Neck Pain and Disability Scale: a critical evaluation. *Expert Rev Pharmacoecon Outcomes Res* 2003;3:379–82.
- Salo P, Ylinen J, Kautiainen H, et al. Reliability and validity of the Finnish version of the neck disability index and the modified neck pain and disability scale. *Spine* 1976 2010;35:552–6.
- Wu S, Ma C, Mai M, et al. Translation and validation study of Chinese versions of the neck disability index and the neck pain and disability scale. *Spine* 1976 2010;35:1575–9.
- Wlodyka-Demaille S, Poiraudou S, Catanzariti JF, et al. The ability to change of three questionnaires for neck pain. *Joint Bone Spine* 2004;71:317–26.
- Goolkasian P, Wheeler AH, Gretz SS. The neck pain and disability scale: test-retest reliability and construct validity. *Clin J Pain* 2002;18:245–50.
- Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385–401.
- Oswald WD, Gatterer G, Fleischmann U, et al. *Ist Altern messbar? Gerontopsychologie*: Springer Vienna, 2008:59–77.
- Lara-Munoz C, De Leon SP, Feinstein AR, et al. Comparison of three rating scales for measuring subjective phenomena in clinical research. I. Use of experimentally controlled auditory stimuli. *Arch Med Res* 2004;35:43–8.
- Scherer M, Blozik E, Himmel W, et al. Psychometric properties of a German version of the neck pain and disability scale. *Eur Spine J* 2008;17:922–9.
- Morfel M, Bullinger M, Nantke J, et al. [The version 2.0 of the SF-36 Health Survey: results of a population-representative study]. *Soz Präventivmed* 2005;50:292–300.
- Campbell R, Quilty B, Dieppe P. Discrepancies between patients' assessments of outcome: qualitative study nested within a randomised controlled trial. *BMJ* 2003;326:252–3.
- Daley AJ, Copeland RJ, Wright NP, et al. 'I can actually exercise if I want to; it isn't as hard as I thought': a qualitative study of the experiences and views of obese adolescents participating in an exercise therapy intervention. *J Health Psychol* 2008;13:810–19.
- Manderson L, Bennett E, Andajani-Sutjahjo S. The social dynamics of the interview: age, class, and gender. *Qual Health Res* 2006;16:1317–34.
- Flick U. *Qualitative Forschung—ein Handbuch*. Reinbek bei Hamburg: rowohlt's enzyklopädie, 2007.
- Mayring P. *Einführung in die Qualitative Sozialforschung*. Weinheim und Basel: Beltz Studium, 2002.
- Miles M, Huberman M. *Qualitative data analysis*. An Expanded Sourcebook. London: Sage Publications, 1994.
- Morse JM, Tashakkori A, Teddlie C. Principles of mixed methods and multimethod research design. In: Tashakkori A, Teddlie C. eds. *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage, 2003:189–208.
- Krosnick JA. Survey research. *Annu Rev Psychol* 1999;50:537–67.
- Holen JC, Hjermstad MJ, Loge JH, et al. Pain assessment tools: is the content appropriate for use in palliative care? *J Pain Symptom Manage* 2006;32:567–80.
- Basler HD. [The Marburg questionnaire on habitual health findings—a study on patients with chronic pain]. *Schmerz* 1999;13:385–91.
- Girtler R. *Methoden der qualitativen Sozialforschung*. Wien: Böhlau, 1984.
- Lamnek S. *Gruppendiskussion. Theorie und Praxis*. Weinheim: Psychologie Verlag Union, 1998.



50. Warme CA, Marshall HM, Hoffman AJ, *et al.* There are a few things you did not ask about my pain: writing on the margins of a survey questionnaire. *Rehabil Nurs* 2005;30:248–56.
51. Slade SC, Molloy E, Keating JL. 'Listen to me, tell me': a qualitative study of partnership in care for people with non-specific chronic low back pain. *Clin Rehabil* 2009;23:270–80.
52. Meyer T, Deck R, Raspe H. [Validity of patient self report data in rehabilitation research: identifying circumstances of the patients' completion of questionnaires]. *Rehabilitation (Stuttg)* 2006;45:118–27.
53. Verhoef MJ, Casebeer AL, Hilsden RJ. Assessing efficacy of complementary medicine: adding qualitative research methods to the "Gold Standard". *J Altern Complement Med* 2002;8:275–81.
54. Paterson C. Seeking the patient's perspective: a qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Qual Life Res* 2004;13:871–81.
55. Kost RG, Lee LM, Yessis J, *et al.* Assessing participant-centered outcomes to improve clinical research. *N Engl J Med* 2013;369:2179–81.
56. Kost RG, Lee LM, Yessis J, *et al.* Assessing research participants' perceptions of their clinical research experiences. *Clin Transl Sci* 2011;4:403–13.
57. Yessis JL, Kost RG, Lee LM, *et al.* Development of a research participants' perception survey to improve clinical research. *Clin Transl Sci* 2012;5:452–60.
58. Willis GB. *Cognitive interviewing: a tool for improving questionnaire design*. Sage Publications, Inc, 2005.
59. Litcher-Kelly L, Martino SA, Broderick JE, *et al.* A systematic review of measures used to assess chronic musculoskeletal pain in clinical and randomized controlled clinical trials. *J Pain* 2007;8:906–13.