Original article

Designing an effective questionnaire in wound care

Keywords analysis, measures, questionnaire design, response rate, validation

For referencing Stephenson J. Designing an effective questionnaire in wound care . WCET® Journal 2022;42(2):24-29

DOI https://doi.org/10.33235/wcet.42.2.24-29

Submitted 29 April 2022, Accepted 6 May 2022

INTRODUCTION

Quantitative data collection via questionnaire is common practice in wound care. Questionnaires are a relatively inexpensive and quick way of amassing data, and do not necessarily require the researcher to be present while the data is being collected. Very often they are the only viable way to collect the data required. Common uses of questionnaires in wound care, which can include questionnaires administered to clinical staff, patients or both, include:

- To assess the effectiveness of a clinical training programme in increasing staff knowledge of a certain condition.
- To assess the extent of the use of particular dressing in a certain clinical setting.
- To evaluate a new piece of equipment.
- To monitor wound healing under a new treatment regime.
- To assess a patient-related outcome, such as pain, quality of life or satisfaction with treatment received.

While many fully validated questionnaires are available 'off-the peg', researchers in wound care may find that the specific measures captured by these questionnaires do not match the aims of their proposed study, and hence it may be necessary for a bespoke instrument to be designed. Questionnaire-based research involves careful thought regarding selection of the study sample, maximising the response rate, identifying the measures to be assessed, formulating and scoring the constituent items, framing the items for analysis, considering the outcome measures and item scoring, and piloting the questionnaire.

WHO IS THE QUESTIONNAIRE TO BE GIVEN TO?

The concept of generalisability – the ability to infer beyond sample data (those who have completed the questionnaire) to a typically much wider parent population – is key to most quantitative research studies. This requires a representative sample of respondents. It is almost impossible to create a sample which exactly reflects the population it is supposed to represent on all aspects. Clinical knowledge is needed to establish important traits – such as job level, patient co-morbidity, or wound type – which will vary from one study to another. Determination of whether a sample does indeed reflect the parent population on the characteristics deemed to be most important to the study may require knowledge of at least the approximate distribution of categories of units in the population of interest: for example, the composition of a typical tissue viability nursing team in a typical organisation may be known, and researchers may seek to reflect that composition in the personnel invited to complete our questionnaire. Failure to ensure that the sample does not differ in some important way from the population it purports to represent may lead to selection bias, which may weaken or invalidate findings.

Some specific features apply to data collected in many wound care studies. First, data must often be collected concurrently on both clinical staff and patients. An example might be a study of the caseload of a community nursing team in which both nurses and their patients will be surveyed; typically, different sets of questionnaire items will be applicable to the nurses and the patients. This often leads to clustered data, where one staff member will be treating several patients. Second, the unit of analysis in wound care studies is not always an individual person, as is often the case in other branches of clinical sciences. It may be a wound, such as a pressure injury, and one patient may supply multiple wounds to the same study. Again, this leads to the issue of clustering of data; here with pressure injuries clustered within individual patients.

MAXIMISING THE RESPONSE RATE

Data collection via questionnaire is particularly susceptible to *response bias,* bias introduced by differences in characteristics between those who choose to complete the questionnaire and those who do not. Although computational methods exist for imputing missing data values, these methods may not be viable in all situations and it is generally preferable to maximise both the proportion of potential responders who actually respond, and the proportion of those who respond who give a complete set of responses. Low response rates also lead to reductions in the power of the analysis – the ability to detect any effect that may exist.

John Stephenson
PHD FRSS(GradStat) CMath(MIMA)
Senior Lecturer in Biomedical Statistics
University of Huddersfield, United Kingdom
Email J.Stephenson@hud.ac.uk

There are some obvious methods of increasing response and completion rates:

- Use of electronic formats instead of, or as well as, paper-based questionnaires (polite emailed reminders may be sent to non-respondents at appropriate intervals).
- Avoidance of questionnaires with excessive items. All included items should be included for a specific purpose: each superfluous item increases the chance that a respondent will not complete the questionnaire properly. For example, respondents should not be asked to directly provide information on quantities such as BMI which can be calculated by the researchers from other information provided by respondents.
- Avoidance of ambiguously worded items. Items should be quick for the respondents to answer by offering a selection of options or visual analogue scales rather than asking for free text. Provision of conditional items can introduce confusion and should be limited.
- Assurance of participant anonymity, if this is appropriate for the information collected.

Some studies will require questionnaire-based data to be collected on multiple occasions, for example, to monitor quality of life or pain in patients with chronic wounds. A common issue here is that the proportion of completed questionnaires generally decreases at each data collection point. This can introduce further bias in the form of attrition bias, when those lost to follow-up are somehow systematically different from those who return their questionnaires. While little can be done about patients moving away or dying during the follow-up period, attrition loss can nonetheless be minimised by not over-burdening respondents in terms of the frequency of questionnaire mailings, nor the length or complexity of the questionnaires they are required to complete.

VALIDATION / MEASURES TO BE ASSESSED

Devising appropriate items to efficiently encapsulate outcome measures of interest is often the most difficult part of effective questionnaire design. It is generally preferable to use a questionnaire that is validated for implementation on similar participants. However, full validation is an extensive process: Price and Harding[1] reported the development and validation of a questionnaire to measure the impact of chronic wounds (leg ulcers and diabetic foot ulcers) on patient health-related quality of life (HRQoL) and identify areas of patient concern. This involved a three-stage process: a focus group and a series of semi-structured interviews to generate items for the questionnaire; a pilot process of the questionnaire with analysis of data via factor analysis; and assessment of reliability, validity and reproducibility of the resulting scale in a 3-month follow-up period.

While full validation of a self-designed questionnaire is a significant undertaking that may not be within the resources of a clinician who needs to design, implement and analyse data in a limited period of time, some common validation steps may be plausible. Often this will involve input to item wording from a panel of expert clinicians, with clarity of wording possibly assessed via focus groups or other means. The aim is to derive a series of items which each contribute to a different facet of the outcome of interest and, when assessed in conjunction with each other, provide a meaningful measure of the overall outcome. Expert advice may be needed to confirm that an item really is contributing to the measurement of the construct intended, and not some other construct. Barakat-Johnson et al[2] developed and evaluated the psychometric properties of an instrument used to assess clinician knowledge of incontinence-associated dermatitis with item development using the input of an expert panel of clinicians as the first stage of a three-stage process; this was then followed by an evaluation of content validity of the instrument via a survey of clinicians and stakeholders, and a pilot multi-site cross-sectional survey design to determine composite reliability.

Content and construct validity should also be addressed during the development process. Items that are too self-similar should be avoided. Rather than each capturing a unique facet of the construct of interest, such items are capturing the same facet, and hence this facet is being double counted, and it is very likely that respondents will respond in the same way to both items. Conversely, however, items which are very different from each other may not be measuring the same construct at all. Another common issue is the 'overlapping' of facets of a construct captured by different items. Evaluation of content and construct validity using recognised summary measures and statistical methods were utilised by Barakat-Johnson et al.[2] in subsequent stages of the development of their tool.

ITEM FORMULATION AND SCORING

Derivation of quantitative data via questionnaire requires 'closed' responses (numbers or categories); 'open-ended' responses are not generally suitable for quantitative reporting. Closed-form questionnaire items may be formulated in a number of ways. Some of the more common item formulations are:

- Items eliciting a numerical quantity directly, such as '*What is your age in years?*'
- Items which yield a numerical quantity indirectly, by requesting respondents to provide a response on a visual analogue scale which is subsequently processed by the researcher. A typical example might be to present a line of given length (say 10cm) with both ends clearly labelled as representing extreme values; for example: '*No pain at all*' and '*The worst pain imaginable*'; and accompanied by an instruction such as '*Please put a mark on this line corresponding to the level of pain your wound is causing you today*'.
- Items allowing respondents to choose one option from a list of possible options offered.
- Items allowing respondents to choose as many options as are applicable from a list of possible options offered.

The first two of these types elicit numerical responses; the second two elicit categorical responses. Both types of responses may be potentially of use for subsequent analysis, and the questionnaire should be formatted so that it is possible for respondents to report either a numerical response, or choose from a list of options, as appropriate, to a particular item.

Items eliciting direct or indirect numerical responses are potentially the most straightforward to include in subsequent analysis procedures. However, subsequent data pre-processing can be made easier by framing a question such that respondents do not feel the need to add in unnecessary words: a question such as '*How long have you worked in this organisation?*' may elicit a range of responses such as '*Less than 1 year*'; '*18 months*'; '*About 5 years*' and so forth, which will be interpreted by most computer software as text, rather than numerical responses, and need extensive editing before they can be used for analysis. A simple re-wording such as '*Please state the number of years (round to the nearest year) that you have worked for this organisation*' might save a lot of pre-processing time. Also, a simple instruction to leave blank any non-applicable items, or items for which the respondent cannot give a correct response, may save more time in deleting various instances of '*Not applicable*'; '*Don't know*'; '*Not sure*' and so forth.

It is common practice to introduce artificial categorisation in items yielding numerical data. For example, an item requesting respondents to report their age might offer a choice of age range options: *'18–30', '31–40', '41–50'* etc. Such approaches are not generally recommended: first, information is lost about the distinction between respondents of different ages within the same age range (there may be considerable differences in the responses of an 18-year-old and those of a 30-year-old); and second, multiple categories in a grouping variable means multiple comparisons are needed in the analysis (outcomes in those aged 18–30 versus those aged 31–40, outcomes in those aged 18–30 versus those aged 41–50 and so on), potentially leading to technical issues and problems of interpretation.

However, for items which capture a construct truly measured at the categorical level, there is no alternative to offering a list of options for respondents to select. The list of options offered should be exhaustive. A respondent who is requested to supply their role in an organisation, for example, only to find that their role is not represented in the options offered, may lose confidence that their participation in the study will result in accurate recording of their views or situation and may be less inclined to complete the rest of the questionnaire accurately.

A similar issue arises when options overlap. If the options for the item '*How many patients are in your weekly caseload?*' are, say, '*10 or fewer*'; '*10–20*'; '*20–30*' etc., then someone with a caseload of 10 or 20 patients exactly will not know which option they should select. Another example might be a respondent who is asked to select their job role from a list of options when they actually have two or more roles. This situation can be simply avoided with better item wording, for example: '*Please select the role from the following list that most closely corresponds to your main job role*'.

In formulating items of this kind, it can be tempting to allow respondents a free text response. This may prevent accidental omission of a respondent's preferred option, or confusion arising from multiple options which are similar, but not identical, to the response that the respondent would prefer to make. However, this allowance may necessitate extensive subsequent pre-processing of free text data into defined groups, which may not always be easy if respondents are not sufficiently explicit in their free-text responses. This situation can often be avoided by offering an '*Other*' option in the list of options.

The options offered to a categorical item may be nominal (no underlying ordering; in which case the ordering of options is unimportant) or ordinal (in which case options should be presented in a logical order). The 'classic' ordinal questionnaire item is the Likert item, the simplest and, by some margin, the most popular formulation for questionnaire items, found in many, if not most, questionnaires. A Likert item is a question which typically asks respondents to choose an option from an ordered list of five options representing the strength of agreement with a particular statement, such as, for example, '*Product X is an effective treatment for over-granulation*'. Typical options to such an item might be '*Strongly disagree', 'Disagree', 'Neither agree nor disagree', 'Agree'* and '*Strongly agree'*. Other Likert items may ask respondents to assess the frequency or magnitude of an event, such as, for example, '*Has the area around the wound become swollen?*' Here, typical options might be '*Not at all', 'A little bit', 'A moderate amount', 'Quite a lot', 'A great deal*'.

Likert items do not have to offer five options, but in general do offer an odd-number of options, of which five is probably the most common number, to allow for a 'neutral' middle option. While items with larger number of options may appear to offer more granularity of response, the distinctions between the points on the scale can be increasingly hard for respondents to discern ('*Some of the time*', *'Much of the time'*, *'Most of the time'*, *'Almost all the time'* etc.). A visual equivalent of the Likert item is a question worded something like: *'On a scale of 0 to 10, how much has your wound prevented you from carrying out daily household tasks?'*. This is an 11-point item: a common error is to allow the scale in questions of this kind to run from 1 to 10 (rather than 0 to 10). The neutral response in such cases would be represented by a response of 5.5, not 5; although many who respond with the value 5 to items of this kind would no doubt be intending to report a response in the exact centre of the available scale. Items with a wide set of ordinal responses behave in some ways like items yielding numerical responses indirectly via a visual analogue scale.

Items that request respondents to select 'as many options are applicable' are acceptable, but such items can be significantly harder to analyse than corresponding items which request

only a single option to be chosen. For example, an item such as '*Which of the following wound dressings do you use on a regular basis – please select all that apply*' followed by a list of 26 options (Product A, Product B, Product C … Product Z), is actually equivalent, in analysis terms, to a series of 26 questions: '*Do you use Wound Dressing Product A on a regular basis – yes or no?*'; '*Do you use Wound Dressing Product B on a regular basis – yes or no?*'… '*Do you use Wound Dressing Product Z on a regular basis – yes or no?*'. This series of items will probably lead to a wide range of combinations of responses and give rise to dozens of pairwise comparisons, all of which will be difficult to interpret.

FRAMING THE ITEMS FOR ANALYSIS

A typical questionnaire may begin with some basic demographic questions, eliciting respondents' demographic and lifestyle attributes, such as age, sex, family status etc.; and/or items relating to their health condition (presence of various mental or physical health conditions, duration of pre-existing wound) or employment status (length of service, staff grade etc.). Some of these items may be included to help illustrate the diversity or characteristics of the sample but will take no further part in the analysis itself.

Within reason, items measuring such 'background variables', which are typically factual questions eliciting numerical or categorical responses, rather than from Likert-style or similar items, can be recorded in whatever way is desired. Questionnaires which are designed to present data descriptively, but will not involve any kind of inferential analysis (i.e. inferring from sample data to a parent population), may be limited to items of this kind. Such studies are typically designed to assess the prevalence or proportion of a quantity, such as a study to ascertain the proportion of nurses using a particular wound care product, or the proportion of clinical staff who respond to a visual prompt such as skin reddening. Brown and Sneddon[3] implemented a questionnaire, comprised of mostly 'stand-alone' items with ordinal responses, to understand how lymphoedema services are funded and delivered across the UK and their level of resource. The questionnaire data yielded estimates of proportions (for example, the proportion of clinicians surveyed who treated open wounds) but the researchers did not attempt to generalise beyond the sample data.

However, inferential analysis is generally within the scope of most quantitative studies, and hence most questionnaires eliciting quantitative data will include items which are needed for subsequent inferential analysis. For example, with respect to a certain outcome or outcomes, it may be desired to compare experienced and novice staff, or ICU patients who are turned regularly and those who are not, or a new piece of equipment and standard equipment. These analyses are examples of *comparative studies*, in which two or more groups are compared against each other: many standard research study designs, such as cohort studies, case-control studies and randomised controlled designs, fall into this bracket. Ousey et al[4] used questionnaire-based data to compare a novel design of mattress against a standard mattress on a range of patient experience metrics (comfort, temperature and sleep quality) of patients. The researchers used standard inferential statistical methods to compare the significance and magnitude of effects, with groups defined by mattress type.

Items used to define grouping variables in these studies are categorical. Categorical variables which can take one of only two categories (or 'levels', as they are sometimes known) are known as binary variables, as in the study of Ousey et al.[4]. Some grouping variables may comprise more than two categories. For example, a study comparing outcomes in patients who may be classified as being underweight, normal weight, overweight, having obesity or having morbid obesity, might use a grouping variable '*Obesity status*' to classify each questionnaire respondent into one of the above five categories.

Such multi-categorical grouping variables should be specified with caution; while a binary grouping variable leads to a single analysis (for example, outcome in males versus outcome in females), the number of analyses required quickly increases with the introduction of multiple-level grouping variables. Another reason to limit multiple-level grouping variables is that although items recording grouping variables should, in general, allow respondent selection of any possible item, researchers should be prepared for the eventuality of thinly-spread data across multiple categories, leading to some groups which are really too small to meaningfully analyse. In such circumstances, it may be necessary to merge certain categories together before analysis.

OUTCOME MEASURES

In most questionnaires, the majority of items relate to the elicitation of outcome measures. Many outcomes are categorical, often binary, for example, the probability of a wound proceeding to 50% healing by 30 days after treatment; or multi-categorical, for example, predominant tissue type in wound bed. Such outcomes can generally be easily captured in a questionnaire with a single binary or ordinal item. Dhoonmoon[5] surveyed the experience of 56 healthcare professionals (HCPs) of the use of a debridement pad via a feedback questionnaire. Most items, including those related to pad performance (removing slough debris, debridement action etc.) were assessed using categorical items, with options from *'excellent'* to *'poor'*. Such measures lend themselves naturally to ordinal categorical assessment. For ease of analysis or other purpose, many ordinal outcomes are dichotomised – for example, one of the measured outcomes in the Ousey et al[4] study (sleep quality) was processed for analysis from its original five options (*'excellent', 'very good', 'good', 'adequate', 'poor'*) into a dichotomous measure comparing the responses of *'excellent'* or *'very good'* with any other response. Numerical outcomes, such as the percentage of patients healed, or the time for pain levels to reach a certain pre-specified value, may also be found but are less common in questionnaire-based analysis in wound care.

ITEM SCORING

Questionnaires are typically used to evaluate quantities for which no simple objective measure exists. In the context of a wound care study, these may be, for example, a clinician's evaluation of a new pressure re-distributing mattress, or a patient's opinion as to how much their wound prevents them from carrying out everyday tasks. Such quantities typically cannot be encapsulated within a single item; a series of items, all of which tap into the construct of interest, may be needed. Examples include the knowledge of dermatitis of a trainee nurse who has recently completed a workshop session on this subject, or the quality of life experienced by a patient living with a chronic wound. Typically, these constituent items may be Likert-style or similar. In such cases, interest is almost invariably centred on the processed score of a set of items, and not on any of the individual items themselves. Hence while, in theory, each item on a questionnaire item could represent a single measure, the number of distinct measures captured on a typical questionnaire is usually a lot less than the number of items in the questionnaire, with several items contributing to the evaluation of each construct.

Limitation of the number of outcomes is generally desirable: extensive presentation of results of individual outcomes in the form of, for example, pie charts may give little insight into the relative importance of the various findings. There are also certain analysis issues which may make large numbers of primary outcomes undesirable. Just like studies which collect data through other means, the ideal questionnaire probably captures information on a single, pre-specified primary outcome, and a small number of secondary outcomes.

A score is needed for all items which contribute to the evaluation of a particular measure. Typically, the scoring for 5-point Likert items is very simple – from 1 point for '*Strongly disagree*' to 5 points for '*Strongly agree*', with intermediate options scored accordingly. Likert items with other numbers of options are scored in a similar way. Many researchers prefer to use a coding such as: –2 points for '*Strongly disagree*', –1 point for *Disagree* and so on up to +2 points for '*Strongly agree*', possibly with the idea that negatively worded responses require negative scores. This coding is exactly equivalent to the 1–5 coding mentioned above – the score for each option is reduced by 3 points for all options. As long as this scoring is applied consistently, inferences will be the same under either scoring systems.

It is normally assumed that item scores are additive, that it is meaningful to derive an overall score by adding up the scores obtained on individual items which contribute to the same measure. This assumption is often easier to justify if there is consistency in the formulation of items. It not obvious how an overall score should be derived with a series of items with a number of options that varies from, say, 2 to 3 to 5 to 7. Scores from the items with the largest number of options will swamp those from items with fewer responses if, for each item, responses are simply coded as 1 up to the value of the number of the options.

It is also harder to justify that summing scores from multiple items leads to a meaningful measure, even if the number of options in each item is the same, if the options are different. If one set of items offers the options '*Strongly disagree', 'Disagree…' 'Strongly agree*' and another set offers the options *'Not at all', 'A little bit…' 'A great deal'*, it may be difficult to argue that the scores from the two sets of items can be meaningfully combined.

To ensure a meaningful total, the above coding may need to be reversed if some items are in the opposite sense to others, for example, if 5-point Likert items such as '*My wound has forced me to limit my activities with others*' and '*The wound has affected my sleep*' are coded using the 1–5 scale above, with 1 point awarded for a response of '*Strongly disagree*' and 5 points awarded for a response of '*Strongly agree*', then the implication is that higher scores indicate worse outcomes. Hence if an additional item in the same scale such as, for example, '*I am able to carry out everyday tasks without difficulty*' is to be included, this item could be coded such that '*Strongly agree*' is awarded 5 points, '*Strongly disagree*' 1 point, and other points of the scale scored accordingly, for consistency with the remaining scale items.

PILOTING THE QUESTIONNAIRE

Pilot implementation can be a useful tool in the refinement of questionnaire items and can reveal issues which may impact on subsequent response rate and response reliability such as poor clarity of item wording or excessive time taken for questionnaire completion. If a questionnaire includes a set of Likert-style or similar items which are designed to tap into the same construct, the internal consistency of the pilot responses to these items can be assessed easily and quickly using the most statistical software. This process can identify items which are not responded to in a similar manner to other items purporting to be measuring the same construct, and hence may require amendments to their wording (if the wording is unclear or has been misunderstood by respondents), deletion from the questionnaire, or possibly moving to the measurement of another construct. The pilot stage is generally the only opportunity to make such amendments if they are needed.

SUMMARY

Good questionnaire design is driven by the research question, and the analysis that proceeds from it. Consideration of the end point is in fact generally the starting point. Issues to be considered include determination of the outcomes to be measured; how are they to be measured; whether outcomes are objective measures that can be adequately captured using items eliciting simple numerical responses or categories, or require multiple items to capture a series of specific facets of the measure.

The level(s) at which the analysis is to be conducted must also be determined – in wound care studies, analyses at the patient, clinician or wound level are all commonplace. It must

also be determined whether or not outcomes are to be linked to any other variables, and whether the desired groups for comparison are featured in the items functioning as grouping variables to classify units of analysis (whether patients, clinicians or wounds) appropriately.

Data collection via questionnaire should be approached just as data collection via medical devices or other means – it is necessary to ensure that the data collection instrument is fit for purpose. This means that as many steps as possible along the validation road are taken (assuming that a pre-validated instrument is not being used) to ensure that we are measuring the outcomes we think we are measuring, via carefully worded items grouped and scored appropriately. Care should be taken that only as many items as are necessary are used to capture demographics, other background information and outcome measures. It is necessary to ensure that respondents are, as far as possible, a representative sample of the population to which generalisations are to be made. Response rates are maximised by making the items as clear as possible, and by asking as little as possible of respondents in terms of the length of time and the amount of effort they will need to complete the questionnaire, just as might be done using other means of data collection.

While it is easy to under-estimate the effort required to facilitate effective questionnaire-based data collection, when conducted properly, questionnaire-based data collection can be a highly effective means of data collection and form a sound base for research studies.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

FUNDING

The authors received no funding for this study.

REFERENCES

1. Price P, Harding K. Cardiff Wound Impact Schedule: the development of a condition-specific questionnaire to assess health-related quality of life in patients with chronic wounds of the lower limb. Int Wound J. 2004 Apr;1(1):10-17.
2. Barakat-Johnson M, Beeckman D, Campbell J, Dunk AM, Lai M, Stephenson J, Coyer F. Development and Psychometric Testing of a Knowledge Instrument on Incontinence-Associated Dermatitis for Clinicians: The Know-IAD. J Wound Ostomy Continence Nurs. 2022 Jan-Feb 01;49(1):70-77.
3. Brown L, Sneddon MC. Lymphoedema service provision across the UK: a national survey. J Lymphoedema. 2020;15(1):16-21.
4. Ousey K, Stephenson J, Fleming L. Evaluating the Trezzo range of static foam surfaces: results of a comparative study. Wounds UK 2016;12(4):66-73.
5. Dhoonmoon L. Experiences of healthcare professionals using Prontosan® debridement pad. Wounds UK 2021;17(1):118-123.