Designing an effective questionnaire in wound care

KEY WORDS
- Complainer
- Difficult people
- Emotional intelligence
- Hostile
- Negativist
- Silent type

Designing a questionnaire for effective quantitative data analysis in wound care requires careful thought. We are usually trying to 'measure the unmeasurable' i.e. capture one or more constructs for which no objective measure exists. Hence, we need to ensure that the items we include to measure those constructs, which can elicit either numerical or categorical responses, adequately capture the essence of the constructs without duplication, omission of some important aspect of the construct or inclusion of irrelevant items. We need a scoring system that is easy to manage and results in meaningful measures. If inferential analysis is to be conducted from our questionnaire data, this needs particular thought: for example, we may need to collect demographic or job-related information recorded in suitable categories to facilitate comparative analysis of construct scores. Pragmatically, we should also be concerned with the length of the questionnaire, the clarity of the items, be clear what is the unit of analysis, who the questionnaire is designed for and the population to which any generalisation is to be made.

JOHN STEPHENSON
Senior Lecturer in Biomedical Statistics, School of Human and Health Sciences, University of Huddersfield, Queensgate, Huddersfield, UK.

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 we need. Common uses of questionnaires in wound care, which can be 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.

Unfortunately, writing a questionnaire that will give analysable data in the required form is significantly harder than many people realise. The fundamental issue, in my opinion, is that questionnaires are often trying to measure the unmeasurable. We would never use a questionnaire to estimate someone's height ('How strongly would you agree with the statements that you get backache when sitting in cinema seats for long periods, or that you have trouble reaching the highest shelf in the cupboard?') because a simple objective measure exists to measure height. But there is no single objective way to measure many of the objectives in a typical wound care study, such as 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: we may need a series of items, all of which, hopefully tap into the construct of interest.

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.

Derivation of quantitative data via questionnaire requires "closed" responses (numbers or categories); "open-ended" responses are not generally suitable for quantitative reporting.

Many people (including me!) use the terms "survey" and "questionnaire" interchangeably, and although strictly speaking they are not synonymous terms, for most of us, this usage leads to few issues. However, to be precise, a questionnaire is simply a list of questions, while a survey is broader in definition and involves the entire data collection process.

Who is the questionnaire to be given to?

Assuming you are using a questionnaire to conduct quantitative research, the concept of generalisability will be key to your study — the ability to infer beyond your sample data (those who have completed the questionnaire) to a typically much wider parent population. 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 — we need to use our clinical knowledge to decide which are the important traits — such as job level, patient comorbidity, or wound type, which will vary from one study to another. Determination of whether our 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, we may know the composition of a typical tissue viability nursing team in a typical NHS Trust, and seek to reflect that composition in the personnel we invite 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 our findings.

There are a couple of unusual features that apply to data collected in many wound care studies. First, we often need to collect data 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 the issue of nested or 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 ulcer (PU), and one patient may supply multiple wounds to the same study. Again, this leads to the issue of clustering of data, here with PU clustered within individual patients.

Maximising the response rate

Data collection via questionnaire is particularly susceptible to response bias. This is 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, our ability to detect any effect that may exist.

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

- Use electronic formats instead of, or as well as, paper-based questionnaires (send polite emailed reminders to non-respondents at appropriate intervals)
- Don't include too many items in the questionnaire. All included items should be included for a specific purpose: each superfluous item increases the chance that a respondent will not bother to complete the questionnaire properly

- Limit "conditional" items to avoid confusion ('if you answered 'yes' to Item 17, go to Item 24 and answer items 24–29; if you answered 'no' to item 17, go to item 18 and answer items 18–23, unless you also answered "maybe" to item 12, in which case skip item 21 altogether')
- Avoid ambiguously worded items; make the items quick for the respondents to answer by offering a selection of options or visual analogue scales rather than asking for free text
- Do not ask respondents for responses that may be difficult for them to assess, particularly if you can derive the required information from other responses: for example, if you have collected data on respondents' height and weight, but wish to know their body mass index (BMI), you need not ask this question directly but can calculate it yourself as part of the pre-processing of data subsequent to data collection. This also rules out the possibility of respondents giving responses which are not self-consistent!
- Assure your participants of anonymity, if this is appropriate for the information you are collecting.

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 gets gradually lower 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 we can do little about patients moving away or dying during the follow-up period, we can nonetheless minimise attrition loss by not over-burdening our respondents in terms of the frequency of questionnaire mailings, or the length or complexity of the questionnaires we ask them to complete.

Item formulation

Closed-form questionnaire items used for quantitative analysis 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 request respondents to provide a numerical quantity indirectly, via a visual analogue scale, which is subsequently processed by the researcher into the required value. A typical example might be to present a line of given length (say 10 cm) 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 Trust?' 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 Trust' 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.

From an analysis point of view, items formulated to elicit numerical responses are

generally preferable to alternative formulations eliciting categorical responses. I have seen many questionnaires asking respondents to select the age range (for example, 18–30, 31–40, 41–50 etc.) corresponding to their actual age. I don't recommend this practice: first, we lose information 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). 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 a Trust, 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: however, if subsequent analysis reveals 'Other' to be the most frequent response to a particular item, that might indicate that not enough consideration was given to the range of the options offered for that item.

Items that request respondents to select 'as many options are applicable' are acceptable, but you should be aware that these 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.

The 'classic' questionnaire item is the Likert item, sometimes referred to as a Likert-style item. Many, if not most, questionnaires include a series of these items. A Likert item is a question which typically asks respondents to choose an option from an ordered list of 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 5 options as in the example above, but in general do offer an

odd-number of options, of which 5 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?'. 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 people 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.

Likert-style items are the simplest and, by some margin, the most popular formulation for questionnaire items, and while other methods of formulating items exist, such as Thurstone and Guttman items, these will not be discussed here.

Item scoring

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. Actually, 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 system.

We normally assume 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 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.

Implications for analysis

A typical questionnaire may begin with some basic demographic questions, eliciting respondents' demographic and lifestyle attributes, such as age, sex, BMI 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 our 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 we wish. Questionnaires that 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.

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, we may wish 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 we compare two or more groups against each other. Many standard research study designs, such as cohort studies, case-control studies and randomised controlled designs, fall into this bracket.

A comparative study requires a grouping variable, one which defines the groups to be compared. This would normally be determined via a single item on a questionnaire, of the type allowing respondents to choose one option from a list of possible options. In general, items allowing respondents to select multiple options are not suitable for use as grouping variables.

Grouping variables are categorical. They do not take numerical values but take a specific category. Categorical variables that can take one of only two categories (or "levels'" as they are sometimes known) are known as binary variables. A simple example is gender, the use of which as a grouping variable would lead to a comparison of outcomes in male and female patients or staff. Another example of a binary grouping variable might be "pressure ulcer status on admission", a variable which classifies ICU patients as either having, or not having, one or pressure ulcers on admission to ICU. In the context of wound care, grouping factors may be defined at many levels: at the wound level, such as pressure ulcer category or exudate level; at the patient level, such as ethnic group, sex or comorbidity; at the clinician level, such as staff role; or (rarely) at the institution level, such as setting. It is perfectly possible to analyse the concurrent effects of multiple grouping variables using questionnaire-based data.

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 5 categories (a classification that would probably be made post-data collection via calculation of BMI values from items eliciting respondents' height and weight, rather than being obtained directly from an item on the questionnaire). Such multi-categorical grouping variables should be specified with caution, while a binary variable leads to a single analysis (for example: outcome in males versus outcome in females); 3-category, 4-category and 5-category grouping variables lead to, respectively, 3, 6 and 10 possible pairwise comparisons. This is generally too many to analyse effectively. Another reason to think twice about specifying grouping variables with multiple options 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.

Less commonly, we may wish to relate numerical quantities to certain outcomes. For example, we may wish to relate wound length (in cm) at baseline, or the age of a patient, to a healing outcome such as the probability of wound closure within 30 days. In the context of a questionnaire, such numerical variables can often be easily captured and used in their raw form.

Other types of quantitative studies aim to assess the prevalence of a quantity, to a certain level of precision, without aiming to link this to any grouping variable. Examples might be 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. Questionnaires designed

for such studies will generally not need to include items designed to capture grouping variables.

Outcome measures

In most questionnaires, the majority of items relate to the elicitation of outcome measures. Outcomes can take many types:

Type 1: A categorical outcome, which could be 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 item.

Type 2: A simple numerical outcome, such as the percentage of patients healed, or the time for pain levels to reach a certain pre-specified value. Such outcomes can also be easily captured with a single item.

Type 3: An outcome that is evaluated by processing responses to a number of items; such as 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.

The number of outcome measures in a questionnaire should be limited. Most of us have had the dubious experience of wading through results presented in the form of several dozen pie charts or extensive tables which give little insight into the relative importance of the various findings. Just like studies that 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. Besides the difficulties in interpreting and summarising the findings of studies with multiple outcome studies, there are certain analysis issues which may make large numbers of primary outcomes undesirable.

Selecting items to contribute to outcome measures

Devising appropriate items to efficiently encapsulate outcome measures of interest is often the most difficult part of effective questionnaire design. Much work goes into the validation of questionnaires capturing outcome measures of this kind, and if you can find a questionnaire that measures what you need to measure and is validated for implementation on the same sort of participants that you wish to study, you should probably use it. However, if no such questionnaire is available, with careful thought it should be possible to derive a series of items that appropriately capture the constructs of interest.

While full validation of a self-designed questionnaire is a significant undertaking that is unlikely to 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 that 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. This is not at all easy to accomplish, Items that are too self-similar should be avoided, such as: "During the last 7 days, my wound left me in pain" and "During the last 7 days, I would describe my wound as being painful to touch". Essentially such items, rather than each capturing a unique facet of the construct of interest, are capturing the same facet and hence this facet is being double counted: 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. 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.

Another common issue is the "overlapping" of facets of a construct captured by different items. For example, many questionnaires use an item representing an overall, non-specific measure (for example, "I am satisfied that Product X performs well in a wound care setting") and also more specific items (for example, "I am satisfied that Product X is effective in reducing wound exudate"). In such cases, we need to decide

whether a particular item is tapping into an outcome that is distinct from other outcomes under consideration, or whether it is tapping into the same outcome as are other items and the items scores hence can be combined accordingly.

Piloting the questionnaire

Questionnaire design is hard to get right first time around. It's a good idea to pilot your questionnaire on a small group of colleagues or friends before general implementation, and to be prepared to act on any advice you receive regarding the clarity of item wording, time taken for questionnaire completion, or any other issue which may be thought to impact on subsequent response rate and response reliability. If your questionnaire includes a set of Likert-style or similar items that are designed to tap into the same construct, you can assess the internal consistency of the pilot responses to these items easily and quickly using most statistical software. This process can identify items that 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. Start with the end point, what outcomes are to be measured and how are they to be measured. Are outcome objective measures that can be adequately captured using items eliciting simple numerical responses or categories, or may they thought of as constructs which cannot be objectively measured using single items and will hence require multiple items, each of which will make a distinct contribution to the measure? At what level(s) do we intend to conduct the analysis in wound care studies, analyses at the patient-, clinician- or wound-level are all commonplace. Are outcomes to be linked to any other variables and if so, are the desired groups for comparison 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 we approach data collection via medical devices or other means. We need to ensure that the data collection instrument is fit for purpose. This means that we take as many steps as possible along the validation road (assuming we are not using an instrument that has already been validated) to ensure that we are measuring the outcomes we think we are measuring, via carefully worded items grouped and scored appropriately, and we include as many items as are necessary (but no others) to capture the demographics and other background information, and the variables that we will use to relate to our outcome measures. We ensure our respondents are, as far as possible, a representative sample of the population to that we wish to generalise. We maximise our response rate by making the items as clear as possible, and by asking as little as possible of our respondents in terms of the length of time and the amount of effort they will need to complete the questionnaire, just as we might do using other means of data collection.

Questionnaire-based methods of data collection are very common in many research fields other than wound care. They are probably most often found in the realms of social research. I always think that it is slightly ironic that such disciplines are often referred to as the 'soft' sciences (in contrast to the 'hard' sciences of clinical disciplines: questionnaire data collection is, in general, much harder to get right than other methods. Many people underestimate the effort required to facilitate effective questionnaire-based data collection, and it is certainly very easy to get it badly wrong, but when conducted properly, questionnaire-based data collection can be a highly effective means of data collection and form a sound base for research studies. WUK