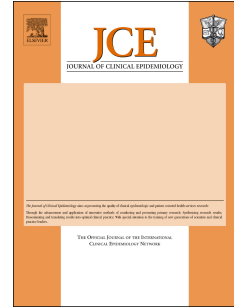


# Accepted Manuscript

Point of data saturation was assessed using resampling methods in a survey with open-ended questions

Viet-Thi Tran, Raphael Porcher, Bruno Falissard, Philippe Ravaud



PII: S0895-4356(16)30261-X

DOI: [10.1016/j.jclinepi.2016.07.014](https://doi.org/10.1016/j.jclinepi.2016.07.014)

Reference: JCE 9219

To appear in: *Journal of Clinical Epidemiology*

Received Date: 7 March 2016

Revised Date: 18 July 2016

Accepted Date: 23 July 2016

Please cite this article as: Tran V-T, Porcher R, Falissard B, Ravaud P, Point of data saturation was assessed using resampling methods in a survey with open-ended questions, *Journal of Clinical Epidemiology* (2016), doi: 10.1016/j.jclinepi.2016.07.014.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Point of data saturation was assessed using resampling methods in a survey with open-ended questions**

Viet-Thi Tran<sup>1, 2, 3</sup>, Raphael Porcher<sup>2, 3, 4</sup>, Bruno Falissard<sup>5, 6</sup>, Philippe Ravaud<sup>2, 3, 4, 7</sup>

<sup>1</sup> Department of General Medicine, Paris Diderot University, Paris, France

<sup>2</sup> Centre de recherche en Epidémiologie et Statistiques (CRESS), INSERM U1153, Paris, France

<sup>3</sup> Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, Paris, France

<sup>4</sup> Paris Descartes University, Paris, France

<sup>5</sup> Paris Sud University, Paris, France

<sup>6</sup> Centre de recherche en Epidémiologie et Santé des populations (CESP), INSERM U1018, Paris, France

<sup>7</sup> Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA

**E-mail addresses:**

Tran Viet-Thi

[thi.tran-viet@aphp.fr](mailto:thi.tran-viet@aphp.fr)

Porcher Raphael

[raphael.porcher@aphp.fr](mailto:raphael.porcher@aphp.fr)

Falissard Bruno

[bruno.falissard@gmail.com](mailto:bruno.falissard@gmail.com)

Ravaud Philippe

[philippe.ravaud@aphp.fr](mailto:philippe.ravaud@aphp.fr)

**Corresponding author:**

Viet-Thi Tran - MD, PhD

Hôpital Hôtel Dieu, Centre d'Épidémiologie Clinique, Paris, France

1 place du Parvis Notre-Dame, Paris 75181, France

Tel: +33 1 42 34 89 87, fax: +33 1 42 34 87 90, email: [thi.tran-viet@htd.aphp.fr](mailto:thi.tran-viet@htd.aphp.fr)

**Abstract (200)****Objective**

To describe methods to determine sample sizes in surveys using open-ended questions and to assess how resampling methods can be used to determine data saturation in these surveys.

**Study design**

We searched the literature for surveys with open-ended questions and assessed the methods used to determine sample size in 100 studies selected at random. Then, we used Monte Carlo simulations on data from a previous study on the burden of treatment to assess the probability of identifying new themes as a function of the number of patients recruited.

**Results**

In the literature, 85% of researchers used a convenience sample, with a median size of 167 participants [interquartile range (IQR) = 69-406]. In our simulation study, the probability of identifying at least one new theme for the next included subject was 32%, 24% and 12% after the inclusion of 30, 50 and 100 subjects, respectively. The inclusion of 150 participants at random resulted in the identification of 92% themes [IQR = 91-93%] identified in the original study.

**Conclusion**

In our study, data saturation was most certainly reached for samples > 150 participants. Our method may be used to determine when to continue the study to find new themes or stop because of futility.

## 1. Background

Surveys with open ended-questions are simple and low-cost methods to explore the diversity and the variation of a phenomenon in a given population [1]. These surveys allow recruitment of large number of geographically distant participants through postal, telephone or internet, but the rigid order and formulation of questions remove the flexibility with which questions may be phrased, elaborated, or followed up with supplementary interrogations [2]. Similarly to the other qualitative inquiry methods, the purpose of these surveys are to describe all facets of a topic of interest. It is the point of data saturation. According to Guest et al., “*Data saturation is the point in data collection and analysis when new information produces little or no change to the codebook*” (the codebook representing the collection of codes that links expressions found in text to all abstract constructs identified by the researchers) [3].

“How many participants are necessary to include to obtain data saturation” is a frequent question in qualitative research [3, 4], but difficult to answer, even for experienced researchers [5]. Indeed, a number of factors can affect the sample size needed to obtain data saturation [6]: 1) the topic of interest and study aim [7], 2) the study participants (e.g., individual factors, differences between participants) [3], 3) the existence of an established theory [6]; 3) the methods of data collection (e.g., interview techniques, experience of investigators); and 4) the methods of data analysis [8, 9]. Although predefining a number of participants to be included is counterintuitive in qualitative studies, researchers usually need to state approximate sample sizes to obtain funding [10] or to meet ethics committees’ requests [11]. In those cases, target sample size is generally determined according to studies with similar subject of interest[12], empirical recommendations and researcher’s experience [5, 10], with sample sizes around 30 participants [11].

Surveys using open-ended questions differ from interviews and focus groups because: 1) structured questionnaires restrict descriptions of the experience gained from a participant; 2)

they are usually sent to large samples of participants determined according to predefined criteria instead of recruiting participants purposefully chosen; and 3) data collection and analysis are sequential instead consisting of iterative cycles of collection/analysis. Thus, usual recommendations for sampling are not adequate and larger sample sizes are needed than when using interviews or focus groups. To our knowledge, no study has yet assessed how decisions about sample size in these surveys are reported, and there is no transparent and reproducible method to determine the point of data saturation in such surveys.

The present study had two objectives. First, we wanted to describe the methods used by researchers to ascertain the point of data saturation and determine sample sizes in surveys using open-ended questions. Second, we wanted to present a new reproducible method to estimate the point of data saturation, using resampling methods.

## **2. Methods**

Our study involved two steps: 1) we performed a literature review on sampling methods used in surveys with open-ended questions; and 2) we described how resampling methods could be used to objectify point of data saturation and illustrated our method with data collected in a survey about the burden of treatment.

### **2.1. Literature review**

In the first step of this research, we identified methods used in the literature to determine sample size in surveys with open-ended questions. For that, we 1) looked for guidelines and methodological articles on sample size in qualitative research; 2) systematically reviewed methods used by researchers in these surveys. Our search was conducted on MEDLINE via PubMed from inception to the 18<sup>th</sup> of December 2015. We identified all studies using surveys using open-ended questions with the following criteria: 1) the qualitative inquiry was the main objective (i.e. we excluded mixed methods studies); 2) they only used qualitative surveys

using open-ended questionnaires (i.e. we excluded studies using multiple data collection methods); and 3) they were published in English. Our search used the words “survey”, “questionnaire”, “open-ended”, “qualitative”, “content”, “thematic” and “grounded”.

One investigator (V-TT) assessed the sampling methods used in 100 randomly selected studies and assessed: 1) the number of participants included in the study; 2) if researchers estimated a target sample size prior to the beginning of the study and its size; 3) how participants were selected to participate in the study; and 4) if researchers claimed to have reached data saturation and how they assessed it.

## **2.2. Assessment of the point of data saturation in surveys with open-ended questions using simulation methods**

In the second step of this research, we estimated point of data saturation in surveys using open-ended questions with resampling methods. We used data from a study on the burden of treatment[13].

### **2.2.1. Materials**

In a previous study on the burden of treatment [13], we used a survey with open-ended questions to determine the causes, consequences and aggravating factors of the burden of treatment in patients with chronic conditions. English-, Spanish- and French- speaking patients with at least one chronic condition answered a series questions in an Internet survey with 1) a broad open-ended question to elicit patient views about their treatments and the associated burden globally and 2) 16 open-ended questions about different aspects of the burden of treatment identified in the literature.

Participant answers were analyzed by content analysis [14]. In a first step, two investigators independently identified for the first 200 responses in French and English, “in vivo codes”:

literal terms used by participants to describe their burden of treatment. During meetings, the investigators reached consensus on the initial codes and grouped them into an initial set of themes. In the present study, we refer as “theme” to describe ideas directly derived from “in vivo codes”. For example, “*I will have to take medication for the rest of my life, there aren’t holidays for treatment*”, was coded as the theme entitled “Treatment is for whole life”. In a second step, this initial set of themes was used for analysis of the remaining responses: each participant’s response was read by two investigators (at least one researcher native in the given language), who independently assigned data segments to each theme. During meetings, the investigators compared their analyses and reached consensus on coding. Whenever a new idea emerged, researchers discussed the idea, thereby refining and enriching the list of themes. As a result, for each participant, we were able to describe the different themes he mentioned and thus create a table opposing participants and themes elicited. All participants gave electronic informed consent before participating in the study. The study was reviewed and approved by the Institutional Review Board (IRB) of Cochin Hospital in France (no. 00001072) and the Mayo Clinic IRB (Rochester, MN, USA).

In this study on the burden of treatment, a total of 1,053 participants answered the open-ended questions from May 2013 to March 2014. The mean (SD) age was 46 (14) years. In total, 671 patients resided in France (64%), 140 in the United States (13%), 66 in Canada (6.3%), 56 in the United Kingdom (5.3%), 34 in Spain (3.2%), 30 in Australia (2.8%), and 56 (5.3%) in a different country. Self-reported main chronic conditions ranged from rheumatologic diseases (33%) and diabetes (16%) to cancers (8%). The mean (SD) number of chronic conditions was 2.4 (1.6, range 1–10). A total of 662 patients (63%) had two or more chronic conditions.

Answers to open-ended questions formed an overall corpus of 408,625 words, in English (148,707 words), French (243,558 words), and Spanish (16,360 words). Median (Q1-Q3) length of patients’ answers were 298 (129-526) words globally (maximum 2,699) and 73 (38-

133) words for the first question (maximum, 1,267). Content analysis described a list of difficulties patients could have when performing healthcare-related tasks. Examples of themes identified and associated verbatim are presented in **Appendix 1**. More details about the study methods and complete results are presented elsewhere [13].

### **2.2.2. Resampling of data**

We used the data collected during the previous study on the burden of treatment to illustrate a method to operationalize point of data saturation using Monte Carlo simulations. Resampling data allowed us to assess the cumulative number of themes identified as a function of the number of patients included in the study with confidence intervals.

Because of there is no cut-off for determining data saturation as a function of the number of themes elicited, we defined that data saturation was reached in simulations when at least 90% of all themes coded in the original study had been identified. This limit was chosen based on results from retrospective studies reporting that 90% of themes were identified with the first 12 to 30 participants [3, 8]. We also defined a “frequent theme” as a theme mentioned by more than 2.5% of patients in the original study. This limit was chosen to highlight themes mentioned by a substantial proportion of patients while keeping enough precision for thematic exploration of the concept.

Simulations were repeated 10,000 times, which ensured a precision of the estimated 95<sup>th</sup> percentile with standard error=0.002 and involved two different datasets: 1) one considering the identification of themes using the complete online survey (i.e., both the first broad open-ended question and specific questions about topics of interest identified in the literature); and 2) the other considering the identification of themes by using only the first open-ended question, afterwards defined as “spontaneous answers by participants.”



It is to be noted that results from this study rely on assumption of 2 hypotheses: 1) the original study sample was sufficiently diverse to ensure that all themes had been elicited at least once (i.e. in the original study, we achieved “true” data saturation); 2) if analysis had only been conducted using a subset of the initial corpus, we would have found the exact same themes (i.e. identification of themes was independent of the order of inclusion of participants).

Analyses were carried using the R program v3.1 (<http://www.R-project.org>, the R Foundation for Statistical Computing, Vienna, Austria).

### **2.2.3. Comparison between sampling strategies to decide when to stop data collection and which participants to include**

We used our simulations to compare different strategies to: 1) decide when to stop data collection and 2) enrich an existing sample of participants with patients presenting different characteristics.

First, we compared two strategies to decide when to stop data collection: 1) *a priori* fixed sample size or 2) use of a stopping criterion. To simulate the use of an *a priori* fixed sample size to approach data saturation, we selected random samples of 15, 30, 50, 100, 150, 200 and 250 subjects among our participants and assessed the cumulative number of themes elicited in each sample. To simulate the use of a stopping criterion to decide when to stop data collection, we reproduced a strategy described in the literature relying on 1) specification of an initial analysis sample of fixed size; 2) specification of how many more interviews are needed without new themes emerging before stopping data collection; and 3) analysis until the stopping criterion is met [15]. We simulated this strategy by randomly selecting an initial sample of subjects and then iteratively adding new participants. For each addition of participants, we assessed 1) whether the stopping criterion was met and 2) the cumulative number of themes elicited by the sample. We tested two different settings: 1) initial sample of

10 + the addition of iterative groups of 3 and 2) initial sample of 10 + the addition of iterative groups of 10.

Second, we assessed whether enriching an initial sample with patients presenting different characteristics would improve the number of themes identified. For this, we randomly selected an initial sample and iteratively enlarged it with patients differing by country of residence (addition of 5 French vs. 5 non-French participants) or conditions (addition of 5 patients with the least represented conditions vs. 5 random participants). These criteria were chosen because they were considered the most meaningful for diversification of the sample in the original study.

### 3. Results

#### 3.1. Results from the literature review

In the literature, we found no specific guidance on sample sizes for surveys with open-ended questions in textbooks and/or methodological articles. Thus, we systematically reviewed the methods used to determine sample size in research articles. Our literature search identified 1410 references, among which 437 corresponded to studies involving surveys with open-ended questions as the main way to collect data (**Figure 1**). These studies tackled a variety of topics ranging from clinical topics (n= 40 studies, 40%) (Investigating how patients perceived fibromyalgia flares and alleviating factors associated with flares [16]) to public health topics (n=25, 25%) (Trialists' opinions on the causes and methods to prevent publication bias [17]). In all selected studies, data collection and analysis were sequential. Median sample size was 167 participants (IQR= 69 – 406). Surveys used a median of 3 open-ended questions (IQR 1-5). Characteristics of included studies are presented in the **Appendix 2**.

Fifteen studies (15%, Confidence Interval (CI) [8.6-23]) determined a target sample size prior to the beginning of the study but all sample sizes were arbitrarily determined. For example, in

a study exploring the determinants of physicians' empathy [18], researchers stated that: "*We assumed that more than four respondents per specialty would constitute a sufficient sample size for our exploratory study*". Target sample sizes varied from 60 [19] to 500 [20] participants.

Six (6%, CI [2-13]) studies reported to have reached data saturation but only two reported the number of participants' answers analyzed before reaching data saturation. In the first study, about cancer patients' views of the constituents of quality of life, data saturation was met after analysis of 75 answers (out of 248 patients recruited) [21]. In the second, about fathers' experiences on breastfeeding challenges, data saturation was reached after analysis of 15 participants (out of 117) [22].

### **3.2. Results from the simulation study**

In the second part of this study, we used simulation methods on data from a previous study on the burden of treatment to determine the point of data saturation. In the original study on the burden of treatment, we identified 123 different themes from participants. All themes had been mentioned at least once after the 681<sup>th</sup> patient had been included. Data saturation, that is, the identification of 111 (90%) themes, was achieved after inclusion of the 192<sup>th</sup> subject. Of all 123 themes, 88 (71.5%) were elicited by more than 2.5% of participants and were considered "frequent"; all frequent themes had been mentioned at least once after inclusion of the 149<sup>th</sup> participant.

When considering only the first open-ended question, 118 themes (96%) had been elicited. Data saturation occurred only after inclusion of the 976<sup>th</sup> participant. All frequent themes had been elicited at least once in the first open-ended question after inclusion of the 447<sup>th</sup> participant (**Appendix 3**). The speed of discovery of new themes did not differ by sex, level of education, or conditions (**Appendix 4**).

### **3.2.1. Probability of identifying a new theme or missing an important theme as a function of the number of patients already included in the study**

The median probability of identifying at least one new theme for the next included subject as a function of the number of participants decreased quickly and was 32%, 24% and 12% after the inclusion of 30, 50 and 100 subjects, respectively. Comparatively, the median probability of missing at least one frequent theme was 17%, 6% and 1% after the inclusion of 30, 50 and 100 subjects, respectively (**Figure 2 and Appendix 5**). When considering only the first open-ended question, these probabilities increased to 50%, 29% and 16% (**Figure 2 and Appendix 5**).

### **3.2.2. Comparison between sampling strategies to decide when to stop data collection and which participants to include**

After the inclusion of 150 participants, we achieved data saturation, with a median proportion of 92% identified themes [interquartile range (IQR) was 91-93%]. After inclusion of the 100<sup>th</sup> participant, we identified a median of 99% [IQR= 99-100%] frequent themes. When considering only the first open-ended question, we were not able to achieve data saturation with 250 participants [IQR=83-86%]. After the inclusion of the 150<sup>th</sup> participant, a median of 92% [IQR=89-94%] frequent themes had been spontaneously elicited at least once (**Figure 3**).

The strategy with specification of an initial sample and stopping criterion did not allow for achieving data saturation for the two settings we predefined. Specification of an initial sample of 10 patients and the iterative addition of 3 participants resulted in the inclusion of a median of 37 [IQR =28-43] patients, eliciting a median of 72% [IQR=66%-77%] themes until stopping criterion was met. When considering a stopping criterion of 10 participants without any new theme elicited, we obtained a median sample size of 100 [IQR=90-120] patients,

eliciting a median of 89% [IQR=85%-91%] themes until the stopping criterion was met (**Table 2 and figure 3**).

Enrichment of an initial sample with participants from different countries allowed identification of more themes for initial samples <100 patients. However, increase was very small with, at most, a mean of 0.66 additional themes identified by adding 5 non-French patients to an initial sample of 50 French participants (**Figure 4**). Enlarging the sample with iterative addition of 5 participants with conditions less represented in the sample did not improve the identification of new themes as compared with adding random participants, whatever the initial sample of patients or study design (i.e., one open-ended question only or one broad question + specific questions on predefined topics) (**Figure 4**).

#### **4. Discussion**

The concept of data saturation is easy to understand but difficult to ascertain in practice because it depends on the topic, purpose of the research, participants, way data is collected (e.g. interviewer characteristics, method of data collection, context, etc.) and analysis [4].

In this study, we first reviewed the literature for methods to determine point of data saturation in surveys with open-ended questions. We found that approximately 95% of studies did not report to have assessed or reached data saturation neither as how their sample sizes were determined. Second, we showed that assessment of the point of data saturation is possible in surveys using open-ended questions, using simulation methods. We re-sampled 10,000 times the 1,053 participants from a single Web-based study and found that, even in extreme cases, recruitment of samples > 150 participants would not likely expand significantly the number of themes identified.

Results from our literature review concur with those from other methodological reviews of sampling methods in studies focused on achieving data saturation. In focus group studies,

authors found that only 17% focus group studies attempted to explain how the number of participants included had been determined [23]. Similarly, in a review of 83 interview studies in Information System research, authors found that many researchers invoked point of data saturation but none explained how it had been assessed [24]. This opacity in the methods used to assess data saturation impacts the credibility of research as it is not possible to ascertain if some researchers did not invoke the criterion of saturation to justify small samples with thin data, in fact resulting from lack of time or funds [23]. In this article, we provide a reproducible and objective method, using resampling techniques to objectify that point of data saturation has been reached [6, 25, 26].

When illustrating the use of our method with data collected in a survey with open-ended questions on the burden of treatment, we found that point of data saturation was attained after inclusion of 150 participants. This figure contrasts with what is usually recommended in textbooks and methodological articles on data saturation, with recommended sample sizes of approximately 30 participants for ethnographies, grounded theory studies; and 100 participants in qualitative ethology [3, 5]. This difference was expected because we worked on data from surveys with open-ended questions where the data collection process is linear: other designs, using purposeful sampling and/or in-depth, unstructured inquiry methods, would provide richer answers per participant and require smaller sample sizes to achieve data saturation. Similarly, we showed in our simulations that the addition of specific open-ended questions prompting topics of interest to a broad general open-ended question enhanced the ability to reach data saturation. This result was not surprising (i.e. asking more questions increases the chances for the participant to discuss relevant topics of the study), however, it contrasts with the relatively small number of open-ended questions in the surveys examined during our literature review.

In the literature, some authors have suggested the use of an “initial sample” and “stopping criterion” to determine point of data saturation [15]. We tested this method with our data using as stopping criterion 3 or 10 consecutive participants not eliciting a new theme; but results were unreliable. However, as the above mentioned method was developed for face-to-face theory-based interview studies, further research is needed to assess whether the method may be used for surveys using open-ended questions and which stopping criterion should be considered.

This study has several limitations. First, we analyzed a random sample of only 100 studies during our literature review, so it is possible that we missed some methods used by researchers to determine sample sizes in surveys with open-ended questions. In addition, estimates of the proportion of studies adequately reporting sampling methods may be inaccurate, even though this would not change the fact that, in general, sampling methods are poorly reported in surveys with open-ended questions. Second, our simulation study was based on a single dataset obtained using an online questionnaire answered by participants without the intervention of researchers. The results would likely differ in other contexts, with different participants, methods of data collection and analyses, especially with designs involving researcher-participants interactions as participant’s contribution are richer [27]. However, our study provides a proof of concept approach to data saturation for qualitative studies using web-based questionnaires and may help researchers plan this type of study. Third, we relied on an arbitrary definition for data saturation (90% of themes identified in the original study covered) or frequent themes. Because this was the first study to re-analyze by simulation the concept of data saturation in surveys using open-ended questions, we had to define limits for analysis.

Data saturation is an elastic notion. The number of participants depends on both the purpose of research and the desired analytic level. As for all qualitative studies, new data will always

add something new, but there are diminishing returns: how many participants researchers are willing to include for a single new theme or variation of a theme? In this study, we offer the first transparent method to assess if sample size in surveys using open-ended questions is adequate. Similar researches with different datasets, in other domains of inquiry and/or with different participants are needed.

## **5. Conclusion**

In our study using a web survey with open-ended questions about the burden of treatment, we showed that a sample of 150 participants ensured data saturation even in worst cases scenarios. This may help researchers determine the correct cut-off between recruiting more patients to find new themes or stopping because of futility.

## **6. Acknowledgments**

The authors thank Laura Smales (BioMedEditing) for editing.

## **7. Author Contributions**

Conceived and designed the experiments: V-TT, RP, BF and PR. Analyzed data: VT-T and RP. Wrote the first draft of the manuscript: V-TT. Contributed to the writing of the manuscript V-TT, RP, BF, PR. ICMJE criteria for authorship read and met: V-TT, RP, BF, PR. Agree with manuscript results and conclusions: V-TT, RP, BF, PR. PR is the guarantor, had full access to the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis

## **8. Competing interests**

The authors have declared that no competing interests exist



## 9. Funding

This study was funded by the French Health Ministry (PHRC AOM13127). Our team is supported by an academic grant from the program “Equipe espoir de la Recherche”, Fondation pour la Recherche Médicale, Paris, France (No. DEQ20101221475). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 10. References

1. Jansen, H., *The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods*. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 2010. **11**(2).
2. Feinstein, A., *Clinimetrics*. 1987, New Haven and London: Yale University Press.
3. Guest, G., A. Bunce, and L. Johnson, *How Many Interviews Are Enough? An Experiment with Data Saturation and Variability*. Field Methods, 2006. **18**(1): p. 59-82.
4. Baker, S. and R. Edwards, *How many qualitative interviews are enough? Expert voices and early career reflections on sampling and cases in qualitative research*. 2012, NCRM National Centre for Research methods.
5. Sandelowski, M., *Sample size in qualitative research*. Res Nurs Health, 1995. **18**(2): p. 179-83.
6. Malterud, K., V.D. Siersma, and A.D. Guassora, *Sample Size in Qualitative Interview Studies: Guided by Information Power*. Qual Health Res, 2015.
7. Charmaz, K., *Constructing grounded theory: a practical guide through qualitative analysis*. 2006, Thousand Oaks, CA: SAGE.
8. Rat, A.C., et al., *Content of quality-of-life instruments is affected by item-generation methods*. Int J Qual Health Care, 2007. **19**(6): p. 390-8.
9. Morse, J., *Styles of collaboration in qualitative enquiry*. Qualitative Health Research, 2008. **18**(1): p. 3-4.
10. Cheek, J., *The Politics and Practices of Funding Qualitative Inquiry*, in *The SAGE Handbook of Qualitative Research*, N. Denzin and Y. Lincoln, Editors. 2011, SAGE: Thousand Oaks, CA.
11. Mason, M., *Sample size and Saturation in PhD Studies Using Qualitative Interviews*. Forum: Qualitative Social Research, 2010. **11**(3).
12. Creswell, J., *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. 2002, Upper Saddle River, NJ: Pearson Education.
13. Tran, V.T., et al., *Taxonomy of the burden of treatment: a multicountry Web-based qualitative study of patients with chronic conditions*. BMC Med, 2015. **13**: p. 115.
14. Lingard, L., M. Albert, and W. Levinson, *Grounded theory, mixed methods, and action research*. BMJ, 2008. **337**: p. a567.
15. Francis, J.J., et al., *What is an adequate sample size? Operationalising data saturation for theory-based interview studies*. Psychol Health, 2010. **25**(10): p. 1229-45.
16. Vincent, A., M.O. Whipple, and L.M. Rhudy, *Fibromyalgia Flares: A Qualitative Analysis*. Pain Med, 2015.
17. Malicki, M. and A. Marusic, *Is there a solution to publication bias? Researchers call for changes in dissemination of clinical research results*. J Clin Epidemiol, 2014. **67**(10): p. 1103-10.

18. Ahrweiler, F., et al., *Determinants of physician empathy during medical education: hypothetical conclusions from an exploratory qualitative survey of practicing physicians*. BMC Med Educ, 2014. **14**: p. 122.
19. Rubenfeld, G.D., et al., *Barriers to providing lung-protective ventilation to patients with acute lung injury*. Crit Care Med, 2004. **32**(6): p. 1289-93.
20. McMurray, J.E., et al., *The attractiveness of internal medicine: a qualitative analysis of the experiences of female and male medical students*. Society of General Internal Medicine Task Force on Career Choice in Internal Medicine. Ann Intern Med, 1993. **119**(8): p. 812-8.
21. Costantini, M., et al., *Cancer patients as 'experts' in defining quality of life domains. A multicentre survey by the Italian Group for the Evaluation of Outcomes in Oncology (IGEO)*. Qual Life Res, 2000. **9**(2): p. 151-9.
22. Brown, A. and R. Davies, *Fathers' experiences of supporting breastfeeding: challenges for breastfeeding promotion and education*. Matern Child Nutr, 2014. **10**(4): p. 510-26.
23. Carlsen, B. and C. Glenton, *What about N? A methodological study of sample-size reporting in focus group studies*. BMC Med Res Methodol, 2011. **11**: p. 26.
24. Marshall, B., et al., *Does sample size matter in qualitative research? A review of qualitative interviews in research*. Journal of Computer Information Systems, 2013. **54**(1): p. 11.
25. Paley, J. and R. Lilford, *Qualitative methods: an alternative view*. BMJ, 2011. **342**: p. d424.
26. Malterud, K., *Qualitative research: standards, challenges, and guidelines*. Lancet, 2001. **358**(9280): p. 483-8.
27. Britten, N., *Qualitative interviews in medical research*. BMJ, 1995. **311**(6999): p. 251-3.

**List of tables**

Table 1: Proportion of themes elicited by using random samples of participants; sample size is fixed *a priori*.

Table 2: Proportion of themes elicited by using random samples of participants; sample size is determined with an initial sample + stopping criterion.

**List of figures**

Figure 1: Figure 1: Flow diagram of the literature review

Figure 2: Median probability (interquartile range [IQR]) of identifying at least one new theme as a function of the number of participants already included in the study.

Figure 3: Proportion of themes identified with a sample determined by a fixed number of patients (Panel 1) or an initial sample + stopping criterion (Panel 2).

Figure 4: Proportion of themes identified by enriching an initial sample of participants with participants reporting different conditions (Panel 1) or living in different countries (Panel 2).

**List of appendices:**

Appendix 1: Appendix 2: Examples of themes and verbatim from the original study

Appendix 2: Characteristics of the studies included in the literature review (n=100)

Appendix 3: Number of themes elicited as a function of the number of participants already included in the study using the experimental original data

Appendix 4a: Number of themes elicited as a function of the number of participants already included in the study by sex

Appendix 4b: Number of themes elicited as a function of the number of participants already included in the study by their chronic conditions

Appendix 4c: Number of themes elicited as a function of the number of participants already included in the study by educational level.

Appendix 5: Median probability (IQR) of missing at least one frequent theme as a function of the number of participants already included in the study.

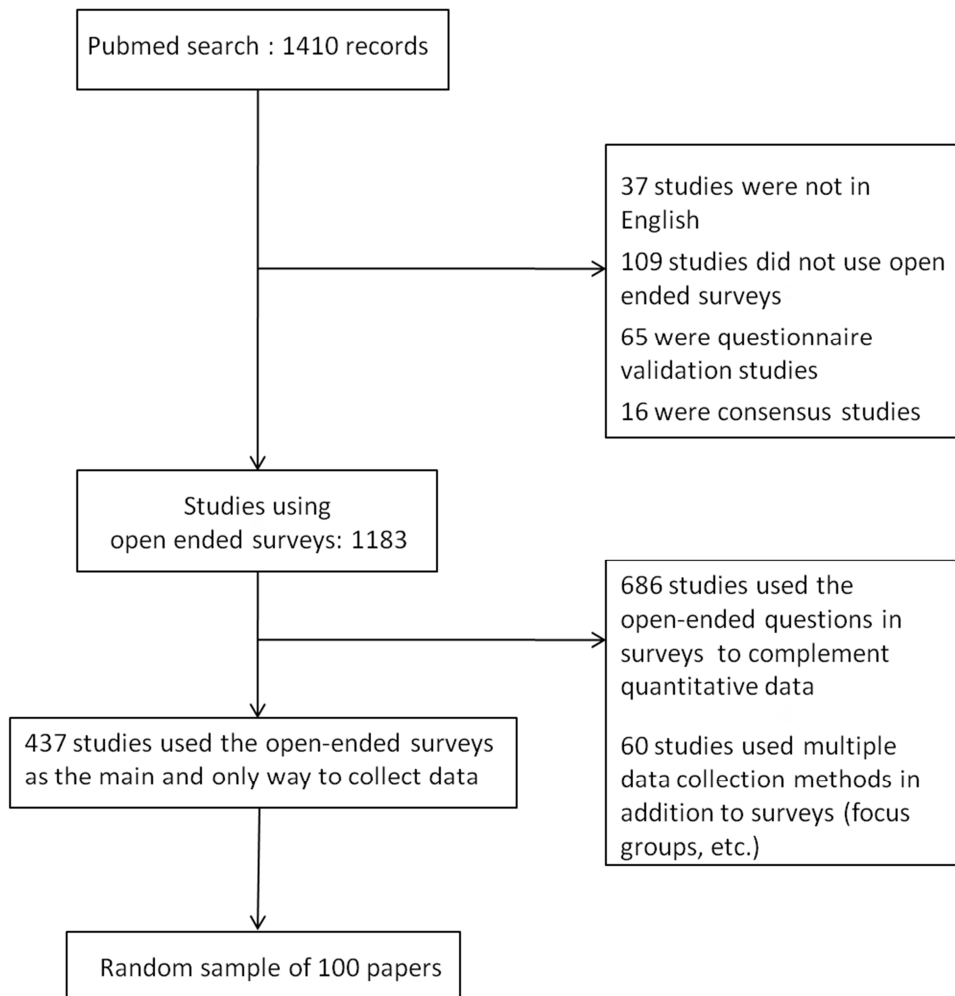
**Table 1: Proportion of themes elicited by using random samples of participants; sample size is fixed *a priori*.**

<b>Study Design</b>	<b>No. of patients included</b>	<b>Proportion of themes elicited Median (%)–(IQR)</b>	<b>Proportion of frequent themes* elicited Median (%)–(IQR)</b>
One broad open-ended question + questions on specific topics	15	54 (51-58)	70 (66-74)
	30	69 (66-72)	86 (83-89)
	50	79 (76-80)	94 (92-95)
	100	89 (86-90)	99 (99-100)
	150	92 (91-93)	100
	200	94 (93-95)	100
	250	96 (95-97)	100
Only one broad open-ended question	15	30 (27-33)	40 (35-44)
	30	44 (41-47)	57 (53-61)
	50	55 (52-58)	70 (67-74)
	100	70 (67-72)	85 (83-89)
	150	77 (75-79)	92 (89-94)
	200	81 (79-84)	95 (93-96)
	250	84 (83-86)	95 (93-96)

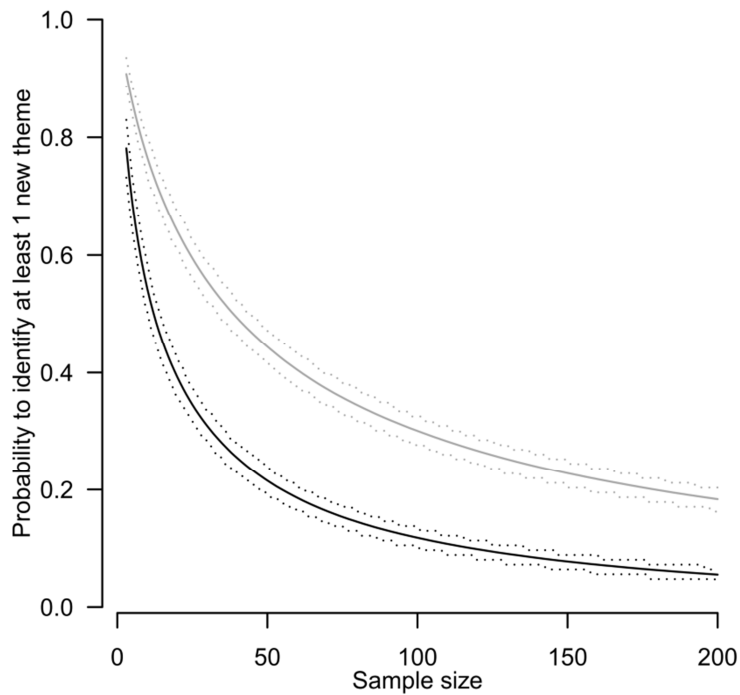
\* Frequent themes: mentioned by more than 2.5% of patients in the original study; IQR: interquartile range

**Table 2: Proportion of themes elicited by using random samples of participants; sample size is determined with an initial sample + stopping criterion.**

<b>Sampling strategy</b>	<b>Study design</b>	<b>Sample size Median–(IQR)</b>	<b>Proportion of themes elicited Median (%)–(IQR)</b>
Initial sample (n=10) + stopping if 3 consecutive participants do not elicit a new theme	One broad open-ended question + questions on specific topics	37 (28-43)	72 (66-77)
	Only one broad open-ended question	37 (28-46)	48 (41-54)
Initial sample (n=10) + Stopping if 10 consecutive participants do not elicit a new theme	One broad open-ended question + questions on specific topics	100 (90-120)	89 (85 -91)
	Only one broad open-ended question	130 (100-150)	74 (70-78)

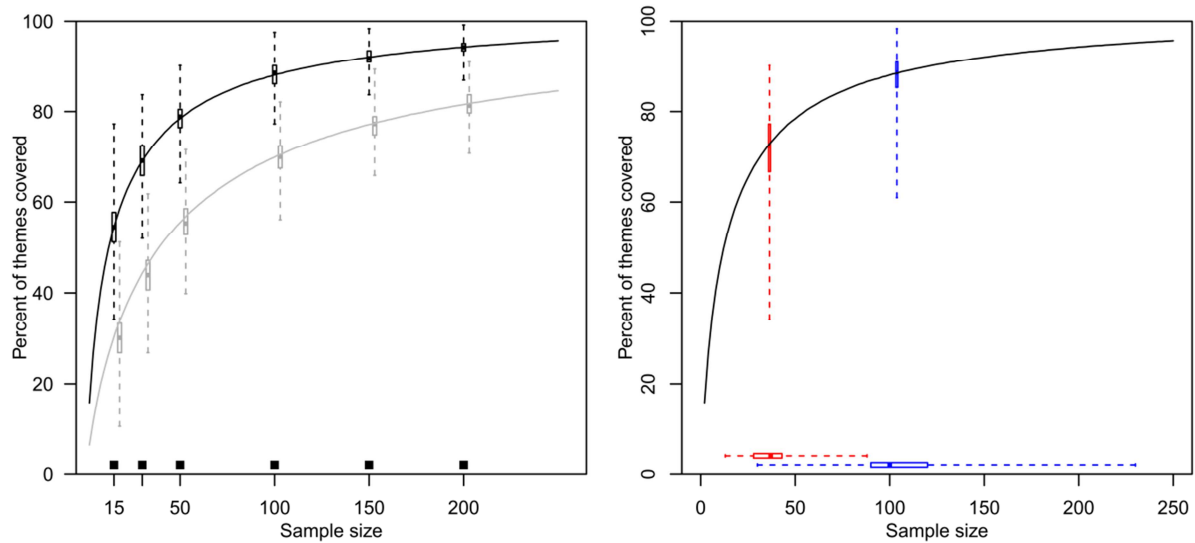
**Figure 1: Flow diagram of the literature review**

**Figure 2: Median probability (interquartile range [IQR]) of identifying at least one new theme as a function of the number of participants already included in the study.**



Dark line represents a study design with one open-ended question and specific questions on predetermined topics. Gray line represents a study design with only one open-ended question.

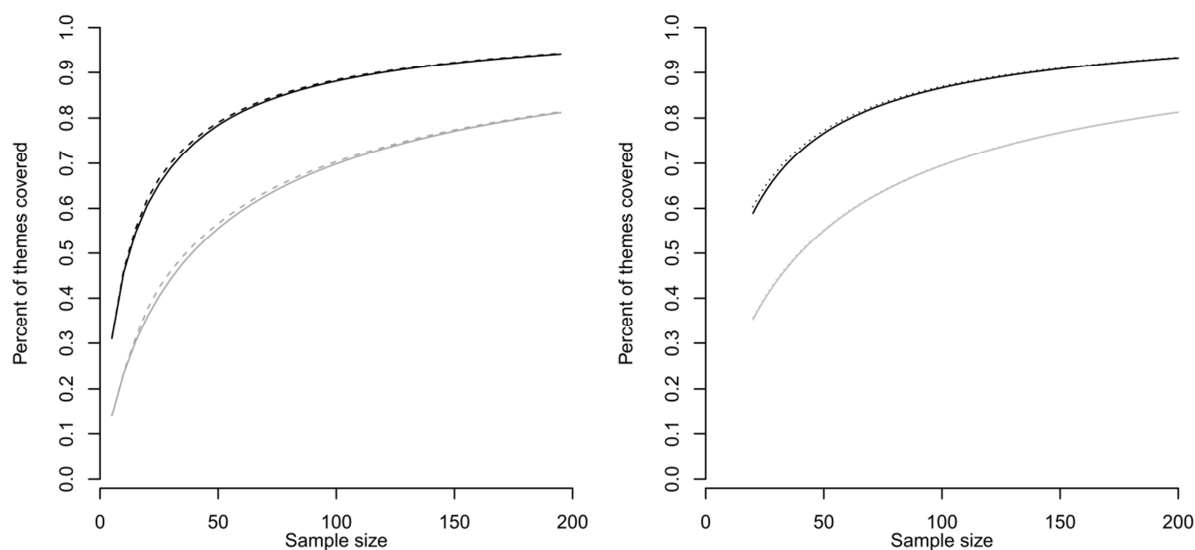
**Figure 3: Proportion of themes identified with a sample determined by a fixed number of patients (Panel 1) or an initial sample + stopping criterion (Panel 2).**



Panel 1 presents the proportion of themes identified with *a priori* fixed sample sizes of 15, 30, 50, 100, 150 and 200 participants with dark box plots representing a study design with multiple open-ended question and gray box plots representing a study design with only one open-ended question. Panel 2 presents the proportion of themes identified (vertical box plots) and number of patients involved (horizontal box plots) when using a sampling strategy based on an initial sample + stopping criterion. Red box plots correspond to an initial sample of 10 participants and stopping criterion of 3 additional subjects not mentioning a new theme. Blue box plots represent an initial sample of 10 participants and stopping criterion of 10 additional subjects not mentioning a new theme.



**Figure 4: Proportion of themes identified by enriching an initial sample of participants with participants reporting different conditions (Panel 1) or living in different countries (Panel 2).**



In Panel 1, the sampling strategy was simulated by the iterative addition of 5 participants with the less represented conditions in the sample. In Panel 2, the sampling strategy was simulated by the addition of 5 random foreign patients to an initial sample of French patients. Dotted lines represent the situation where specific enrichment is performed (according to conditions or country of residence). Plain lines represent addition of participants without specific enrichment. Dark lines represent a study design with one open-ended question and specific questions on predetermined topics. Gray lines represent a study design with only one open-ended question.