A novel comparison of qualitative data sources: content analysis of semi-structured patient interviews versus weblogs (blogs)

Acaster, S and Wild, D. Oxford Outcomes Ltd, Oxford, UK

Introduction

The research potential offered by first person narratives (FPN) posted on the internet has been recognized for some time (1), and the rapid increase in personal weblogs, or blogs, in recent years has only served to increase this potential. With home internet use now exceeding 211 million in the United States, and growing all the time, the demographics of the internet user are also changing, becoming more varied and further adding to the research potential offered by this group (2). Despite the recognition of FPNs or blogs as a rich source of qualitative material, the benefits of blog analysis as an adjunct to literature reviews in the early stages of qualitative analysis are not being capitalized upon within health related quality of life (HRQL) research.

In a recent literature search, only 1 paper was identified that considered unsolicited FPN data as a source of information in HRQL research (3). This paper used FPNs in the development of an item pool for a HRQL instrument. While, the latest FDA guidelines on the development of patient reported outcomes (PRO) instruments would not endorse this methodology for item creation (4), the use of FPN / blog analysis at an earlier stage of the research process could be even more informative.

We believe that considering blog analysis as an adjunct to literature reviews at the early stages of qualitative research could improve the quality of the data collected at the interview stage, by informing the development of interview schedules. As such, the present study was conducted to illustrate the potential of blogs in research development by comparing the thematic content of blogs with that gained from conducting semi-structured interviews.

Methods

The subject of menopausal hot flashes / night sweats was used to demonstrate the comparison. Twenty semi – structured interviews were conducted with women reporting to suffer hot flashes. The interviews focused on a description of the symptoms and their impact on HRQL. Thematic analysis using Atlas.ti was conducted on the interview transcripts to develop a simple conceptual model based on the identified symptoms of hot flashes and their impact on HRQL. The same process was then undertaken on the content of a number of unsolicited blog / FPN entries on the same topic. Four researchers conducted the analysis, two in each data source group. The data and themes identified in each conceptual model were then compared, and the relative advantages and disadvantages of each data source were assessed.

Results

Sample Demographics

The twenty women who participated in the interviews had an age range of 36 – 64 with a mean age of 53, and an average age of onset for the menopause of 47. On average participants had been experiencing hot flashes for approximately 6 years, with a large range between 1 and 16 years, and the average number of hot flashes / night sweats reported in a 24 hour period was 10.

As the blog analysis was passive, conducted on unsolicited information available in the public domain, with no active dialogue between researcher and participant there are no sample demographics for this group. The blog content, however, suggested that all participants experienced hot flashes and night sweats and that, in line with the interview group, these varied in frequency, duration and intensity.

Semi Structured Interview versus Blog Analysis.

Table 1 details some of the advantages and disadvantages of using semi structured interviews and blogs for qualitative analysis.

Interviews	Blogs
Advantages	
• Can probe for clarification or depth • Can capitalise on non verbal cues • Demographic Data • Can apply inclusion/ exclusion criteria • Can ensure a representative sample • Can control the content to address areas of interest • Can probe for clarification or depth	• Inexpensive • Convenient • Less time-consuming • Less invasive (for sensitive topics) • No restriction on content • No geographic barriers • Monitor reactions to recent events / products
Disadvantages	
• More time consuming • More expensive (face-to-face interviews) • Less willingness to discuss sensitive subjects • May miss interesting concepts not addressed in the discussion guide	• Selection bias • Not suitable for all disease areas • Can't probe for clarification or depth • No non verbal cues • Lack of demographic information • Unable to reliably apply inclusion / exclusion criteria • Issues of interest (e.g. impact on QOL) may not be discussed

Conceptual Models of Hot Flashes / Night Sweats

Thematic analysis of the interview and blog content identified similar themes in both data sources. The conceptual models generated from this process are illustrated below.

Figure 1: Conceptual Model Based on Interview Data

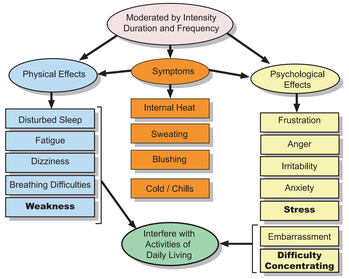
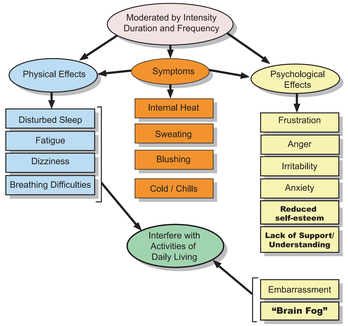


Figure 2: Conceptual Model Based on Blog Data



The same symptom profiles were elicited from both data sources. Likewise, the same broad themes of physical and psychological functioning were identified as having an impact on the participants HRQL, the latter relating to both emotional and cognitive effects. Both of these themes were also related to interference with daily activities for each group, and in both instances the symptoms and their impact on HRQL were moderated by the intensity, frequency and duration of the hot flashes.

There were some differences across groups within the profiles of physical and psychological functioning. The issues of physical weakness and psychological stress were only identified in the interview analysis, and the issues of lack of support/understanding and reduced self esteem were only apparent in the blog analysis.

The additional psychological effects identified within the blogs could be considered as more sensitive issues than those raised in the interviews, as they largely relate to others perception of their experiences rather than just a reaction to their physical experiences. However, it would have been necessary to introduce these concepts to the interview schedule to explore the relationship of these issues to the other elements of psychological functioning mentioned, such as frustration and embarrassment, as the information in blogs does not permit such interaction / probing. Similarly the issue of 'brain fog' would need to be explored in an interview to ascertain whether it is equivalent to the concentration problems mentioned in the interviews, or has some qualitative differences that could be further explored.

Data Quality

The similarity of the conceptual models belies the differences in the quality of the data. These differences fall into two broad categories: ambiguity resolution and freedom of expression. The interview data permitted the clarification of any ambiguous statements and further exploration of subjects as they arose. For example, in the development of the conceptual models, whereas the interview data contained concrete references to each of the themes identified, the blog analysis was in part guided by inference. For example, when an interviewee provided a story that could be interpreted as embarrassing, the interviewer could probe to discuss the participant's feelings, whereas in the blog this can only be inferred, as in this example:

"I was hoping the student didn't notice anything was amiss. Again, no luck. The student just sort of looked at me and said quietly, "are you ok?""

Interview data is thus a more reliable source of information for research purposes; however, the blog data did illustrate greater freedom of expression. This could be useful when determining which aspects of a disease / disorder's symptoms or impact on HRQL seem most pertinent to the patients/ participants when developing an interview schedule. For example, in the present analysis the blog data highlighted the fact that hot flashes were generally ignored or misunderstood by the women's partners friends and colleagues, and that this created anger and frustration;

"…I am flabbergasted…I hope that woman has one doozie of a hot flash when its her time…"

"My husband said , "your night sweats are making me sweat, could you move over when you feel one coming on?", guess who slept on the couch last night?"

As well as a sense that they only talk openly about hot flashes to other women experiencing them:

"I'm afraid they'll think of me as weak…It is the lack of understanding and compassion among my friends that has me keeping my mouth shut"

"She told me it was in my head (the hot flash) …I had no clue how to respond…Why would someone say this? Did I embarrass her by mentioning it? Is this still something you don't talk about in public?"

This information may partly explain why the interviewees' descriptions of their experiences were muted in comparison to those writing blogs. Had this analysis been available at the interview schedule development stage, the issue of others perception or support could have been addressed by the interviewer at the outset, and participants may have felt more comfortable discussing the impact hot flashes have on their quality of life.

Conclusions

Interviews are an essential component of qualitative research, providing a rich source of data and all the elements of control required to produce the standard of research expected by the research community and reimbursement agencies alike. However, when the disease or disorder of interest is suited to blog / FPN based analysis, this relatively inexpensive and immediate data source can prove to be a rich and fruitful adjunct to other preliminary analyses, such as literature reviews. As such, blog analyses could provide an inexpensive and efficient method to help inform interview schedule development, or be used as an exploratory first step into a field under consideration for future research.

References

(1) Robinson, K. (2001). Unsolicited narratives from the Internet: A rich source of qualitative data. Qualitative Health Research, 11, 706-714.

(2) Infoplease.com. (2004). How many online worldwide? Retrieved May 1, 2009, from http://www.infoplease.com/ipa

(3) Adair, C.E., Marcoux, G., Williams, A., Reimer, M. (2006). The internet as a source of data to support the development of a quality-of-life measure for eating disorders. Qualitative Health Research, 16, 538 -546.

(4) US Food and Drug Administration: Guidance for Industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. http://www.fda.gov/cder/guidance/5460dft.pdf.

Oxford Outcomes

www.oxfordoutcomes.com