

Generating qualitative data by design: the Australian Longitudinal Study on Women's Health qualitative data collection

Meredith Tavener^{a,b}, Catherine Chojenta^a and Deborah Loxton^a

^a Research Centre for Generational Health and Ageing, University of Newcastle, NSW, Australia

^b Corresponding author: meredith.tavener@newcastle.edu.au

Article history

Publication date: July 2016

Citation: Tavener M, Chojenta C, Loxton D. Generating qualitative data by design: the Australian Longitudinal Study on Women's Health qualitative data collection. Public Health Res Pract. 2016;26(3):e2631631. doi: <http://dx.doi.org/10.17061/phrp2631631>

Key points

- Free-text comments collected in longitudinal studies can provide a valuable data resource on a wide range of topics
- A variety of strategies for analysis of qualitative free-text comments have been used on data collected by the Australian Longitudinal Study on Women's Health
- Researchers must consider issues surrounding the rigour of the analysis, and ensure the quality of analysis procedures and conclusions made from the data

Abstract

Objectives and importance of study: The purpose of this study was to illustrate how qualitative free-text comments, collected within the context of a health survey, represent a rich data source for understanding specific phenomena.

Study type: Work conducted with data from the Australian Longitudinal Study on Women's Health (ALSWH) was used to demonstrate the breadth and depth of qualitative information that can be collected. The ALSWH has been collecting data on women's health since 1996, and represents a unique opportunity for understanding lived experiences across the lifecourse.

Methods: A multiple case study design was used to demonstrate the techniques that researchers have used to manage free-text qualitative comments collected by the ALSWH.

Results and conclusions: Eleven projects conducted using free-text comments are discussed according to the method of analysis. These methods include coding (both inductively and deductively), longitudinal analyses and software-based analyses. This work shows that free-text comments are a data resource in their own right, and have the potential to provide rich and valuable information about a wide variety of topics.

Introduction

Qualitative research is both a vibrant and contested method of research.¹ It has become increasingly recognised as a way to explain complex social processes, and capture essential information about people's beliefs, values and experiences that, in turn, influence particular health behaviours.² Often, free-text comments within the context of a survey are used to corroborate the content of the survey and to ensure that all relevant issues have been covered.³

Approaches to qualitative enquiry are highly diverse, and can include telephone, face-to-face or online exchanges; diary or journal keeping; photo records and observation; interviews conducted while walking, or at a person's home or business; and analysis of images or personal narratives.

Narrative analysis, in particular, whereby people's accounts or stories are interpreted and reported, has become popular as a qualitative research approach in response to the perceived lack of human stories in the 1960s.⁴ Data for narrative analysis can be collected using a variety of means, as long as it involves capturing an account, and can include invitations for people to record free-text comments on surveys. This approach can provide an opportunity for individuals to 'tell their story', and represents a rich source of data for understanding life experiences and specific phenomena, particularly in a longitudinal context.⁵

The Australian Longitudinal Study on Women's Health (ALSWH) has been inviting women to write free-text comments on their surveys for the past 20 years. The ALSWH has a wide breadth and depth of qualitative information that demonstrates women's life experiences. The purpose of this paper is to illustrate how qualitative free-text comments, collected within the context of a health survey, represent a rich data source for understanding specific phenomena.⁶

Methods

Australian Longitudinal Study on Women's Health

The ALSWH first collected mailed survey data from more than 42 000 women in 1996 across three birth cohorts: 1973–78 (aged 18–23 in 1996), 1946–51 (aged 45–50) and 1921–26 (aged 70–75). A new cohort of women born between 1989 and 1995 was recruited in 2012–13. At the conclusion of each ALSWH survey, participants are asked "Have we missed anything?", and are invited to record free-text comments on their health and any other topics that they felt the survey did not adequately include.

Selected analyses of free-text comments

Between 1996 and 2013, more than 56 000 free-text comments were contributed by more than 30 000 participants from the four ALSWH cohorts across a total of 22 surveys (Table 1).

The sheer number of comments in the qualitative datasets can be daunting. Conversely, the fear that a respondent may not write about a topic of interest can be equally disturbing to researchers. However, a number of techniques have been used to manage the data.

Firstly, feasibility of topic areas can be facilitated by initial keyword searches – that is, searching the database for key terms to see whether any appropriate data are identified.

Secondly, it is useful to determine the broad inclusion and exclusion criteria of the data to be extracted – for example, identification of the cohorts to be included and the years when data were collected, and any experiences recorded in the quantitative data that need to be considered (e.g. participants with a particular condition might be included or excluded). Determining these criteria is the equivalent of determining a sampling frame from which a sample of data will be drawn.

Thirdly, the sample data can be extracted by selecting those data most relevant to the enquiry, or by determining keywords and only extracting those data that include the keywords. Although the latter option is more time- and cost-efficient, there is the potential to miss relevant data. For example, research conducted by Rich et al.⁷ using data from all three original cohorts showed that searching for the keyword 'drought' would have resulted in a loss of relevant data where terms such as 'dry' had been used to indicate drought. Nevertheless, the size of the data sample drawn and the resources available are likely to determine the methods used.⁸

To date, free-text comments on ALSWH surveys have been analysed to explore issues such as emotional distress, relationships, health-seeking behaviours and work-life balance. Numerous approaches have been taken to analyse ALSWH free-text comments, and selected examples are highlighted below.

Results

There were four main findings from our review of ALSWH analyses using women's free-text comments, as detailed in the following results.

Table 1. Number and percentage of participants who commented by cohort over each survey

ALSWH cohort (birth year)	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5	Survey 6	Survey 7	Total number of words recorded by each cohort	% of cohort to have ever written a free-text comment in the ALSWH
1921–26	2978	4695	3955	4353	2481	698	–	1 221 713	59
1946–51	2392	2058	2672	2967	3731	3745	3021	1 155 298	61
1973–78	2423	1948	2266	2394	2415	1656	–	629 113	57
1989–95	1415	1150	740	–	–	–	–	106 924	16
Total	9208	9851	9633	9714	8627	6099	3021	3 113 048	na

– = survey not held; ALSWH = Australian Longitudinal Study on Women's Health; na = not applicable

Combining inductive and deductive analysis

ALSWH data have been analysed both inductively (identifying emergent themes from the free-text comments) and deductively (looking for predetermined key issues or themes). The method chosen by Tavener et al.⁹, when exploring the phenomenon of life after stroke in older ALSWH participants, was a hybrid of a pre-existing framework of expected categories of stroke experience¹⁰ and a data-driven inductive approach.¹¹ The existing World Health Organization International Classification of Functioning, Disability and Health (ICF) core dataset for stroke was examined against older women's free-text comments about living with stroke. The thematic analysis showed that, for older Australian women, concerns about post-stroke living were not adequately classified by the ICF, and that the women's narratives of stroke experiences were not congruent with the clinical stroke phrasing, potentially impeding a full recovery for the women.

A hybrid approach was also taken by Tavener et al.¹² to explore older women's longitudinal experiences of health and ageing. They applied the salutogenic model of wellness to explore adaptive strategies of women living with chronic conditions. Women's free-text comments, recorded from 1999 to 2011 on 3-yearly ALSWH surveys, were subjected to longitudinal analysis for contrasting groups. Inductive thematic analysis identified categories of information described by the women, which were compared with deductive analysis findings of salutary expressions such as vulnerability, growth, coping and security. The hybrid approach enabled a rational and detailed longitudinal synthesis of women's health experiences to be compared with the salutogenic model.

Pattern, theme and content analysis

Establishing categories or patterns of common textual meaning and/or counting the number of instances in which they are used are common ways to analyse free-text comments. Analyses can result in categorised representations of accounts, patterns of categories, or themes or particular narratives being identified, and a numerical description of textual features.

Thematic analysis of ALSWH data has explored the experiences of older women who take on caregiving roles.¹³ The analysis examined data from Survey 1 of the 1921–26 cohort, after 10% of respondents were identified from quantitative survey data as being caregivers and recorded free-text comments about their caregiving experiences. Following an empirical literature review on the topic¹⁴, which provided a conceptual framework for the content analysis to follow, researchers analysed the free-text comments of 168 women. This work identified that caregivers had significantly lower levels of emotional wellbeing than noncaregivers, and felt stressed, rushed and pressured. The women's free-text comments provided insight into the forces that drive older women into providing family care.

A similar approach was repeated with the 1946–51 cohort.¹⁵ A content analysis of comments made by 185 women was examined for themes regarding caregiving. A cross-check step was included, where two researchers examined the comments, with one categorising the comments and the second recoding a subset. Emerging themes not already encompassed by the empirical literature search¹⁴ were then identified. Conducting a review of caregiving issues in older¹³ and then mid-aged women¹⁵ also allowed lifecourse nuances to emerge, such as differences in employment concerns, and caring for children and parents.

A thematic analysis of widowed women's free-text comments identified that both short- and long-term effects followed the death of a husband.¹⁶ The work followed on from the findings of a quantitative examination of the health impacts of widowhood for the same cohort of women in the 12 months following bereavement¹⁷, which formed the theoretical framework for the qualitative analysis. Comments from more than 700 women elicited at Survey 1 of the 1921–26 cohort were examined. Key themes of health, social relationships and support, and financial and structural issues were identified as being of concern to the women.

Chojenta et al.¹⁸ used the free-text comments elicited by the 1973–78 cohort in Surveys 2–5 to identify risk factors for postnatal depression: examining those risk factors already identified, and exploring any new or emerging concepts in the qualitative data. The researchers analysed the initial content using the principles of directed content analysis, and codes were developed using deductive category application.¹⁹ Using the principles of directed content analysis, they used the results of previous research to identify initial categories. This approach enabled direct comparisons with the quantitative findings and consideration of the theoretical implications of the findings. The thematic analysis also involved a deeper level of analysis than the content. These findings indicate that, for some women, symptomatology might be a more suitable strategy for discussion by health practitioners rather than direct labelling of the event as 'postnatal depression'.

Use of medicines was compared across three ALSWH cohorts by Walkom et al.²⁰, who conducted a thematic analysis of women's experiences with day-to-day expenses that relate to medicines and their healthcare. Along with three broad themes of costs of medicines, doctor visits and complementary medicines, the researchers revealed age-specific issues with medicine costs, which included the importance of contraceptive medicines (for younger women), hormone replacement therapy (for mid-aged women) and osteoporosis medicines (for older women). This is critical information for life-stage affordability of health services, including access to medicines.

Longitudinal analyses of qualitative data

Building upon earlier work on ALSWH women's experiences of widowhood, Kennaugh et al.²¹ conducted a longitudinal analysis of free-text comments to understand widowhood 15 years beyond the initial period of bereavement. The data were analysed inductively and deductively (as an extension of the work by Feldman et al.¹⁶). Two of the researchers reviewed the coding process, and the third author then reviewed Kennaugh's coding against the full codebook. The differences were then reconciled by taking the strongest meaning held on each coded segment of text. Results confirmed that the original themes continued to persist over time, highlighting the importance of continuing to support women and to be responsive to their changing needs well after the initial period of bereavement.

Rich et al.⁷ used the free-text comments of women from the 1946–51 cohort to explore their experiences of living through drought. The use of longitudinal data in this context was ideal, because drought is known to be a long-lasting event, and the long-term impact of living through drought would be evident with longitudinal data. Women's comments were examined in chronological order, guided by Riessman's narrative analysis framework²² of discrete story and extended-account models.

Use of software to analyse data

Numerous computer-based software packages can assist with data organisation for qualitative data analysis. Software can help a researcher deal with large amounts of data, increase the flexibility of data organisation and allow auditability of the work.²³

Research by Tavener²⁴ made use of Leximancer software to contrast social narratives between groups of baby boomer women. Leximancer software uses word-association information to produce relationship and concept maps from excerpts of text.²⁵ For this work, the software was trialled as a 'hands-off' alternative approach to researcher-led inductive coding, to identify and describe the top three concepts identified from different groups of baby boomer women, categorised according to health, income and marital status. However, although the software did map women's comments into different concepts of health, work and life, the primary researcher did not enjoy the hands-off experience and felt removed from the richness of the free-text comments. In contrast, Rich⁸ found the use of Leximancer to analyse free-text comments about the experience of drought across the original three ALSWH cohorts to be very useful, particularly for distinguishing the different themes that emerged over time for the three different cohorts.

ALSWH free-text comments have also been used to investigate the stability of propositional density (i.e. a measure of how much information is conveyed in an utterance or text) as a predictor of language decline with ageing.²⁶ More than 600 texts from the 1946–51

cohort were analysed longitudinally, using automated propositional density rating software. Results revealed propositional density to be a stable measure over time when comparing grouped data, and revealing between- and within-subject variation over time.

The above examples demonstrate the versatility of free-text data, which can be analysed using many of the techniques and tools that are used with more common qualitative data. Although the scope of the topics that have been investigated is wide, there remains a vast untapped mine of information in free-text comments in the ALSWH and many other survey-based studies.

Discussion

This paper has illustrated the value of free-text analysis and cohort longitudinal data. For example, differences exist in key themes between the ALSWH birth cohorts and over time.⁵ The authors concluded that free-text comments represent a valuable data source that is suitable for content, thematic and narrative analysis, particularly when examined longitudinally. Of particular significance is the fact that the women chose to record their comments on a research survey. The selection of stories chosen by the women to share with the ALSWH team reflects the relative level of importance of that information to them and their need to share that part of their experience. The women have written about a wide range of topics of interest to them, with both positive and adverse events recorded.

Although qualitative research methods are continually evolving, it is important that researchers continue to address issues of rigour – that is, reconsidering traditionally quantitative terms, such as validity and reliability, in qualitative terms instead. Rigour in qualitative research refers to the quality of the research process, evidenced by features such as credibility, dependability, confirmability and transparency.²⁷

For the ALSWH, the concept of rigour is considered in terms of the quality of the research process: from initial steps to anonymise the data through to iterations of coding and categorising, with the key being analytic reflexivity. Rigorous qualitative research must be both transparent and explicit. Researchers need to describe what they did in clear, simple language, account for their own influence on the work and be open about the analysis procedures. In this regard, analysis of ALSWH data is no different from any qualitative analysis.

Checklists encourage researchers to report results in a way that is considered both useful and credible. Following a systematic review of qualitative studies, Tong et al.²⁸ have published the consolidated criteria for reporting qualitative research (COREQ). The checklist aims to help researchers report key aspects of the research team, study methods and context, and analysis and interpretation, with reference to interviews and focus groups. Kitto et al.²⁹ published a checklist to address

quality reporting across the domains of clarification, justification, procedural, rigour, representativeness, interpretation, reflexivity and transferability. Both the Tong and Kitto checklists have been used by researchers who analyse ALSWH free-text comments. However, we also acknowledge the caution offered by Barbour³⁰, to not reduce qualitative research to a “list of technical procedures” only. Furthermore, existing checklists tend to give consideration to the most common data collection methods, such as focus groups and interviews, and give less attention to methods that are equally valid but less common, such as diary studies and analysis of free-text comments (such as those collected by the ALSWH). This is a potential problem when seeking to demonstrate rigour to journal reviewers, if the methods chosen are not adequately dealt with in the recommended (or required) checklist.

Conclusion

It has become common practice to include an open-ended question at the conclusion of surveys, whether they are online, written or verbal. The purpose has largely been to gather feedback on the survey itself. ALSWH researchers have realised the potential of this resource, where participants can communicate with researchers without being confined to particular quantitative response options. These free-text comments are data in their own right, and have the potential to provide rich and valuable information about a wide variety of topics in the person's own words.

Acknowledgements

The research on which this paper is based was conducted as part of the ALSWH at the University of Newcastle and the University of Queensland. We are grateful to the women who provided the survey data, the staff of the ALSWH, and the Australian Government Department of Health for funding. Researchers in the Research Centre for Generational Health and Ageing at the University of Newcastle are members of the Hunter Medical Research Institute.

Competing interests

None declared

Author contributions

All authors contributed equally to this work.

References

1. Leavy P. The Oxford Handbook of Qualitative Research. New York: Oxford University Press; 2014.
2. Curry LA, Nembhar IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. *Circulation*. 2009;119(10):1442–52.
3. O'Cathain A, Thomas KJ. “Any other comments?” Open questions on questionnaires – a bane or a bonus to research? *BMC Med Res Methodol*. 2004;4:25.
4. Riessman CK. Narrative methods for the human sciences. Thousand Oaks, CA: Sage Publications; 2008.
5. Rich JL, Chojenta C, Loxton D. Quality, rigour and usefulness of free-text comments collected by a large population based longitudinal study – ALSWH. *PLoS ONE*. 2013;8(7):e68832.
6. Elliott J. The narrative potential of the British Birth Cohort Studies. *Qualitative Research*. 2008;8(3):411–21.
7. Rich J, Wright SL, Loxton D. ‘Patience, hormone replacement therapy and rain!’ Women, ageing and drought in Australia: narratives from the mid-age cohort of the Australian Longitudinal Study on Women's Health. *Aust J Rural Health*. 2012;20:324–8.
8. Rich JL. The nature of things: an interdisciplinary investigation into the experiences and impacts of drought for three generations of Australian women. Newcastle, NSW: The University of Newcastle; 2014.
9. Tavener M, Thijsen A, Hubbard I, Francis JL, Grennal C, Levi C, Byles J. Acknowledging how older Australian women experience life after stroke: how does the WHO 18-Item brief ICF Core Set for Stroke compare? *Health Care Women Int*. 2015;36(12):1311–26.
10. Crabtree B, Miller W. Using codes and code manuals: a template organizing style of interpretation. In: Crabtree B, Miller W, editors. *Doing qualitative research*. Newbury Park, CA: Sage Publications; 1999. p. 163–77.
11. Boyatzis R. Transforming qualitative information: thematic analysis and code development. Thousand Oaks, CA: Sage Publications; 1998.
12. Tavener M, Byles JE, Curryer C, Forder P. Applying the salutogenic model for better health outcomes in older Australian women. Paper presented at: 47th Australian Association of Gerontology National Conference ‘50 not out aiming for a century’. Adelaide: Australian Association of Gerontology; 2014.
13. Lee C. Experiences of family caregiving among older Australian women. *J Health Psychol*. 2001;6(4):393–404.
14. Lee C. Health, stress and coping among women caregivers: A review. *J Health Psychol*. 1999;4(1):27–40.
15. Lee C, Porteous J. Experiences of family caregiving among middle-aged Australian women. *Fem Psychol*. 2002;12(1):79–96.
16. Feldman S, Byles JE, Beaumont R. ‘Is anybody listening?’ The experiences of widowhood for older Australian women. *J Women Aging*. 2000;12(3–4):155–76.

17. Byles JE, Feldman S, Mishra G. For richer, for poorer, in sickness and in health: Older widowed women's health, relationships and financial security. *Women Health*. 1999;29(1):15–30.
18. Chojenta C, Loxton D, Lucke J. The perfect mother wouldn't have that: Australian women's experiences of motherhood and postnatal depression. Paper presented at: Challenging the Boundaries: 16th Qualitative Health Research Conference. Vancouver: International Institute for Qualitative Methodology, University of Alberta; 2010.
19. Mayring P. Qualitative content analysis. *Forum: Qualitative Social Research*. 2000;1(2):20.
20. Walkom EJ, Loxton D, Robertson J. Costs of medicines and health care: a concern for Australian women across the ages. *BMC Health Serv Res*. 2013;13:484.
21. Kennaugh R, Byles J, Tavener M. Beyond widowhood: do prior discovered themes that describe the experiences of older Australian widowed women persist over time? *Women Health*. 2015;1:1–16.
22. Riessman CK. *Narrative Analysis. Narrative, Memory & Everyday Life*. Huddersfield: University of Huddersfield; 2005. p. 1–7.
23. St John W, Johnson P. The pros and cons of data analysis software for qualitative research. *J Nurs Scholarsh*. 2000;32(4):393–7.
24. Tavener M. *Your bloomin' lot: an empirical study of the popular baby boomer stereotype*. Newcastle, NSW: The University of Newcastle; 2010.
25. Smith AE, Humphreys MS. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behav Res Methods*. 2006;38(2):262–79.
26. Spencer E, Craig H, Ferguson A, Colyvas K. Language and ageing – exploring propositional density in written language – stability over time. *Clin Linguist Phon*. 2012;26(9):743–54.
27. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation*. 1986;30:73–84.
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349–57.
29. Kitto SC, Chesters J, Grbich C. Quality in qualitative research. *Med J Aust*. 2008;188(4):243–6.
30. Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ*. 2001;322(7294):1115–7.