Education and debate

Checklists for improving rigour in qualitative research: a case of the tail wagging the dog?

Rosaline S Barbour

Qualitative research methods are enjoying unprecedented popularity. Although checklists have undoubtedly contributed to the wider acceptance of such methods, these can be counterproductive if used prescriptively. The uncritical adoption of a range of "technical fixes" (such as purposive sampling, grounded theory, multiple coding, triangulation, and respondent validation) does not, in itself, confer rigour.

In this article I discuss the limitations of these procedures and argue that there is no substitute for systematic and thorough application of the principles of qualitative research. Technical fixes will achieve little unless they are embedded in a broader understanding of the rationale and assumptions behind qualitative research.

Checklists in quantitative research

In medical research the question is no longer whether qualitative methods are valuable but how rigour can be ensured or enhanced. Checklists have played an important role in conferring respectability on qualitative research and in convincing potential sceptics of its thoroughness.[1–3] They have equipped those unfamiliar with this approach to evaluate or review qualitative work (by providing guidance on crucial questions that need to be asked) and in reminding qualitative researchers of the need for a systematic approach (by providing an aide-mémoire of the various stages involved in research design and data analysis[4]).

Qualitative researchers stress the importance of context but sometimes forget that research itself is carried out against an ever-changing backdrop. Now that it has secured a place in the methodological mainstream, qualitative research is increasingly being influenced by funding and editorial policies. Despite disclaimers by authors that their checklists should be viewed as being "reflective rather than constitutive of good research,"[5] there is evidence that checklists are sometimes being used prescriptively.

Over the past two years, several researchers have informed me that they must comply with various procedures (such as respondent validation, multiple coding, etc) in order to satisfy the requirements of specific journals where they hope to publish their work. (I am not concerned here with the accuracy of such claims, although my own experience suggests these are exaggerated.) While we all attempt to tailor our writing to match the style and format of the journal in question, the strategic adoption of such technical fixes

Summary points

Checklists can be useful improving qualitative research methods, but overzealous and uncritical use can be counterproductive

Reducing qualitative research to a list of technical procedures (such as purposive sampling, grounded theory, multiple coding, triangulation, and respondent validation) is overly prescriptive and results in "the tail wagging the dog"

None of these "technical fixes" in itself confers rigour; they can strengthen the rigour of qualitative research only if embedded in a broader understanding of qualitative research design and data analysis

Otherwise we risk compromising the unique contribution that systematic qualitative research can make to health services research

Department of General Practice, University of Glasgow, Glasgow G12 0RR
Rosaline S Barbour
senior lecturer in primary care R&D

rsb2g@clinmed.gla.ac.uk

has wider repercussions. The complex dilemmas in research design that qualitative researchers face with regard to sampling, choice of methods, and approaches to analysis cannot be solved by formulaic responses. If we succumb to the lure of "one size fits all" solutions we risk being in a situation where the tail (the checklist) is wagging the dog (the qualitative research).

From reading recent journals and my experience of reviewing journal articles and grant submissions, I find that the five technical fixes currently enjoying the greatest popularity are purposive sampling, grounded theory, multiple coding, triangulation, and respondent validation (table). The rest of this article outlines their limitations and provides a more realistic appraisal of their potential.

Purposive sampling

Rather than aspiring to statistical generalisability or representativeness,[6] qualitative research usually aims to reflect the diversity within a given population.[7] In the past qualitative research often relied on convenience samples, particularly when the group of interest was difficult to access. Purposive (or theoretical[8]) sampling, however, offers researchers a degree of control rather

Technical fixes used to confer rigour on qualitative analysis, the concerns they are thought to address, and their realistic potential

Technical fix	Concerns addressed	Realistic potential
Purposive sampling	Bias	Enhancing sample coverage and providing a framework for analysis
Grounded theory	Original theorising	Developing existing theory or, occasionally, new theories
Multiple coding	Inter-rater reliability	Refining interpretations or coding frameworks
Triangulation	Confirmation or refutation of internal validity	Corroborating or, more often, refining findings
Respondent validation	Confirmation or refutation of interpretations	Corroborating or, more often, refining findings

than being at the mercy of any selection bias inherent in pre-existing groups (such as clinic populations). With purposive sampling, researchers deliberately seek to include "outliers" conventionally discounted in quantitative approaches.[9] It allows for such deviant cases to illuminate, by juxtaposition, those processes and relations that routinely come into play, thereby enabling "the exception to prove the rule."[9] [10]

Some strategies claimed as examples of purposive sampling in effect involve hybrids, which retain elements of random or convenience sampling and which are unlikely to yield the spread of respondents required. When they are provided at all, details of sampling are often dealt with in the methods section of papers and are disregarded in the analysis section, which often consists of little more than a description of undifferentiated themes that emerged during data analysis. For example, we are likely to be told in the methods section that a third of the sample were men, but the analysis section does not discuss how their perspectives differed from those of female respondents.

Such approaches do not use qualitative datasets to full advantage. That would involve applying the constant comparative method[11] to continuously compare the views and experiences of respondents who have been selected precisely—indeed, purposively—in order to illuminate subtle but potentially important differences. In other words, samples may have been selected purposively, but they are not being used purposefully to interrogate the data collected.

Grounded theory

In its purest form the grounded theory approach to data analysis alleges that all explanations or theories are derived from the dataset itself rather than from a researcher's prior theoretical viewpoint.[12] In practice, however, you are unlikely to obtain research funding without having carried out a thorough literature review or having formulated some idea of the content of the data you are likely to collect.

According to many researchers who invoke the concept of grounded theory, coding categories reflect the content of data collected rather than the questions on the interview schedule or focus group topic guide and often use concepts or vocabulary borrowed from respondents. However, few published papers yield the surprises likely to be a feature of analyses driven entirely by respondents' concerns, and the terminology and theories to which papers appeal generally bear an uncanny resemblance to current disciplinary concerns and debates.

Bryman and Burgess have criticised the use of grounded theory as "an approving bumper sticker" invoked to confer academic respectability rather than as a helpful description of the strategy used in analysis.[13] Melia claims that most researchers use a pragmatic variant, whereby they can achieve added value by identifying new themes from the data alongside those that could have been anticipated from the outset.[14] All too often, however, the tension between these two different sorts of insight—and its potential to illuminate the topic being studied—is not explored in the presentation of findings.

In the absence of an attempt to systematically analyse the commonalities and contradictions reflected in the data, many researchers produce an artificially neat and tidy account that is descriptive rather than analytical and which militates against formulating in-depth analyses. Uncritical adoption of grounded theory can result in explanations tinged with the "near mysticism" that Melia derides in the original text on grounded theory.[12] A sleight of hand produces a list of "themes," and we are invited to take it on trust that theory somehow emerges from the data without being offered a step by step explanation of how theoretical insights have been built up.

Multiple coding

Multiple coding concerns the same issue as the quantitative equivalent "inter-rater reliability" and is a response to the charge of subjectivity sometimes levelled at the process of qualitative data analysis. Although multiple coding does not usually demand complete replication of results, it does involve the cross checking of coding strategies and interpretation of data by independent researchers. While I would caution against multiple coding of entire datasets (on the grounds of economy in both cost and effort), some element of multiple coding can be a valuable strategy. It can be useful to have another person cast an eye over segments of data or emergent coding frameworks, and this is a core activity of supervision sessions and research team meetings.[15]

Although six experienced researchers who independently coded one focus group transcript showed substantial agreement, Armstrong et al found considerable variation in the ways that they packaged coding frameworks (including the language used).[16] This is not surprising, given the complexity of qualitative data and the range of disciplinary backgrounds and interests of qualitative researchers. Indeed, Mauthner et al have shown how researchers' original interpretations may shift when they revisit previously collected data.[17]

However, the degree of concordance between researchers is not really important; what is ultimately of value is the content of disagreements and the insights that discussion can provide for refining coding frames. The greatest potential of multiple coding lies in its capacity to furnish alternative interpretations and thereby to act as the "devil's advocate" implied in many of the checklists[1–3] in alerting researchers to all potentially competing explanations. Such exercises encourage thoroughness, both in interrogating the data at hand and in providing an account of how an analysis was developed. Whether this is carried out by a conscientious lone researcher, by a team, or by involving independent experts is immaterial: what matters is that a systematic process is followed and that this is rendered transparent in the written research project.

Triangulation

The current heavy reliance on triangulation in grant applications testifies both to the respect accorded to this concept and to its perceived value in demonstrating rigour. Triangulation addresses the issue of internal validity by using more than one method of data collection to answer a research question. In principle, it sounds eminently feasible to combine, say, observational fieldwork and interviews or focus groups in order to get a broader view. However, triangulation is difficult to perform properly: data collected using different methods come in different forms and defy direct comparison. This is true for different types of qualitative data, such as interview and focus group transcripts, as well as for the more obvious differences between qualitative and quantitative data.

The production of similar findings from different methods merely provides corroboration or reassurance; the absence of similar findings does not, however, provide grounds for refutation. This is because different methods used in qualitative research furnish parallel datasets, each affording only a partial view of the whole picture.

Triangulation relies on the notion of a fixed point, or superior explanation, against which other interpretations can be measured. Qualitative research, however, is usually carried out from a relativist perspective, which acknowledges the existence of multiple views of equal validity.[1] Therefore, it does not readily lend itself to the production or observance of such a hierarchy of evidence.[18]

Richardson suggests that it is more helpful to conceive of complementary rather than competing perspectives and offers the term "crystallisation" as an alternative to triangulation.[19] Qualitative research, with its distinctive approach to harnessing the analytical potential of exceptions, allows a research question to be examined from various angles. As Mays and Pope conclude, comprehensiveness may be a more realistic goal for qualitative research than is internal validity.[20] According to this approach, apparent contradictions (or exceptions) do not pose a threat to researchers' explanations; they merely provide further scope for refining theories.

Respondent validation

Given the current focus on consumerism, respondent validation, which involves cross checking interim research findings with respondents, has a ready appeal. Respondents' reactions to emerging findings can certainly help refine explanations—as can key informants'—but several commentators have questioned whether it is always appropriate.[20] [21] As Mays and Pope point out, researchers seek to provide an overview whereas respondents have individual concerns, and this can result in apparently discrepant accounts.[20]

Sometimes researchers choose to disregard their own interpretations and to accept those of respondents at face value. This can be cosy but may lead to collusion: Atkinson has warned of the dangers of "romanticising" respondents' accounts.[22] Respondent validation exercises, such as reading of drafts, make considerable demands on participants' time and,

depending on the research topic and content of transcripts, can even be exploitative or distressing.[23]

Respondent validation can be particularly valuable in action research projects, where researchers work with participants on an ongoing basis to facilitate change. Most health services research, however, involves a one-off data collection exercise, in which respondent validation may be more trouble than it is worth.

Conclusion

Although some of the technical fixes discussed here may seem appealing in the face of the dual imperatives of securing grant funding and publication, each has limitations. Reducing qualitative research to a list of technical procedures, however extensive, is overly prescriptive and results in "the tail wagging the dog." None of these technical fixes, in itself, confers rigour. They can strengthen the rigour of qualitative research only if they are embedded in a broad understanding of qualitative research design and data analysis. Otherwise we run the risk of compromising the unique contribution that systematic and thoughtfully carried out qualitative research can make to health services research.

1 Popay J, Williams G, Rogers A. Rationale and standards for the systematic review of qualitative literature in health services research. *Qual Health Res* 1998;8:341-51.
2 Hoddinott P, Pill R. A review of recently published qualitative research in general practice: more methodological questions than answers? *Fam Pract* 1997;14:313-20.
3 Seale C, Silverman D. Ensuring rigour in qualitative research. *Eur J Public Health* 1997;7:379-84.
4 Seale C. Quality in qualitative research. *Qual Inq* 1999;4:465-78.
5 Miller G, Dingwall R, eds. *Context and method in qualitative research*. London: Sage, 1997.
6 Harding G, Gantley M. Qualitative methods: beyond the cookbook. *Fam Pract* 1998;15:76-9.
7 Kuzel AJ. Sampling in qualitative inquiry. In: Crabtree BF, Miller WL, eds. *Doing qualitative research*. London: Sage, 1992:31-44.
8 Mays N, Pope C. Rigour in qualitative research. *BMJ* 1995;311:109-12.
9 Barbour RS. The case for combining qualitative and quantitative approaches in health services research. *J Health Serv Res Policy* 1999;4:39-43.
10 Frankland J, Bloor M. Some issues arising in the systematic analysis of focus group materials. In: Barbour RS, Kitzinger J, eds. *Developing focus group research: politics, theory and practice*. London: Sage, 1999:144-55.
11 Green J. Commentary: grounded theory and the constant comparative method. *BMJ*, 1998;316:1064-5.
12 Glaser B, Strauss A. *The discovery of grounded theory*. Chicago: Aldine, 1967.
13 Bryman A, Burgess RG, eds. *Analyzing qualitative data*. London: Routledge, 1994.
14 Melia KM. Producing 'plausible stories': interviewing student nurses. In: Miller G, Dingwall R, eds. *Context and method in qualitative research*. London: Sage, 1997:26-36.
15 Barry C, Britten N, Barber N, Bradley C, Stevenson F. Using reflexivity to optimize teamwork in qualitative research. *Qual Health Res* 1999;9:26-44.
16 Armstrong D, Gosling A, Weinman J, Marteau T. The place of inter-rater reliability in qualitative research: an empirical study. *Sociology* 1997;31:597-606.
17 Mauthner NS, Parry O, Backett-Milburn K. The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology* 1998;32:733-45.
18 Barbour RS. Mixing qualitative methods: quality assurance or qualitative quagmire? *Qual Health Res* 1998;8:352-61.
19 Richardson L. Postmodern social theory: representational practices. *Sociol Theory* 1991;9:173-9.
20 Mays N, Pope C. Assessing quality in qualitative research. *BMJ* 2000;320:50-2.
21 Bloor M. Techniques of validation in qualitative research: a critical commentary. In: Miller G, Dingwall R, eds. *Context and method in qualitative research*. London: Sage, 1997:37-50.
22 Atkinson P. Narrative turn or blind alley? *Qual Health Res* 1997;7:325-44.
23 Barbour RS. Engagement, representation and presentation in research practice. In: Barbour RS, Huby G, eds. *Meddling with mythology: AIDS and the social construction of knowledge*. London: Routledge, 1998:183-200.

(Accepted 9 November 2000)