Making words count: the value of qualitative research

ROS JOHNSON University of the West of England, Bristol, UK
JACKIE WATERFIELD Keele University, Staffordshire, UK

ABSTRACT In the current climate of evidence-based practice, physiotherapy is urged to prove its worth via rigorous scientific research. However, there are concerns that limited methodologies are used to explore complex therapeutic issues, and that the profession relies too heavily on quantitative research studies to provide its evidence base. Qualitative research methods are able to explore the complexity of human behaviour and generate deeper understanding of illness behaviours and therapeutic interactions. Nevertheless, there is still a sense of distrust of qualitative research, related to the challenge of evaluating both the quality and usefulness of findings derived through qualitative methods. This discussion paper explores these issues. It examines some of the most frequently used techniques aimed at ensuring quality and value in qualitative research, such as sampling, triangulation, multiple coding, respondent validation and the use of audit trails, as well as addressing reflexivity. Because of the pluralistic and interactive nature of qualitative inquiry, the criteria used to judge quality need to be appropriate to each piece of research and should provide evidence to help readers to evaluate the calibre of the study and its relevance to their own area of work.

Key words: physical therapy, qualitative research, rigour

INTRODUCTION

Some years ago, a number of papers appeared in physiotherapy journals promoting the desirability of qualitative research methods (Jensen, 1989; Parry, 1991; Shepard et al., 1993). As a profession, however, physiotherapy still appears to be reluctant to embrace qualitative research. Gibson and Martin (2003) searched four physiotherapy journals, from 1996 to 2001, and found that only four per cent of research papers were qualitative studies, compared with 30% in occupational therapy journals. This may be due, in part, to the physical nature of physiotherapy practice, although the World Confederation for Physical Therapy (WCPT) does emphasize the interpersonal qualities of therapists (WCPT, 1999). Also, historically, physiotherapy has sought to emulate the established profession of medicine, and its presumed scientific
Evidence-based practice is a driving force behind contemporary healthcare, and physiotherapy is urged to prove its worth via rigorous scientific research, in particular randomized controlled trials, systematic reviews and meta-analyses (Ritchie, 1999). However, there is concern that limited methodologies are used to explore complex issues (Culpepper and Gilbert, 1999) and that research designs devised for investigating the efficacy of pharmaceutical therapy are not necessarily appropriate for validating therapies that have human interaction at their core (Bithell, 2000).

It is widely accepted that to understand and respond to the multiple social and physical facets of health and disease, research methods must be able to explore the complexity of human behaviour beyond the scope of quantitative data and experimental models (Hammell and Carpenter, 2000), and that qualitative studies are a means of allowing evidence to be elicited from diverse populations and contexts (Barton, 2000). Nevertheless, there is still a sense of distrust of qualitative research, related to its perceived inability to produce useful and valid findings, beyond a supplementary role to quantitative research (Sandelowski, 1997).

This distrust may stem from insufficient understanding of the philosophical background for qualitative work, where emphasis is placed on meaning and experience, the relationship between the researcher and the researched, and the social construction of reality (Denzin and Lincoln, 1994). The way in which people interact and, through this, construct their social world is variable and context-specific (LeCompte and Schensul, 1999). Hence, qualitative research is underpinned by the belief that there is no one truth, and thus consensus is neither necessarily achievable nor a necessary goal.

The challenge appears to be that of evaluating both the quality and the usefulness of findings derived through qualitative methods within evidence-based practice (Hammell, 2001). It is an area of vigorous debate and evolving ideas within the sociological, nursing, medical and occupational therapy literature (Dingwall et al., 1998; Emden and Sandelowski, 1998; Barbour, 2001; Hammell, 2002). The purpose of the present article is to raise the issue of evaluating qualitative research within a physiotherapy arena. It examines some of the most frequently used techniques aimed at ensuring quality and value in qualitative research, and discusses their strengths and limitations.

**WHAT IS QUALITATIVE RESEARCH AND HOW IS IT EVALUATED?**

The term ‘qualitative research’ is used widely throughout the literature, often as a mutually exclusive opposite to ‘quantitative research’ (Kelly and Long, 2000). However, it is a poorly defined descriptor for a wide range of research activity that is underpinned by a number of different theoretical perspectives and methodologies (Carpenter, 1997; Smith-Sebasto, 2000). But, despite these differences, the emphasis for this type of research is on capturing in detail and/or depth something significant in the social world (Goodwin and Horowitz, 2002). Qualitative approaches also have the capacity to allow assessment of researchers’ subjective experiences and their impact on the setting (Gubrium, 1999, p. 8). Traditional research evaluation criteria were designed for quantitative inquiry (Carpenter and Hammell, 2000). Qualitative data are descriptive, unique to a particular
context and therefore cannot be reproduced time and again to demonstrate ‘reliability’ (Bloor, 1997). Instead of trying to control extraneous variables, qualitative research takes the view that reality is socially constructed by each individual and should be interpreted rather than measured; that understanding cannot be separated from context. Thus, qualitative data cannot be tested for ‘validity’ using criteria based on assumptions of objective reality and positivist neutrality (Shopard et al., 1993; Angen, 2000; Barbour, 2001).

In an attempt to equate to the quantitative research concepts of validity and reliability, numerous terms are used (Table 1) and promoted as desirable components of qualitative research (Murphy et al., 1998; Krefting, 1991). A number of strategies are frequently advocated to achieve these components, and cited as criteria by which to evaluate the worth of a qualitative study (Mays and Pope, 2000). However, the use of quality checklists or guidelines is controversial, and some authors challenge the appropriateness of specific methodological

### TABLE 1: Assessment of qualitative research: some terminology

<table>
<thead>
<tr>
<th>Quantitative research criteria</th>
<th>Qualitative research equivalents</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal validity</strong></td>
<td>Credibility</td>
<td>When multiple realities revealed by study participants are represented as adequately as possible, so that those who live the experience instantly recognize its description and interpretation (Sandelowski, 1996)</td>
</tr>
<tr>
<td></td>
<td>Truth value</td>
<td>Has the researcher established confidence in the truth of the findings for the subject … and the context in which the study was undertaken? (Krefting, 1991, p. 215)</td>
</tr>
<tr>
<td><strong>External validity</strong></td>
<td>Transferability</td>
<td>The ability to transfer working hypotheses between different settings (Murphy et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>Applicability</td>
<td>The degree to which findings can be applied to other settings or groups; the ability to generalize to larger populations (Krefting, 1991)</td>
</tr>
<tr>
<td></td>
<td>Fittingness</td>
<td>When findings from the study fit contexts outside it (Mulhall, 2000)</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>Dependability</td>
<td>Accounting for variability in the phenomena studied or changes in research design employed because the iterative process of inquiry (Marshall and Rossman, 1999)</td>
</tr>
<tr>
<td></td>
<td>Auditability</td>
<td>The ease with which the reader can check the pathway of decisions, or ‘audit trail’, taken by the researcher (Cutliffe and McKenna, 1999)</td>
</tr>
<tr>
<td><strong>Neutrality</strong></td>
<td>Confirmability</td>
<td>Neutrality of the data, not the researcher, so that others reach the same interpretations of meaning and significance as the original researcher (Chard and Gaberson, 2001)</td>
</tr>
</tbody>
</table>
criteria (Murphy et al., 1998; Barbour, 2001). Five strategies that are commonly used in qualitative research — sampling, respondent validation, triangulation, audit trail and reflexivity — are discussed in more detail.

SAMPLING

In qualitative research the sampling strategy does not seek to achieve statistical representativeness, but usually reflects the diversity within the study population (Popay et al., 1998) and the underlying theoretical framework. The sampling strategy is determined by the research question, the scale of the study and the type of material to be collected, and sample size is dictated by the specific purpose and context of the research. The sample must be sufficient to generate depth rather than breadth, and may comprise only a small number of participants (Sandelowski, 1995), or just one participant followed over a period of time.

Rather than probability sampling, purposive sampling is often employed (Curtis et al., 2000). This generally involves the deliberate selection of information-rich cases that will give rise to data relevant to the research aims (Patton, 1999). Researchers may explicitly and consciously select outliers, or ‘deviant cases’ that can help to generate deeper understanding of more general processes and interactions (Sim and Wright, 2000, p. 120; Barbour, 2001; ‘looking at what rarely happens in order to illuminate what usually happens’ (Gomm et al., p. 261). As an example of purposive sampling, Jensen et al. (1990), in their study to develop a conceptual framework of physical therapy practice, selected a sample of therapists working within one setting to enhance transferability, but included a range of novices and experts so as to explore the characteristics and boundaries of practice in this setting.

Qualitative studies frequently use a convenience sample, essentially drawn from those who happen to be in the right place at the right time, and often self-selecting volunteers (Sim and Wright, 2000, p. 120). This does not preclude theoretical representativeness, whereby theoretical insights and conceptualizations derived from the research may be applied elsewhere. However, the characteristics of the sample must be thoroughly and clearly presented so that the credibility and transferability, or fittingness, of findings to other appropriate settings may be assessed (Sim and Wright, 2000, pp. 113, 120; Malterud, 2001). Collecting data in the belief that more is better, rather than guaranteeing transferability may overwhelm the researcher, resulting in superficial analysis and thereby failing to elicit the unique meanings of the sample. Conversely, too little data may result in the loss of important perspectives and themes (Sim and Wright, 2000, p. 50; Malterud, 2001).

Whilst it is appropriate for the qualitative researcher to select an information-rich sample, Murphy et al. (1998) highlight that giving voice to the disadvantaged has a long tradition in qualitative research. More recently, feminist researchers have championed the concepts of advocacy, empowerment and emancipation within qualitative inquiry (Grbich, 1999, pp 53–57). However tempting it is to champion the underdog, ideological or political correctness may be criticized for unfair treatment of the powerful and privileged, leading to inadequate representation of the phenomena under study. Dingwall (1992) argues convincingly that social interactions and processes need to be understood and elucidated without taking sides, and calls for evidence of fair dealing within a study. That is, a research design that explicitly includes and lends equal weight to a diversity of
perspectives, and which seeks to understand both perceived villains and heroes (Dingwall, 1992), without portraying one sub-group as owning the ‘truth’ (Mays and Pope, 2000).

Skelton (1997) discusses a study on the education of low back pain patients in general practice. Interview data from patients and general practitioners revealed that doctors controlled both content and process. Further, ‘normal educational interactions became a vehicle for implicitly punishing patients for their (assumed) non-compliant behaviour’ (Skelton, 1997, p. 154). However, he goes on to discuss the apparent paradox evident in the data, of patients being willing participants in this medicocentric model, thus challenging the interpretation of domineering doctors and instead exploring socially acceptable health-related discourses.

RESPONDENT VALIDATION
Respondent validation, or member checking, occurs when participants review collected data or data analysis and confirm or challenge their validity. These responses are included in the research findings (Krefting, 1991). Cook and Hassenkamp (2000) collected interview data from seven individuals attending a back rehabilitation group and asked two of the participants to be key informants, with whom the researcher reviewed the emergent themes.

In the current climate of patient-centred care and consumerism, the concept of respondent validation is persuasive (Barbour, 2001). However, many authors highlight difficulties, such as people changing their minds over time, poor recall, the effect of the data collection process itself and the effect of new experiences in the intervening period (Bloor, 1997; Angen, 2000; Long and Johnson, 2000). Respondent validation of the emerging analysis is criticized for relying on the notion of a fixed ‘superior’ truth (Angen, 2000). Participants are most likely to seek themselves and their own reality in the data, whereas researchers aim to produce a conflation of many realities in a way that still represents each one, from specific situations to more general and underpinning theory (Cutcliffe and McKenna, 2002; Horsburgh, 2003).

Respondent validation may be regarded as a means to establish dependability of data and credibility of findings, but it also gathers additional original data, which require further analysis (Mays and Pope, 2000). This may lead to confusion rather than confirmation, and researchers may be faced with moral rather than analytical decisions when respondents retract or change data (Sandelowski, 1993). The value of seeking validation from participants may be questionable, but it is expected that they should recognize themselves and aspects of their world within the research findings (Krefting, 1991; Horsburgh, 2003). Cutcliffe and McKenna (2002) add that this recognition is derived through the dialogue and interaction that takes place within a reciprocal research relationship.

TRIANGULATION
Another strategy employed to enhance both dependability and credibility is triangulation. The term 'triangulation' has been borrowed from land surveying and navigation, where it refers to two landmarks being used to enable accurate location of a third point (Angen, 2000). In qualitative research, triangulation requires the researcher to examine data collected from different sources or by different methods or researchers, or findings derived from different analytical procedures (Shih, 1998; Seale, 1999). Doody and McAteer (2002) explored clinical reasoning of physiotherapists. They collected data from multiple sources: non-participant observation and audio-taping of physiotherapy sessions;
One purpose for which triangulation has been used is to seek confirmation of the findings of a study by seeking evidence to corroborate or refute an overall interpretation (Mays and Pope, 2000). The assumption is that each strategy used will expose different aspects of reality (Patton, 1999), and thus research has higher internal validity if more than one method, investigator or data source give rise to consentient findings (Bloor, 1997).

However, it may be costly and time-consuming to institute combined data collection strategies, such as individual and group interviews, observations and diaries, and it may be difficult to pool and compare different types of qualitative data (Sim and Sharp, 1998; Patton, 1999). Triangulation may then give rise to as many inconsistencies as agreements (Angen, 2000).

Triangulation, as identified earlier, is based on the concept of a fixed point, or the notion that an essential reality can be identified (Angen, 2000; Barbour, 2001), whereas qualitative research acknowledges multiple views of reality, each of equal validity (Bloor, 1997). According to Sim and Sharp (1998), some workers claim that by using triangulation the strengths and weaknesses of different research strategies will balance each other out and it is possible to arbitrate between different sources or accounts. However, when different sets of findings appear contradictory, it may be impossible to determine which, if any, have more validity than the others (Bloor, 1997). Similar or identical findings may be reassuring and confirming, but their absence does not provide grounds to discard unique findings (Cutcliffe and McKenna, 1999).

Another approach, and perhaps more in keeping with a qualitative philosophy, is the suggestion that triangulation may bring increased understanding or ‘completeness’ to a study (Redfern and Norman, 1994; Sim and Sharp, 1998), so that inconsistencies are considered to be parallel data sets that each contribute a partial yet equally valuable understanding of a phenomenon (Barbour, 2001). Thus, apparent contradictions can stimulate interaction and dialogue as they allow different, perhaps unconsidered, aspects to be examined, and they inspire further development of both research strategies and emergent theories (Bloor, 1997). Therefore, diverse sources can provide richer data, encourage reflexivity and help to increase the comprehensive understanding of phenomena, rather than provide a pure criterion-based test in which agreement among different sources confirms internal validity (Mays and Pope, 2000; Malterud, 2001).

Importantly, Shih (1998) argues that if triangulation is used the researcher must make explicit to the reader its form and purpose.

‘Multiple coding’ is a form of triangulation when two or more researchers analyse the data independently and compare their findings (Patton, 1999). It is a common procedure used to demonstrate rigour, and it attempts to establish a level of inter-rater reliability (Barbour, 2001). However, when the purpose is not to define a single reality, the credibility of an interpretation does not automatically increase with the number of people in agreement (Cutcliffe and McKenna, 1999). Also, other researchers may not have the same level of intimacy with the data and may be less able to judge whether phenomena have been adequately described (Angen, 2000).

Multiple coding should not be confused with peer review, which occurs when others review data and their analysis to evaluate whether the identified themes clearly
emerge from, or are grounded in, the data (Sandelowski, 1998). Peers are not experts brought in to ‘confer the validity stamp of approval on a project, but they can provide expert criticism’ (Sandelowski, 1998, p. 470). They may be other researchers or clinicians working in the same or a similar field, whose role is to ask questions, not to provide answers. Peer review tests the robustness and completeness of the emerging themes, and thus helps to validate the findings. It enables researchers to question and justify their interpretations, to ask different questions of their data and to explore different analyses (Sandelowski, 1998; Cutcliffe and McKenna, 1999).

May (2001) interviewed patients about their satisfaction with management of low back pain. He used framework analysis to elicit themes from the data and, striving to ‘ensure that the final themes identified were comprehensive and all-inclusive’ (May, 2001, p. 18), he discussed the findings with peers, thus enhancing dependability.

Peer review can also stimulate the search for contradictory evidence in the data, by highlighting alternative interpretations (Barbour, 2001). This assists in refining the analysis and in distinguishing it from purely emotional reactions of the researcher (Dingwall, 1997). Understanding the data may be enhanced by considering contradictory cases, in that they may broaden or transform meanings, or enable greater understanding of a particular sub-group within a given population (Patton, 1999). Hence, an explicit search for data that seem to contradict emerging meanings is considered a sign of rigorous analysis (Murphy et al., 1998).

Qualitative data represent meanings, and meanings are analysed through conceptualization. Thus, qualitative data analysis is a creative process that depends on the insights, background and informed value judgements of the researcher (Dey, 1993, p. 3; Patton, 1999). Whilst others may not share the interpretations, they should be able to follow how the researcher arrived at them (Koch, 1994). One means of achieving this is through an audit trail.

**AUDIT TRAIL**

The concept of an audit trail (sometimes called a ‘decision trail’) derives from a fiscal audit that looks for sources of error or deception by examining the way in which accounts are kept. An auditor reviews a sample of records and supporting documents in order to confirm accuracy (Koch, 1994). Qualitative data cannot be replicated to prove reliability but they can be audited, and a qualitative study should produce detailed, representative data and a pathway of decisions made during their collection and analysis that can be followed by others (Giacomini and Cook, 2000).

An audit trail requires clarity about the reasons for theoretical, methodological and analytic choices so that others can understand how and why decisions were made. This is said to enhance the reliability, or dependability, of the findings (Krefting, 1991; Koch, 1994). In other words, the trustworthiness of a study is enhanced by a satisfactory audit of the process by which its findings have been achieved as well as the end products: data, interpretations and recommendations (Krefting, 1991; Horsburgh, 2003). In the study by Jackson (1996), of patients’ experiences of hydrotherapy, she describes a clear pathway of decisions taken throughout the design, methodological and interpretative stages of the research, illuminating the context and thus enhancing both credibility and transferability of findings.

The term ‘thick description’ or ‘dense description’ is used when context, meanings and interpretation that elucidate the research process are provided, rather than mere
statement of facts independent of intentions or situations (Krefting, 1991; Popay et al., 1998). This richness allows the reader to gauge both the reliability of the data and the extent to which findings can be generalized to other settings (Krefting, 1991).

However, Cutcliffe and McKenna (1999) argue that a decision trail may stifle the recognition of hunches or ‘felt sense’ of emerging theory that a researcher would normally follow. It is also a concern that researchers may be unable to communicate these and other less tangible decisions that are based on intuition and even inspiration, but as part of research expertise may be perfectly reliable (Cutcliffe and McKenna, 1999). It is apparent that, whilst desirable, making sufficient material easily available may be a practical challenge and a threat to participants’ confidentiality (Horsburgh, 2003). Another problem encountered by those striving to leave a discernible decision trail in published studies is the limit on word count imposed by many academic journals (Horsburgh, 2003).

REFLEXIVITY

The researcher is an integral part of the world studied, and explicit reflexivity can clarify the audit trail (Cutcliffe and McKenna, 1999; Horsburgh, 2003). Rather than an attempt to attain an objective distance from the research, reflexivity seeks to recognize and value the researcher’s participation in shaping data and their analysis (Angen, 2000). Researchers cannot avoid taking value positions with them into the research process, just as readers will have their own perspectives. The resulting dialogues between perspectives and experiences can make research meaningful (Koch, 1994). The inevitable subjectivity is a resource rather than a source of error or bias (Sim and Wright, 2000, p. 134), and the researcher’s reflexivity lends plausibility to the findings.

Different researchers might reach different conclusions when examining the same data as part of a triangulation process. This does not have to mean invalid and unreliable research, but, rather, enhanced understanding of complex phenomena (Malterud, 2001). The influence of the researcher’s experiences, beliefs and personal history should be acknowledged (Krefting, 1991). Preconceptions are not the same as bias unless the researcher fails to identify them (Malterud, 2001). Indeed, Mays and Pope (2000) argue that personal and intellectual biases that are made explicit enhance the credibility of research findings. With the help of critical self-reflection, these subjective prejudices provide a basis from which further understanding develops (Angen, 2000).

Koch (1994) demonstrates the strength of reflexivity in her study of older patients’ experiences of acute care. As a qualified nurse, she collected data through participant observation and patient interviews on the wards. Koch makes explicit her own background and personal experiences that inevitably influenced her collection and interpretation of data. She reports the blurring of her roles as nurse and researcher and the unexpected effects of one upon the other, in particular the dilemma and ethical tensions of researching a setting dominated by inadequate care: ‘I disliked intensely being involved in nursing care that pushed me towards working in a non-caring routine’ (Koch, 1994, p. 982).

Thus, the researcher is central to the construction of valid findings, and Popay et al. (1998, p. 348) comment: ‘given the involvement of the researcher, the question is not whether the data are biased but to what extent the researcher has rendered transparent the processes by which data...’
have been collected, analysed and presented'. This reiterates the importance of an audit trail and adequate, thick, description. Reflexivity does not aim to demonstrate neutrality and objectivity, but makes explicit the researcher’s contribution to the entirety and complexity of the interpretative research process (Popay et al., 1998), not least as the ‘instrument through which the topic is revealed’ (Angen, 2000, p. 391). The impact of the researcher throughout the research should be explored during the discussion of limitations and strengths of the study, and this will contribute to the reader’s own assessment of its value and relevance (Mahurud, 2001).

**SUMMARY**

Qualitative studies enable therapists to reflect on their own personal and professional beliefs, to look further than treatment techniques to explain therapeutic effect, to gain richer insight into patients and their experiences, and to appreciate the importance of the patient–therapist relationship. Qualitative research has much to offer a profession dominated by one-to-one personal communication and therapeutic touch, yet despite its increasing contributions to the evidence base of health and social care, qualitative research is underrepresented in the physiotherapy literature. Gibson and Martin (2003, p. 353) suggest this is in part due to ‘mistaken attempts to evaluate qualitative studies according to the evidence-based hierarchy where the status of qualitative research is not acknowledged’. Professional interpretation of best evidence does not encompass the means by which to assess and incorporate qualitative studies.

In order to provoke dialogue and debate among physiotherapists, this article has discussed some of the strategies employed to enhance the rigour of qualitative research and has provided examples of recent research papers to illustrate these. Because of the pluralistic and interactive nature of qualitative inquiry, the criteria used to judge quality need to be appropriate to each piece of research (Hammell, 2002). Rather than a checklist of strategies, a transparent and auditable research process, systematic and fitting methods of data collection and analysis, thick description and reflexivity should provide the evidence for sound qualitative research and help readers to evaluate the calibre of the study and its relevance to their own areas of work.

**REFERENCES**

Carpenter C. Conducting qualitative research in physiotherapy. Physiotherapy 1997; 83: 547–552.
Johnson and Waterfield

Cutcliffe JR, McKenna HP. When do we know that we know? Considering the truth of research findings and the craft of qualitative research. *International Journal of Nursing Studies* 2002; 39: 611–618.


Pope J, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. Qualitative Health Research 1998; 8: 341-351.


Sandelowski M. ‘To be or to use’: enhancing the utility of qualitative research. Nursing Outlook 1997; 45: 125-132.


Sandelowski M. Rigor or rigor mortis: the problem of rigor in qualitative research revisited. Advances in Nursing Science 1996; 19: 37-44.


Address correspondence to: Jackie Waterfield, Senior Lecturer, School of Health and Rehabilitation, Keele University, Keele, Staffordshire ST5 5BG, UK (E-mail: j.waterfield@physio.keele.ac.uk).

(Submitted September 2003; accepted February 2004)