

Evaluating qualitative research in social geography: establishing 'rigour' in interview analysis

Jamie Baxter and John Eyles

A review of 31 empirical and eighteen substantive papers by qualitative social geographers mainly using in-depth interviews reveals little explicit reference to the principle(s) adopted to enhance 'rigour' and to ensure meaningful inference. Given the modest explicit discussion of evaluative criteria in these papers, a scheme from evaluation research itself is critically reviewed. A set of evaluation questions derived from this review and their application to an empirical piece of qualitative work frame an argument for a general set of criteria rather than rigid rules for assessing qualitative work. Such criteria can serve as anchor points for qualitative evaluation.

key words methods qualitative evaluation social geography

Department of Geography, McMaster University, Hamilton, Ontario, Canada L8S 4K1
e-mail: baxterjw@mcmaster.ca

revised manuscript received 19 August 1996

Introduction

This paper addresses the issue of evaluating the designs and findings of qualitative research in social geography through establishing a set of questions and criteria to be asked by and of such work. It is not our purpose to trace the history of the use of qualitative methods which may be seen as part of the humanistic and cultural turns in geography (see Eyles and Smith 1988; Ley and Samuels 1978; Smith 1984). We recognize that there is an apparent tension between the creativity of the qualitative research process – which implies contingent methods to capture the richness of context-dependent sites and situations – and evaluation – which implies standardized procedures and modes of reporting. Despite the problems implied by this tension, evaluation is critical if qualitative evidence or findings are to gain acceptance outside the community of practitioners, especially given the widespread recognition of the appropriateness of different approaches to investigation (see, for example, Clark *et al.* 1987). Nevertheless, the significance of authorship and the characteristics

of the researcher in shaping the interpretation of findings have become increasingly influential as writing in social and cultural geography seeks to re-present the subjects of inquiry (see Barnes and Duncan 1992; Livingstone 1992). And yet, once reflexivity is acknowledged, the important concerns over what and how we present has in some respects overshadowed the problems connected with obtaining and interpreting interview texts.

Three starting points guide our discussions of evaluation. First, despite problems in the research act (e.g. power relations), academic discourse is sufficiently different from lay accounts to merit attention in its own right; the analytic views of 'outsiders' are important for understanding. Secondly, we recognize that all qualitative researchers reflect actively on what they do and how they relate to their subjects. The researcher is herself a 'positioned subject' (Rosaldo 1989) – consciously thinking about what and where he is and what and how she does things. Such reflexivity is a strength for evaluating qualitative work, allowing a conscious deliberation of what we do, how we interpret and how we relate to subjects. Thirdly, as

interpretive geographers 'concerned with the understanding and analysis of meanings in specific contexts' (Eyles 1988a, 2), we set out to learn to view the world of individuals or groups as they themselves see it. In Schwartz and Jacobs' (1979) phrase, interpretative geography is in the reality reconstruction business, attempting to develop representations and constructions to describe the representations and constructions that take place within the social world. Although there are accessible literatures on the methods involved (e.g. Burgess 1984; Eyles and Smith 1988), little attention has been paid in human and social geography to judging what makes sense and what is plausible in the findings and the designs on which they are based. This is especially true of approaches such as in-depth interviewing and the conversational part of participant observation which involve talking/conversing with people on a face-to-face basis. More evaluative work has been undertaken on the analyses of texts (see Barnes and Duncan 1992; Cosgrove and Daniels 1988) which may require different kinds of evaluation (Eyles 1988a). These are not examined in this paper.

The purposes of this paper are three-fold: to review existing strategies for rigour and the criteria of evaluation currently employed in qualitative research in social geography; to provide a set of criteria and strategies which may be used for evaluation and a set of questions to guide implementation; and to provide a critique of an example of qualitative research using the evaluation questions.

It is important to be clear about our terms, in particular 'evaluation', 'rigour' and 'evaluative criteria'. 'Evaluation' refers to the judgement of empirical scientific work according to what has traditionally been called 'rigour'. In general terms, this is a process of critical appraisal to determine whether or not a study is worthy of attention. Evaluation takes place in at least three ways: by addressing the research methodology, methods and analysis (plausibility of research design); via the corroboration or refutation of research findings (plausibility of accounts); and through the fit with an existing body of literature or theory (appeal to interpretive community). While all are essential to evaluation, it is the first two that are central to this paper, although most appeals to plausibility are based on the third which may be expressed cynically as 'trust me, I agree with you'. 'Rigour' has come to mean the satisfaction of the conventional criteria of validity, reliability and objectivity within

quantitative research. Yet we must not forget the general context of rigour around the principles of academic integrity (see Frost and Stablein 1992; Knafelz 1994; Teich and Frankel 1992) including responsibility and honesty: dimensions of self-reflection, essential to qualitative research. An important dimension of rigour concerns the extent to which a piece of research is believable and hence worthy of attention, a notion similar to Lincoln and Guba's (1985, 290) trustworthiness: 'findings which are worth paying attention to, worth taking account of'. 'Evaluative criteria' are the basic principles used to guide the judgement of the integrity or trustworthiness of a study. They are guiding principles rather than rigid standards and can be satisfied only in a restricted number of ways. This is particularly important for assessing qualitative research which has relatively few standardized procedures for evaluation and whose practitioners are encouraged to be flexible and to utilize novel methodological and analytical procedures. Indeed, techniques range from passive observation and personal reflection to intervention, with a common theme of shared meanings and subjective understanding (Smith 1994). Qualitative researchers are encouraged to allow the research situation to guide research procedures in order that they may gain access to human experiences. Yet for the research to be evaluated, there must be clarity of design and transparency in the derivation of findings.

Current strategies for rigour in qualitative social geography

To discover how social geographers judge their qualitative findings, we carried out CD-ROM and manual searches¹ of the (English-language) geography literature for empirical studies and discussion papers published within the last twelve years which deal with qualitative research. We wanted to discover what qualitative geographers, and especially those using in-depth interviews, were doing to make their work plausible and deserving of attention.

Table I shows that, among the 31 empirical papers identified, the most common ways to ensure rigour are the provision of information on the appropriateness of the methodology, the use of multiple methods, information on respondent selection and the presentation of verbatim quotations. Most of the studies include a rationale for

Table I Strategies for establishing qualitative 'rigour' in geographic work, 1984–95

Source	^a Rationale for methodology	^b Multiple methods	^c Respondent	^d Interview quotations	^e Interview practices	^f Procedures for analysis	^g Immersion/lengthy fieldwork	^h Revisits	ⁱ Verification by respondents	^j Appeals to interpretive community	^k Rationale for verification
Bonnett (1992)		X	X	X						X	
Bridge (1994)		X	X	X		X				X	
Brown (1995)	X	X	X	X		X					
Cooper (1995)	X	X	X	X	X		X	X	X		
Cooper (1994)		X	X	X	X		X	X			
Droogleever Fortuijn and Karsten (1989)	X	X	X	X		X					
Dyck (1989)	X	X	X	X	X		X			X	
Eyles <i>et al.</i> (1993)	X		X	X	X	X					X
Eyles and Perri (1993)	X		X	X		X		X	X		X
Eyles and Donovan (1986)	X		X	X		X					X
Fernandez Kelly (1994)	X	X	X	X	X	X	X			X	
Gregson and Lowe (1995)			X								
Harrison and Burgess (1994)		X		X		X					
Herod (1991)		X		X						X	
Hewitt (1994)	X			X	X						
Katz (1991)	X	X	X		X			X			X
Leckie (1993)		X	X	X							
Mackenzie (1992)	X	X	X	X						X	
McDowell (1994)	X	X	X	X	X					X	
McDowell and Court (1994)	X	X	X	X	X					X	
Newton (1995)	X	X	X			X					
Porteous (1988)	X	X	X	X	X		X	X	X		X
Rollinson (1990)	X	X	X	X							
Rowe and Wolch (1990)	X	X		X	X		X			X	
Rutherford (1995)	X	X	X	X		X				X	
Shute and Knight (1995)	X		X	X	X	X					
Valentine (1995)			X	X		X				X	
Valentine (1993)	X		X	X						X	
Valentine (1989)	X	X	X	X						X	
Wilson (1993)	X	X	X	X			X			X	
Winchester and Costello (1995)	X				X					X	
Totals	23	22	26	27	13	12	7	5	3	15	5

Notes: ^aQualitative methods are argued to be the most (or only) appropriate way to address the research question; ^bmore than one method used for studying the problem; ^ca description of the group(s) of respondents (e.g. number and gender ratio is given); ^dthe words of the respondent may be read; ^edetails of how interviews were conducted (e.g. use of interview schedules and autobiography are provided); ^fa description of how data were converted/condensed into theoretical constructs is given; ^git is argued that long field seasons (self-proclaimed or stated to exceed one year) develop rapport with respondents and/or enable deep understandings of the research situation; ^hrevisits to respondents are made usually (but sometimes unstated) to clarify meanings and build rapport; ⁱrespondents were contacted to verify interpretations/meanings; ^jan existing theory is supported (or refuted) by the findings, i.e. there is more than reference to the literature; ^krationale for showing that there is agreement between constructs/interpretations and the meanings held by respondents is provided.

using a qualitative approach. For example, [Brown \(1995, 162\)](#) claims that ethnography

can reveal such geographies [of AIDS] in spite of the distances perpetuated by spatial science. The linchpin of my argument is that sociospatial distance sacrifices geographic knowledge; that ethnography overcomes distance.

whilst [Droogleever Fortuijn and Karsten \(1989, 366\)](#) suggest

It is evident that because of the selection criteria mentioned above the research population is not representative for all persons whose daily activity pattern is characterized by combining tasks. Instead of making generalizations, the purpose of this study is primarily to obtain more in-depth knowledge about the different ways people combine activities and about the constraints they meet. For the same reason we have chosen a mainly qualitative method of data collection.

But while there are numerous ways to undertake 'ethnography' and use 'qualitative methods', they are not self-explanatory. Elaboration is needed to clarify the ways in which the methodology and methods are carried out to achieve things like 'overcoming distance' and obtaining 'in-depth knowledge'.

All but eight studies involve a combination of methods and many of these use in-depth interviews together with participant observation or textual analysis. The use of multiple methods enables triangulation (see below) but simply using two or three different methods does not necessarily guarantee more rigorous results. Few investigators comment on the reasons for using more than one method. Does each method address the same or different research questions and what are the implications for evaluation?

Most of the papers include a brief mention of who was interviewed. However, only in ten papers² do the researchers mention how respondents were recruited and several do not indicate how many people were interviewed. Sample size is relevant for qualitative researchers. It is the basis of discovery and description and, while an N of 1 can be easily justified ([Dukes 1965](#)), a rationale should be provided. Further, a description of respondent characteristics is critical since experiences crucial to the research question may be unnecessarily overlooked. They offer an indication of who is allowed to speak and, of equal importance, who is not.

All but four of the papers include (what appear to be) verbatim respondent quotations. Quotations

are important for revealing how meanings are expressed in the respondents' own words rather than the words of the researcher. However, these vary considerably in detail in the papers reviewed. For those papers containing quotations, the number ranges from as few as one ([Brown 1995](#)) to over 100 ([Eyles and Donovan 1986](#)); some reports provide quotations of considerable detail with less investigator commentary³ while others provide shorter quotes with considerable commentary⁴. Why? We are uncertain, although we share the disquiet of [Bryman \(1988\)](#) and [Silverman \(1993\)](#) on the anecdotal nature and assumed representativeness of such accounts. While there need not be a model for the size and number of quotations, it is reasonable to expect some discussion of why particular voices are heard and others silenced through the selection of quotes.

There are at least seven other strategies for demonstrating rigour (but these are mentioned in less than half of the 31 articles): details of interview practices, discussions of the procedures for analysis, immersion/lengthy fieldwork, revisits to respondents, verification by respondents, appeals to interpretive communities and the provision of a rationale for verification (validity) of the findings.

Research practices which may enhance rigour include the use of standardized interview guides⁵ and attention to the power relations involved in research interviews.⁶ The guides allow for inter-interview comparisons of emergent phenomena, while detailing the power relations between the interviewer and interviewee helps account for the ways that interview texts are constructed. However, the reader is often left to interpret for themselves statements such as

We are white and middle class. One of us is thirty-something, the other forty-something, and for the purposes of the interviews we wore 'professional suits' that blended in with the clothes of our women respondents. The respondents, bar one, were white; all but six were under 40. They were socially adept and articulate. ([McDowell 1994, 665](#))

The implications should be stated, since similarities between interviewers and interviewees may, for example, foster or stifle interview conversations.

Similar comments may be made about procedures for analysis. We argue that one of the most important issues for analysis is the manner in which interview conversations are constructed into

theoretical concepts. Although most qualitative reports display verbatim quotations, there is rarely a discussion of how particular quotations are selected for presentation from the range of available interview texts. Since there are no standard procedures for analysing interview texts (but see Willms *et al.* 1990), it is necessary to elaborate how data get transformed into concepts/theory(ies) to show readers whose meanings are represented and why.

Lengthy fieldwork or immersion in the research context of interest, evident in seven of papers, is a traditional way of lending credence to the theories that emerge from qualitative interviews. However, while immersion may lead to more sensitivity to the subtleties of meaning in the group(s) being studied, it can also threaten the credibility of findings. For example, the researcher may 'go native', whereby the study group, and not the community of researchers, becomes the main group with whom the researcher identifies.

One of the main threats to ensuring qualitative validity is the misinterpretation of meanings expressed through interview conversations. Re-visits to respondents, undertaken by only five of the researchers, is one way to check interpretations. Revisits are often used to cover new ground rather than to verify what was covered in previous interviews and so it is important to report when they are used to confirm the researchers' interpretations. For example, Cooper (1995, 352) points out

When I involved these individuals in the final selection of illustrative passages of their discourse, they were able to identify with the context in which I intended to use the data. Similarly, I have occasionally introduced themes and questions into the interviews, which had previously been absent from the adolescents' discourse. Through these interviews, where themes the adolescents considered unimportant and important were discussed, a more comprehensive analysis became possible.

In only five cases is there explicit mention of the rationale for verification of the findings. Thus Katz (1991, 495–6) states

The population was selected based on a full enumeration and survey of the entire village. This, along with the diversity of the methodology I developed, the standards I maintained in working with the children, and comparisons of my results with those of the few existing relevant studies, give me confidence in the validity of the information produced.

Such statements reveal what, for the author, are the things about the study that make the findings worthy of attention. Without these clues, the reader may judge work unfairly according to criteria which may not be as relevant to the research. For example, the lack of interview quotations in Katz's study does not necessarily detract from the value of the findings if this represents a trade-off for helping to maintain a rapport with the children and parents with whom she spoke. Such trade-offs, however, should be explicit.

Appealing to an existing body of literature – the 'interpretative community' – our third component of evaluation, is often the overarching and implicit strategy for organizing some research projects. In such cases, the quotes and interpretations are usually forwarded as evidence for the adequacy or lack thereof (e.g. Bridge 1994) of concepts and theories formulated elsewhere. For example, Bonnet (1992), Herod (1991) and Rutherford (1995) seem to use literatures on racism, deindustrialization and restructuring respectively as a good proportion of the support for the trustworthiness of their findings. Reference to a body of literature is a necessary but insufficient condition for plausibility. Verification based solely on appeals to conventional wisdom does not necessarily lead to rigorous findings and it may be counterproductive to the development of new wisdom (often a goal of qualitative research). Valentine (1993), however, not only incorporates feminist literature but, by including both a commentary on her rationale for respondent recruitment and the use of detailed quotations, she goes a long way towards demonstrating the rigour of her findings.

The discussion of the various strategies in Table I will be revisited and reformulated in the next section of the paper which develops a checklist of general criteria for establishing trustworthiness. However, together the 31 papers suggest that there is insufficient mention of the practices (or techniques) and principles for establishing rigour. This is not to suggest that each rigorous study should have a complete row of 'Xs' in Table I since reasonable argument may be made for focusing on only a few (e.g. Katz 1991). But, for the most part, such arguments are rarely evident. Readers are not given much basis for judging the merits of some of these studies since there is scant mention of the principles of good qualitative work to which they hope to appeal. This does not necessarily mean that these studies are not rigorous, merely that

researchers may not be reporting fully on what they are in fact practising. Qualitative papers are often edited substantially to satisfy journal size limitations and so details of the research process may be forfeited in order to focus attention on the particulars of the concepts/theory which emerge from the data. Eliminating sections on research practices not only leaves readers wondering about the trustworthiness of findings but also robs us of helpful clues to ways of doing good research. We submit that there is a need – with no parallel in quantitative research with its relatively well-established and agreed strategies⁷ – for such content.

Current evaluation criteria used in qualitative social geography

To assist in discovering more about the principles/criteria used by qualitative social geographers to guide their work, we turn to papers which focus on qualitative methods and methodology rather than research findings. We found eighteen such papers.⁸ Despite some consideration of a qualitative notion of validity (see Katz 1991), there has been relatively little discussion of the principles which do/should guide qualitative research. Calls for more explicit principles to guide qualitative research in geography date back at least to Smith (1984) and, more implicitly, to Ley (1974) and Ley and Samuels (1978). In her discussion of the numerous Chicago-school ethnographic studies of the city, Smith (1984, 357) protests 'there are rarely any explicit "rules of the game" lending structure to the monographs they produced'. She argues that the enthusiasm for providing novel and rich accounts of life in the city perhaps overshadowed any concerns about evaluation. It may even be argued that 'rules' of any sort may stifle the creativity which helps to make all types of qualitative work so vibrant.

Much attention has been focused recently on the interview process and the need for reflexive consideration of how knowledge is produced through the social relations of the interview: a key element in the postmodern and new cultural turns.⁹ For example, Pile (1991) focuses on the power relations between researcher and researched, and the inter-subjective nature of the interpretative, interview-based research process. Further, he claims that 'by assuming that the subject can make as much sense as the researcher, we can refuse the objective of

capturing the *other*' (*ibid.*, 467, *emphasis added*). But, if there is no 'capture', how can we re-present? Also, if we do not or cannot re-present, are we rendered speechless? What the voice says is as important as recognizing the existence of voice. We prefer not to 'refuse capturing the other'; rather we want to ensure that the other's voice is heard alongside that of the researcher. Deliberations on epistemology, ontology and practice, while necessary for enriching the findings of qualitative research, are not sufficient on their own for evaluation purposes. Principles for evaluation are needed to bridge the gap between the philosophical concerns of a qualitative epistemology and ontology on the one hand and the practice of qualitative methods on the other. These issues are revisited in the discussions of disciplined subjectivity, bracketing and member checking in the next section.

One criterion of evaluation – validity – has received considerable attention. Eyles (1988a, 11) promotes a meaning-centred version of the 'scientific' notion of validity whereby 'principles of validation are internal to the discourse itself' and 'interpretations must be justified in terms of the presented evidence, so much depends on the coherence of argument and the reason, consistency and honesty of the theorist'. Eyles' qualitative interpretation of the conventional (quantitative) face validity refers to the plausibility of connections between data and concepts which appeal to common sense and consensus. He distinguishes the qualitative interpretation of face validity as involving data-to-concept links which not only make sense to scientists but also to the lay people on whose experiences interpretations are based. This notion is advanced by engaging Schutz's (1962, 44) postulate of adequacy which claims that a construct

must be understandable for the actor himself [*sic*] as well as for his fellow-men [*sic*] in terms of common-sense interpretations of everyday life. Compliance with this postulate warrants the consistency of the constructs of the social scientist with the constructs of common-sense experience of the social reality.

In this way, scientific constructs or meanings (second-order typifications) are validated as adequate by lay constructs (first-order typifications). Similarly, Rose (1982) suggests that it is important to distinguish participant concepts (terms used by the respondents themselves) from

theoretical (researcher-derived) concepts. In order to satisfy the postulate of adequacy, the former are preferred to the latter where possible.

Jackson (1985, 170) claims that social scientists are currently wedded to a narrow (quantitative) definition of validity and that 'alternative bases of validity are rarely even considered'. In keeping with Eyles' notion of face validity, Jackson (*ibid.*, 171, *emphasis added*) claims,

The criterion of validity most appropriate to the case study method concerns the *logical* relationship between characteristics rather than their representativeness or typicality. The logicity of this connection is in turn to be judged from the adequacy with which the wider social context is specified. The extent to which generalizations may be made from case studies depends upon the adequacy of the underlying theory and the whole corpus of related knowledge of which the case forms a part rather than on the particular instance itself.

While it is necessary to pay attention to Jackson's concern with the 'logical relationship between characteristics', he seems to confound validity and representativeness which, we will argue, should be separate considerations.¹⁰

McDowell (1992a) promotes a different approach to validity in defence of her claim that Schoenberger (1991) (unsuccessfully) tries to apply the traditional principle of validity to qualitative findings. McDowell maintains that validity is not so much a property of interpretations as it is the collective agreement of intended audience(s) that the interpretations are convincing. This seems to align well with Eyles' promotion of Schutz's postulate of adequacy – that the scientific and lay communities must be convinced by these interpretations – but she perhaps understates the importance of the 'property of interpretations' and the role of the interpretive community. We argue that interpretations which are rigorous, rather than merely compelling/novel, are the most 'convincing'. Smith (1984) makes similar claims in her discussion of 'logical inference' – 'the process by which the analyst draws conclusions about the essential linkage between two or more characteristics in terms of some systematic explanatory schema' (Mitchell and Draper 1982, 200). While she does not refer to this as 'validity', the connection between data (characteristics) and theory (explanatory schema) is evident.

For us, there are two common threads in these discussions. First, it is important to use logical

inferences which emphasize the relationship between phenomena and constructs of these phenomena. Secondly, interpretations should appeal to scientists, those researched and a wide array of lay people. The review of the empirical papers indicates that if there is any attention being paid to qualitative validity, it is to the first issue rather than the second, since only three papers make mention of checking their findings with those researched. While the discussion of validity will be taken up below, it seems that the key to 'validity' is clarity – making the implicit 'rules' explicit.

Clarity about evaluation criteria is evoked by Athens (1984, 260, *emphasis added*) who states

whenever works in science are judged, the evaluator, whether or not admitted, makes his or her evaluation on the basis of *some criteria*.

Although, as Athens argues, there may not be much explicit discussion of qualitative evaluative criteria, there are implicit principles used to guide and evaluate such work. However, as Merton (1968, 71–2) states,

the sociological analysis of qualitative data often resides in a private world of penetrating but unfathomable insights and ineffable understandings; however, science is public, not private.

We argue that a step towards making qualitative analysis more 'public' is to be explicit about the principles for making this work rigorous. Being forthcoming about these criteria will better equip those who do not traditionally work within the qualitative paradigm to judge its approach and findings and, perhaps more importantly, these criteria will be made public for constructive scrutiny and debate.

Towards a set of criteria for establishing qualitative rigour

Table II sets out Lincoln and Guba's (1985) evaluation criteria. These are widely cited among qualitative researchers outside geography (see Burns 1988; Krefting 1990; Sandelowski 1986) and thus deserve to be critically reviewed here. The four criteria in the table extend the somewhat limited discussion of evaluative criteria in social geography which seems to focus mainly on only one: validity/credibility. 'Strategies' for enhancing rigour are also listed against each criterion. The table allows the reader to connect the criteria to

Table II Criteria for evaluating qualitative research

<i>Criteria</i>	<i>Definition</i>	<i>Assumptions</i>	<i>Strategies/practices to satisfy criteria</i>
Credibility	Authentic representations of experience	Multiple realities Causes not distinguishable from effects Empathetic researcher Researcher as instrument Emphasis of the research endeavour	Purposeful sampling Disciplined subjectivity/bracketing Prolonged engagement Persistent observation Triangulation Peer debriefing Negative case analysis Referential adequacy Member checking
Transferability	Fit within contexts outside the study situation	Time and context-bound experiences Not responsibility of 'sending' researcher Provision of information for 'receiving' researcher	Purposeful sampling Thick description
Dependability	Minimization of idiosyncrasies in interpretation Variability tracked to identifiable sources	Researcher as instrument Consistency in interpretation (same phenomena always matched with the same constructs) Multiple realities Idiosyncrasy of behaviour and context	Low-inference descriptors, mechanically recorded data Multiple researchers Participant researchers Peer examination Triangulation, inquiry audit
Confirmability	Extent to which biases, motivations, interests or perspectives of the inquirer influence interpretations	Biases, motivations, interests or perspectives of the inquirer can influence interpretation Focus on investigator and interpretations	Audit trail products Thick description of the audit process Autobiography Journal/notebook

Source: Lincoln and Guba (1985)

philosophical concerns about epistemology, ontology and methodology as well as to the research process itself. Note that plausibility of design and accounts are often dealt with simultaneously since many strategies employ overlapping criteria (e.g. triangulation). This is similar to quantitative research in that a validity issue is also one of reliability and vice versa. Yet in qualitative research there are fewer conventional procedures and more resourceful ones. In fact, qualitative researchers tend to applaud new and innovative ways for making work rigorous.

Credibility

The most important principle for guiding qualitative studies is the notion of credibility. This may be defined as the degree to which a description of human experience is such that those having the experience would recognize it immediately and those outside the experience can understand it (Lincoln and Guba 1985). Credibility refers to the connection between the experiences of groups and

the concepts which the social scientist uses to recreate and simplify them through interpretation. The parallels with some of the recent discussions surrounding 'validity' within geography are evident. In particular, Eyles' (1988a) and McDowell's (1992a, 1992b) concern that interpretations be credible both to those within the research endeavour (including investigators and respondents) and to the wider scientific and lay communities is commensurate with the notion that 'those having the experience would recognize it immediately'. It is apparent that research practices should involve strategies for returning interpretations to respondents for commentary (and perhaps revision).

Credibility is based on the assumption that there is no single reality but rather multiple realities, mentally constructed by ourselves. It is not confirmation that is required from respondents as much as a commentary from them on the plausibility of the interpretations offered. There may be disagreements between respondents and scientists (Borland 1991) or the former may deny, repress, falsify or

otherwise give only partial accounts of their experiences (Miles and Crush 1993). Since outsider accounts may well differ from insider accounts, Eyles and Perri (1993) report unease with some of their comments on the Italian-Canadian family studied by them, although the 'insiders' were asked to confirm that (academic) accounts made sense. Hence, the goal of the researcher is to represent adequately the realities of groups in such a way that not only does the scientific community but also the people who constructed the reality in the first place understand the (re)construction of that reality. The problem of participants providing only partial accounts of their experiences and meanings is more problematic as it is often difficult for the researcher to detect gaps between what is reported and what 'actually occurred'. Since realities are assumed to be multiple and flexible, it is also assumed that there is no way to distinguish between things like 'causes', 'effects' and 'truth' within the social world. This serves to distinguish credibility from its quantitative counterpart: internal validity.

In order to enhance credibility, researchers focus on respondent selection procedures, interview practices and strategies for analysis. Patton (1990) identifies two major strategies for recruiting respondents: random sampling and purposeful sampling. Random sampling, based on statistical representativeness, is used only rarely in qualitative research (e.g. Eyles and Donovan 1990) but may be employed if there is no conceptual reason for directing attention to particular informants at the outset of the research process. Purposeful sampling – the strategy used most often by qualitative researchers – stresses the search for 'information-rich cases'. Such respondents are at ease and talk freely with the researcher such that a great deal can be learned about the research question. Sample size is determined largely by the need to involve as many experiences as possible in the development of a conceptual framework/theory. Recruitment then often occurs until 'redundancy' or 'saturation'; that is, until no new themes or constructs emerge. Thus credibility need not be threatened by low sample sizes. Yet there is still a requirement for qualitative researchers to be mindful of self-selection 'biases' which may come from certain strategies like snowball-sampling. While bias may be used to advantage, the skewing of sample characteristics and types of questions answered and not answered needs to be recognized and

reported (see King *et al.* 1994). Detachment is often seen as being more problematic than observer bias and may itself lead to indifferent responses (LeCompte and Goetz 1982).

Patton (1990) identifies sixteen purposeful sampling strategies. Not all are equally useful. For example, 'stratified purposeful sampling' is considered useful in ensuring that all sub-groups within a research setting are given a voice so that comparisons can be used to construct commonalities and differences in interpretations across groups. On the other hand, 'convenience sampling' – interviewing only those who are easy to access (e.g. all people in one amiable community group) – is problematic since easy contacts are not necessarily the most informative contacts and such people may have only limited capacity to comment on issues relevant to the research question. Further, Wax (1971) urges against choosing participants merely on the basis of personal preference and to select respondents as broadly as possible. In this sense, the qualitative researcher will often try to establish the range of possible respondents (an ongoing process) by selecting proportionally from all groups/types.

Thus sampling is important. At the very least, the essential characteristics of qualitative sampling strategies should be considered, even if the sample design – although preconceived (at least enough to answer the question 'where and with whom do I start?') – is flexible and evolves as the study progresses; who and what comes next depends on who and what came before. The sample is selected serially or adjusted continuously or 'focused' by the concurrent development of theory. Selection continues to a point of thematic saturation and sampling includes a search for negative cases in order to give developing theory greater breadth and strength (Kuzel 1992; Lincoln and Guba 1985).

As indicated in the previous section, the practices of interviewing have probably received the most attention in recent papers on qualitative methods in social geography. As the researcher is the active instrument in qualitative research, interviews raise questions about the role of the 'author' and how her/his characteristics become a formative influence upon them. More generally, it is the interviewer's skill at developing a rapport with respondents and his/her ability to use this to develop information-rich conversations which shape the data gathered. But, beyond this, power relations and the presentation of self in the interview are crucial determinants. Age, gender,

ethnicity and other outward appearances can potentially affect how respondents react in the interview (Pile 1991). Vigilance over these issues has long been advocated by survey researchers (Moser and Kalton 1981) and field workers (Burgess 1984). The practice of being mindful of one's own ethnocentricity and biases has been called 'disciplined subjectivity' (Erickson 1973) and 'bracketing' (Lincoln and Guba 1985). Further potential interviewer effects arise from the researcher's socio-demographic characteristics in relation to those of the respondents (a relationship usually represented through a biographical sketch).¹¹

Many of the strategies for conducting credible analyses are closely wedded to interview practices. In particular, both share the problem of dealing with the implications for credibility of analyses of social relations between the researcher and the researched. Lincoln and Guba (*ibid.*) propose three 'preventative' techniques for dealing with this: prolonged engagement, persistent observation and triangulation. Prolonged engagement involves spending sufficient time in the field to build trust and rapport with the respondents, to learn the 'culture' of the relevant group(s) and to investigate for possible misinformation/distortions introduced by self or respondents. It may result, however, in the problem of going native discussed earlier (see also Vidich *et al.* 1964), although the so-called 'contamination' of going native can provide the basis of vibrant accounts (see Plummer 1983).

Persistent observation is complementary to prolonged engagement in the sense that the latter provides scope while the former provides depth. While prolonged engagement involves being aware of the 'multiple influences and mutual shapers and contextual factors' (Lincoln and Guba 1985, 304), persistent observation involves focusing on the 'things that count' in terms of the research questions being asked. It is closely linked to purposive sampling since both may involve seeking out a diversity of (informative) respondents to ensure that relevant experiences are not omitted.

Triangulation is one of the most powerful techniques for strengthening credibility. It is based on convergence: when multiple sources provide similar findings their credibility is considerably strengthened (Knaf and Breitmayer 1989; Krefting 1990). Denzin (1978) suggests that there are four major types of triangulation involving the use of multiple sources, methods, investigators and

theories. Source triangulation, the most common of the four, refers to the use of more than one report from a data set to corroborate a construct. One of the most frequently used (often implicit) forms of source triangulation is the use of quotations from several different respondents (e.g. Eyles and Donovan 1986). Method triangulation involves corroboration of constructs based on information derived from at least two different methods. Various combinations of qualitative methods are often used and, increasingly, qualitative and quantitative methods are combined in the same study (e.g. Harrison and Burgess 1994; Leckie 1993). To counter methodological eclecticism, Fielding and Fielding (1986) suggest that the methods chosen should make sense within one theoretical perspective. Investigator triangulation involves multiple investigators investigating the same phenomenon and comparing results. This may work well if investigators are part of the 'same team', such that they are, in effect, looking at the same phenomena with similar perspectives, e.g. graduate students and their supervisor(s). Problems may arise, however, when the subtle nuances of the interview (e.g. body gestures) are known only to the researcher who conducted the interview and helped construct the interview text. Variations in interpretation resulting from differential access to what happened in the interview can, however, be resolved by negotiation between researchers. Theoretical triangulation is usually reported as epistemologically unsound and counter to appeals to the interpretive community.

There are also analytical techniques which may be used after data have been collected. We highlight four: peer debriefing, negative case analysis, referential adequacy and member checking. Peer debriefing involves exposing data and interpretations to a respected colleague in order to point up possible sources of misinterpretation and the 'suppression' of themes or voices that do not 'fit' the 'storyline'. There are dangers, however, that 'unresolvable' sources of disagreement may arise or that one person may defer to the other on the basis of unequal power/authority relations (Risteen Hasselkus 1991). Negative case analysis involves a largely inductive process of constantly revising an hypothesis by comparing it with all interview texts until it accounts for all known cases. This process serves to explore numerous dimensions of a theme in order to make it robust and is particularly recommended for the development of constructs

intended for central theoretical status (Kidder 1981). Lincoln and Guba (1985) assert that accounting for all negative cases is 'too rigid' a criterion to satisfy, since some cases may be so obscure that they are of little conceptual/theoretical consequence. Referential adequacy is the practice of verifying the constructs developed through an interpretation of the bulk of the data by the subsequent analysis of a selection of data which has been archived (i.e. not part of the original analysis). Member checking is arguably one of the most important strategies for enhancing credibility since it involves checking the adequacy of analytic categories/constructs/hypotheses with members of the group(s) from which the data were obtained. But Hammersley (1992, 65) is cautious about this strategy:

To assume that respondents can validate or even falsify accounts in some definitive way is to forge the social character of the relationship between researcher and participants and to assume that they have privileged access to the truth. Neither of these assumptions is sustainable.

Whilst respondents do not have privileged access to the truth, they do have privileged access to their own opinions and meanings. It is the adequate representation of these that should be the goal for member checking. The appeal of this strategy is that it is implicit in the qualitative notions of validity (in particular, Schutz's (1962) postulate of adequacy), whereby interpretations are more credible if they are meaningful for both academia and the group studied. There is no definitive procedure for returning information to the respondents for such a credibility check and there is also the risk of what Borland (1991) calls 'interpretive [*sic*] conflict', whereby the participants will largely disagree with the researcher's interpretation of the interview text. We agree with Borland who suggests that checking should be done in the spirit of an 'exchange of ideas' and, if we do not check our interpretations with participants, we are in danger of merely fitting data into the preconceived theories/frameworks with which we are comfortable. There is also an ethical imperative to let participants know how their interviews are being used. It is increasingly the case that 'contracts' with groups and communities that agree to be researched include feedback of some sort (if wanted). Thus reporting back occurred in research on the effects of the Hagersville (Ontario) tyre fire

(Baxter *et al.* 1992). These 'contracts' are now often demanded by ethics committees that sanction research on human subjects.

Ethics aside, a key substantive question remains. How much analytical refinement is appropriate for the information which is to be submitted for respondent review? For example, the information may be interpretations of single interviews (low level of refinement) or interpretations of multiple interviews (high level of refinement). While respondents may feel that they are 'qualified' to comment only on their own transcripts, the higher-level interpretations are often more meaningful for theory development. Porteous (1988) used member checking by giving each respondent a copy of chapters with their own quotes highlighted. In this way, not only does the author verify the highest level of interpretation, he shows each respondent how their comments fit into the analysis.

Transferability

Transferability refers to the degree to which findings fit within contexts outside the study. Elements of research produced in one context may be transferred to others. However, such a strategy is clearly dependent on the correspondence between 'sending' and 'receiving'¹² contexts. It is analogous – in principle at least – to the more familiar notion of generalizability or external validity. The qualitative researcher is rarely as concerned about transferability as she is about the credibility of the findings. Within the qualitative paradigm, experiences and meanings are assumed to be largely bound to the time, people and setting of the particular study. Most qualitative researchers focus on one context in order to discover, describe, hypothesize and otherwise reconstruct the things that are meaningful to the people within it. It is assumed, therefore, that statements will be idiographic rather than nomothetic and it is perhaps not surprising that none of the empirical papers reviewed here makes any claims about the transferability (or generalizability) of findings.¹³

This does not mean, however, that qualitative research need pertain only to the specific cases under investigation. Meanings are often shared by many individuals. While a qualitative study may reconstruct meanings as they apply to the experiences of only a very small sub-group, it is possible that these experiences may be common to a larger group. However, there is nothing built into most qualitative research designs to assess the degree to

which findings are transferable beyond the single case. [Lieberson \(1992\)](#) points to the problems of, and relatively rare circumstances in which, 'generalized' conclusions may be drawn from case studies. Since most qualitative researchers adhere to the notion of idiographic knowledge, claims about transferability are rarely made (at least convincingly so) by the 'sending' researcher, who may enhance transferability by careful ascription of findings to specific sub-groups in the study. [LeCompte and Goetz \(1982\)](#) also argue that the multi-site study is one strategy for increasing the likelihood that findings will transfer.

It is, however, important to recognize the qualitative researcher's responsibility with respect to transferability. There is an onus on qualitative researchers to provide data which allow for transferability, rather than necessarily demonstrating its existence, by providing the database on which such judgements may be made by others. The original researcher must describe the study context as completely as possible because, at root, transferability involves the degree to which constructs are meaningful to other groups (as yet unstudied or not yet compared with the original group). Detailed, thick description ([Geertz 1973](#)) – as a methodological as well as interpretative strategy – of how constructs/hypotheses are developed and what they mean, will be of use to the researcher or layperson who wishes to determine the degree to which they may be transferred to other contexts.

Dependability

Dependability is the degree to which it is possible to deal with instability/idiosyncrasy and design-induced change. [Kirk and Miller \(1986\)](#) see this criterion as being as important as credibility if qualitative research findings are to be taken seriously. We assert that dependability includes the consistency with which the same constructs may be matched with the same phenomena over space and time (see [LeCompte and Goetz 1982](#)) but is largely concerned with documenting the research context. In this way, there are similarities with reliability, although the latter involves standards of stability, consistency and predictability, whereby multiple applications of the same research instrument are expected to yield similar findings ([Streiner and Norman 1989](#)). Few qualitative researchers are willing to concede the 'unreliability' of a study based on changes that have occurred 'naturally' within the study group(s) being

researched. Since qualitative researchers accept the inevitability of changes in 'reality', they tend to focus rather on design/researcher-induced changes. Yet, while quantitative researchers focus on the survey items themselves (e.g. wordings) and the manner in which they are administered, qualitative researchers focus on the interpretations and their consistency from one interview transcript to another. Thus dependability refers to the plausibility of accounts; reliability to the plausibility of design. So, although both types of research are concerned with plausibility, it is seen and treated differently.

There are at least two phenomena which may threaten the dependability of interpretations in qualitative research: poorly defined analytical constructs and premises ([LeCompte and Goetz 1982](#)), and premature closure ([Lincoln and Guba 1985](#)). When analytical constructs are poorly delineated, they may be subject to variable interpretation by both researchers and those being researched. For example, the concepts of 'class' or 'culture' are known to have numerous definitions, all with their own attendant assumptions and implications. Premature closure occurs when the researcher finalizes analytical constructs sooner than the available data warrant.

[LeCompte and Goetz \(1982\)](#) suggest five strategies for guarding against threats to dependability: low-inference descriptors, mechanically recorded data, multiple researchers, participant researchers and peer examination. Analogous to some of these are triangulation and the inquiry audit ([Lincoln and Guba 1985](#)). Low-inference descriptors and mechanically recorded data are two interrelated methods, whereby the degree to which the interpretations 'agree' with the data are authenticated by others ([Pelto and Pelto 1978](#); [Schatzman and Strauss 1973](#)). The most common types of low-inference descriptors are fieldnotes and audio recordings which include verbatim accounts and narratives of behaviours, activities and events. These may be used by other researchers to compare with the interpretations of the original researcher. Multiple researchers (investigator triangulation), participant researchers (a form of member checking) and peer examination (peer debriefing) are all methods, already discussed, of introducing alternative perspectives in data analysis prior to finalizing the set of theoretical constructs. None of the empirical papers gives any indication of the use of such strategies to check for the dependability of construct-to-data matching.

Some may argue that the similarity in strategies for improving dependability and credibility implies that they should be collapsed into one criterion. While the strategies to satisfy both are similar, the criteria themselves are quite different. Credibility refers to the accurate representation of experiences while dependability focuses attention on the researcher-as-instrument and the degree to which interpretation is made in a consistent manner. In support of this argument against merging dependability and credibility, Lincoln and Guba (1985, 317) claim that

Since there can be no validity without reliability (and thus no credibility without dependability), a demonstration of the former is sufficient to establish the latter. If it is possible using the techniques outlined in relation to credibility to show that a study has that quality, it ought not to be necessary to demonstrate dependability separately. But, while this argument has merit, it is also very weak. It may serve to establish dependability in practice, but does not deal with it in principle. A strong solution must deal with dependability directly.

The inquiry audit combines elements of thick description with those of peer examination and is analogous to a fiscal audit (Halpern 1983). The auditor maintains checks on the status of the research to ensure that appropriate decisions are made along the way. It is intended to produce a detailed account of how the research was done and, like the peer examiner, the auditor should be someone intimately familiar with qualitative research and/or the topic area. The inquiry audit is used to ensure that the appropriate decisions are made a priori. Unlike the peer examiner who enters near the end of the inquiry, the auditor is expected to be involved at the outset to look for and evaluate an 'audit trail' of how and why various decisions are made regarding such things as respondent selection, methods used and data interpretation techniques. Since the auditor is in a position to advise on these decisions, he can help to keep idiosyncrasy in the design and interpretation to a minimum. To a certain extent, the graduate student-professor supervisory relationship functions as a convenient, often implicit form of auditee-auditor research relationship, albeit much less formally than Halpern suggests.

Confirmability

Confirmability, similar to the conventional notion of objectivity, focuses attention on both the

investigator *and* the interpretations. Conventional objectivity is usually associated with the following assumptions: there is a single, largely unchanging reality; good data may reflect only that reality; and, when the researcher disturbs the data or the reality, objectivity is compromised. By contrast, confirmability may be defined as

the degree to which findings are determined by the respondents and conditions of the inquiry and not by the biases, motivations, interests or perspectives of the inquirer. (Lincoln and Guba 1985, 290)

Scriven (1971, 95–6) adds that, to be objective, data must be 'reliable, factual, confirmable or confirmed and so forth'. By incorporating concerns about the character of the data, confirmability is more broadly based than the principle(s) of objectivity which focus solely on the accountability of the inquirer. Thus qualitative researchers are expected to account for their interests and motivations by showing how they have affected interpretations.

There are similarities in the techniques used for enhancing rigour between confirmability and credibility, transferability and dependability. Confirmability highlights the research audit which includes audit trail products such as raw data, data reduction and analysis products, data reconstruction and synthesis products, process notes, materials relating to intentions and dispositions, and instrument development information. Confirmability is also an account of the audit process, including how decisions were made regarding the determination of credibility, transferability and dependability (see Halpern 1983). It is questionable how often this does or can occur in practice due to the rather elaborate and apparently strict audit criteria suggested by Halpern (*ibid.*). At the very least, qualitative researchers need to ask some basic questions of all their work so as to assist in evaluating design and findings. Rose (1982), for example, suggests eight (see Table III). A detailed journal with notes on findings, problems and interpretations (see Spradley 1979) is one way of providing material for an audit. The geography literature reviewed in this paper gives an indication of how infrequently audits, however scant in detail, are included in published reports. Rarely is there mention of the biases, motivations and interests of the researcher in relation to the questions asked and the decisions made throughout the research process. One exception is Cornwell (1988),

Table III Checklist for evaluating qualitative interview research

<i>Question</i>	<i>Elaboration/examples</i>	<i>Evaluation of Eyles et al. (1993)</i>
1 What was the natural history of the research?	Original purpose(s) of the research Rationale for methodology How research developed over time Fieldwork relations	Stated – research responses to a tyre fire Stated – to study resident responses in context Little discussion Little discussion
2 What data were collected and by what methods?	Method of note-keeping; Method of tape-recording	None apparent Five interviewers, over one month, three months after event (fire) Debriefing and training of interviewers Verbatim transcription
3 How was the sampling done?	Explicit delineation of sample frame (working universe) Random or purposeful? Rationale for type of sampling used	1300–1400 sample frame, 173 who talked with community outreach programme, 66 who agreed to follow-up=43 respondents Purposeful – opportunistic Stated – mutual trust
4 How was the data analysis done?	Procedures for summarizing and presenting data How data were selected for presentation	Not apparent (implicit – inter-interview themes) Not apparent
5 What results are presented?	Description of researcher's objective for results presentation (e.g. theory-building or description) Differentiation of data-derived as opposed to pre-existing constructs Differentiation of participant concepts as opposed to theoretical (researcher-derived) constructs	Descriptions of how the findings fit in with existing theory and outline policy implications Relate findings to literatures on environmental stress and risk perception Participant (quotations), data derived (researcher commentary), theoretical (Figure 1 in the paper)
6 How credible and dependable are the data–construct links?	Details of the relationship(s) between the data and constructs/concepts derived from data (e.g. member checking)	Analytical validation through investigator triangulation
7 How credible is the theory/hypothesis?	Specification of the relationship between constructs/concepts and theory/hypotheses	Yes, see note 5 Relate literature to this case study
8 How transferable are the findings?	Recognition of the limits imposed by the sampling strategy	Acknowledged case study Depends largely on credibility of research constructs

Source: [Rose \(1982\)](#); and [Lincoln and Guba \(1985\)](#)

who argues that there are many false trails in exploratory qualitative research, whilst the 'warts-and-all' of research are often told in autobiographies of the research process.¹⁴

A case study of rigour and trustworthiness

In our review of qualitative research in social geography, we indicate that there is inadequate mention of the practices and criteria for producing trustworthy results. But can the criteria in [Table III](#) be used reasonably and fairly to evaluate such work? We will address this issue by reviewing one

of our own papers ([Eyles et al. 1993](#)) concerning the social construction of risk in the community surrounding the Hagersville tyre fire in February 1990. We do not choose one of our own works in order to paint it in a favourable light (or, conversely, as an example of what not to do) but because we are in the unique position of being able to comment on the editorial exchanges between the journal editor and the authors which influenced what was included in the published paper.

While the purpose of the research is stated as 'identifying the effects of the fire, particularly through the ways in which people responded to and coped with the event' (*ibid.*, 283), and the

rationale for the methodology as 'a qualitative research design using depth interviews to allow us to examine the types and forms of residents' responses to the tire fire in their societal context' (*ibid.*, 282), there is little else mentioned of the natural history of the research (see Table III). There is a discussion of how the study fits in with the history of community remediation efforts in that the sample of 43 respondents was drawn from those who were visited by public health nurses from the Community Outreach Programme (COP). However, there is not much discussion of what happened after that in terms of the fieldwork relations between the researchers/interviewers and respondents. There is passing mention of interviewer debriefing sessions intended for 'identifying issues arising from the interviews which are not self evident from the transcripts' (*ibid.*, 283) but there are no examples of such issues, which may well have affected the manner in which the transcripts were constructed.

There is substantial information about the data and methods of collection. The interviews were conducted over a four-week period, three months after the fire, by five interviewers who taped each conversation. The interviews were transcribed for thematic analysis. However, the implications of using five interviewers (each with her/his own interviewing 'style' which influences what and how things get said) are not discussed in much detail other than the mention of the existence of interviewer training and debriefing sessions. These sessions seem to have been put in place to ensure consistency in the way residents were interviewed but it is not clear how this was (or was not) accomplished. Consistency is also addressed through the use of an interview checklist:

Depth interviews were guided by a checklist of topics to be covered with all respondents while allowing them considerable freedom to describe their experience and stories in their own terms. (*ibid.*)

Although it is not stated explicitly, covering the same (minimum) of topics facilitates analyses that may compare themes across interview texts. While the checklist itself is not provided – an editorial decision – it is apparently brief enough to allow respondents the power to direct the conversation.

The paper also includes considerable detail of sampling. One of the main concerns in conducting the study was the establishment of 'mutual trust' (*ibid.*) between the residents, who faced potentially

traumatizing circumstances, and the researchers. For this reason, a purposeful/opportunistic sampling strategy was used which took advantage of existing community remediation efforts, instead of a more 'rigorous' sampling strategy. The interviews were conducted as a follow-up to visits by public health nurses. The following excerpt details how the sample was achieved:

As a result, our sample was doubly selected, being drawn from those families who talked to the COP (N=173) and agreed to a follow-up (N=66). We tried to contact all 66. Some had moved and some were away. We talked to all those we could contact (N=43) and their stories form the basis of our account of the impacts of the fire.

There is, clearly, the potential to miss the accounts of certain groups. This is critical for a study claiming to provide stories which represent most groups within the community. Do those who did not have time for a follow-up comprise a group which has a different understanding of the event from those who did? The authors claim that the double sampling is not a threat to providing representative stories (*ibid.*):

These 43 provide a cross-section of people within the evacuation zone, some being upwind of the fire, some in the red zone (closest to the site), some on the neighbouring concession lines, and some downwind but further away.

Yet there is no evidence that the people who were not interviewed have stories of the fire which are similar to those of the respondents. Those who did not agree to a COP visit or were not visited at all by the COP (N=1200), did not agree to a follow-up (N=107) interview or could not be followed-up by one of the five research assistants (N=23) may speak differently of the fire. It is important to recognize that, while this study provides considerable detail about sampling when compared with many of the 30 other studies, more information about who was and who was not interviewed would be useful for revealing where gaps may exist. It is impractical to expect that all stories can be told in any given research situation but readers should know from whom they are and are not hearing.

As is the case with many qualitative papers in social geography, there is scant detail about how the data are transformed, in this case from 43 interviews to a framework for understanding

anxiety, uncertainty and risk in the Hagersville community (*ibid.*, Figure 1). The data are condensed to eleven quotations from only eight of the original respondents, with no indication of how this reduction occurred. What is offered, however, is the following statement regarding the validity of the constructs in the 'framework':

The analyses and interpretations of the stories have been undertaken and discussed by several members of the research team so that the end-product is a set of themes validated by multiple analysts. (*ibid.*)

Although such triangulation by multiple analysts may lead to credible interpretations, there is no mention of how these analysts managed the data. To return to an issue raised earlier, the editorial/peer review process resulted in the elimination of a section of the original paper which specified the strategies for analysing/condensing the transcript data. As is often the case, the authors compromised by directing the reader to a companion methodological piece in another journal which explains the data management and analysis in considerable detail (see also [Porteous 1988](#)).

The results appear in the forms of quotations, author commentary concerning the quotations and the 'framework' outlined in Figure 1 of the paper which shows how theoretical concepts are linked. The authors do differentiate lay concepts (the quotations), data-derived constructs (the commentary) and those constructs which already exist in the literature on environmental stress and risk perception (various parts of the framework). Further distinctions are made through a review of the literature at the beginning of the paper and a subsequent revisit to this literature at the end of the paper insofar as it relates to the constructs for the Hagersville case. These reviews reveal how the researchers relate constructs developed from the stories in the study to a broader literature and how this needed to be adapted in order to understand the Hagersville context. For example, the issues related to evacuation, latency, coping strategies and financial resources 'loomed large' in the aftermath of the fire but are not as prominent in the literature which were reviewed.

The credibility of both data-to-construct and construct-to-theory links is implicitly ensured by the investigator triangulation mentioned above and appeals to the literature of environmental risk and coping. The paper suffers from a lack of convincing elaboration/discussion, particularly of

the process of triangulation/interpretation. There is also an almost complete lack of reference to the relationships between the interviewers and interviewees which, as argued above, can have profound implications for what becomes the 'data' in the first place.

The issue of transferability in the paper is shaped by the fact that it was a case study of the risk, anxiety, uncertainty and coping experiences of the community surrounding the Hagersville tyre fire. The authors admit that the results are context-specific. In fact, that is the point of their endeavour. The limits to transferability are also apparent in the explicit statements regarding sampling. The discussion of the policy implications, however, suggests that the authors consider that some issues should be considered transferable to other events which require similar emergency responses:

The Hagersville case study points to a need for authorities, in difficult circumstances, to recognize and cater to the ways in which lay publics act, think, and talk. The use of local officials with such knowledge may be an answer. (*Eyles et al. 1993, 288-9*)

There is no audit trail of the research process as suggested by [Lincoln and Guba \(1981\)](#), or excerpts from field diaries, for ascertaining how research decisions were made. There is, however, considerable detail in the stories and descriptions themselves to determine the similarity between the Hagersville 'event' and other environmental events, so revealing how credibility and transferability may become intertwined. Yet all that is available to determine the transferability of the findings is the credibility of the constructs/theories presented.

This review of [Eyles et al. \(1993\)](#) indicates that the eight questions in [Table III](#) are general enough to provide a reasonable assessment of qualitative work which uses interviews as the main method of data collection. It serves not only as a guide for what to look for but reveals where gaps exist in reporting information necessary for ascertaining rigour. The table is not intended to provide specific standardized rules for what should be done to produce trustworthy findings. It indicates the basic information requirements for appraisal. The four criteria and the strategies outlined in [Table II](#) provide more detailed ways of assessing rigour and we suggest that they should be incorporated into the research process as a basis for answering the questions posed in [Table III](#).

Summary and conclusions

Until recently, qualitative researchers have tended to focus more on what criteria should *not* be used to evaluate their work – the standards used to judge positivistic-quantitative work – and less on what they should be looking for to determine the rigour of qualitative research. In social geography, this is reflected in both empirical and substantive work. The sample of qualitative geographic research reviewed here indicates scarce explicit mention of the principle(s) which have guided its concern for rigour and ensure meaningful inference. But its strengths should also be recognized in the use of multiple methods, numerous detailed quotations, discussions of validity and appeals to recognized bodies of literature. Nonetheless, researchers need to be more explicit about the research process including the rationale(s) for, among other things, respondent selection, key changes in research direction and analytical procedures. This may prove problematic if journal editors continue to emphasize results at the expense of equally important accounts of strategies for maintaining rigour.

While it may be expected that the criteria which these strategies are intended to address may be found in the substantive literature on qualitative methods, this tends to focus on distinguishing the epistemological and ontological foundations of qualitative from quantitative research as well as being reflexive about qualitative interview practices. There is also appeal, often uncritical, to interpretive communities. Less attention has been paid to why some theories seem to be more appealing than others. Why is their time right? For example, Lamont (1984) gives a fascinating assessment of the plausibility of Derrida and his interpretations and accounts while Fish (1979) writes of the authority of interpretive communities in texts.

The criteria of credibility, transferability, dependability and confirmability for establishing rigour are useful general principles for guiding qualitative evaluation. These criteria are analogous to the traditional quantitative standards of validity, generalizability, reliability and objectivity, and much of the debate is framed in this way (e.g. Bryman 1988; Silverman 1993). Yet similarities regarding the principles to which both sets of evaluative criteria appeal should not be interpreted as a licence to use quantitative criteria to evaluate qualitative work (or vice versa). Beyond the

considerable differences between the criteria themselves, the strategies used to strengthen qualitative rigour are quite different from their quantitative counterparts. It is important to distinguish between principles which may be similar and practices (see Table II) which are quite different when evaluating rigour for qualitative as opposed to quantitative work.

There remain, however, objections to establishing qualitative principles for evaluating rigour. Thus Risteen Hasselkus' (1991, 3–4) claims that

We seem to be circumscribing qualitative research with an orthodoxy of rules to which it must conform. In my view this bend toward a dogma of qualitative inquiry can potentially smother the creative elegance of such research. Further, I do not believe that the writers who have brought forward these guidelines and criteria for judging qualitative research ever intended for those guidelines to be incorporated as mandatory rules and regulations in the qualitative research process.

We agree that there should be no mandatory rules but there should be criteria that enable a judgement to be made concerning honesty, integrity and plausibility of design and accounts. These criteria/principles are intentionally general and, consequently, may be satisfied in numerous ways. Indeed, the list of strategies for strengthening them is quite extensive but should by no means be considered all-encompassing. It is when limits are placed on the types of strategies that may be used to achieve rigour that there is a danger of a 'dogma of qualitative inquiry'. But questioning how things are done – an essential component of self-reflection – allows qualitative research to demonstrate the relevance of the single case (credibility) and to move beyond it (transferability) with a degree of certainty (dependability and confirmability). Context, contingency and the specific positioning of subjects (including researcher-as-instrument) are central to qualitative inquiry and are not threatened by the application of a general set of criteria for evaluating rigour. These criteria provide reasonable anchor points for a paradigm which is often inappropriately accused of engaging in 'anything goes' science:

As long as we strive to base our claims and interpretations of social life on data of any kind, we must have a logic for assessing and communicating the interactive process through which the investigator acquired the research experience and information. If we are to understand the detailed means through which human beings engage in meaningful action and create a world

of their own or one that is shared with others. (Morgan 1983, 397)

We must acknowledge that

insufficient attention has as yet been devoted to evolving criteria for assessing the general quality and rigour of interpretive research. (*ibid.*, 399)

This task still largely awaits qualitative research in social geography. Criteria (establishing ways of thinking) and detailed questioning will help us accomplish this task.

Acknowledgements

Both authors acknowledge the advice given at presentations of this manuscript in draft at meetings of the Medical Geography Specialty Group of the Association of American Geographers, the Sixth International Symposium in Medical Geography, and by the three anonymous referees. Jamie Baxter also wishes to acknowledge the support of an Ontario Graduate Scholarship and a Social Science and Humanities Research Council (Canada) Research Fellowship. John Eyles is supported by the Ecoresearch Program in Environmental Health with contributions from the Social Science, Engineering and Medical Research Councils, Environment Canada and the Regional Municipality of Hamilton-Wentworth.

Notes

- 1 Searching the *Wilson social sciences*, *Wilson humanities* and *Sociofile* indexes as well as twelve geography journals manually, we found 49 papers, 31 empirical and eighteen commentary. While there are undoubtedly books which include studies involving qualitative interview methods, these could not be located in the computer and manual searches of the McMaster libraries. We acknowledge that the longer formats of books/monographs/chapters may allow writers more opportunity to elaborate methodological and analytical procedures and that they may be less subject to the criticisms in this paper but they are seldom subject to as critical peer review as journal articles. Peer-reviewed journal articles thus represent the pinnacle of successful incorporation of qualitative approaches in social geography.
- 2 The following articles detail how people were selected: Bridge (1994); Cooper (1994); Droogleever Fortuijn and Karsten (1989); Dyck (1989); Eyles *et al.* (1993); Leckie (1993); McDowell (1994); Rollinson (1990); Rutherford (1995); and Valentine (1995).

- 3 For example, Cooper (1994); Dyck (1989); Leckie (1993); Rowe and Wolch (1990); Valentine (1993); and Wilson (1993).
- 4 For example, Bonnett (1992); Bridge (1994); Brown (1995); McDowell and Court (1994); and Rollinson (1990).
- 5 For example, Dyck (1989); Eyles *et al.* (1993); Fernandez Kelly (1994); Hewitt (1994); and McDowell (1994).
- 6 Cooper (1995) and McDowell (1994) provide some detail on how their role in the research interviews may have affected responses.
- 7 While there is lively methodological debate among quantitative social scientists, there is consensus about many basic strategies for establishing rigour, particularly those involving basic statistical procedures.
- 8 Eyles (1988b); Eyles and Smith (1988); Jackson (1985); Keith (1992); Ley and Samuels (1978); Livingstone (1992); Lowe and Short (1990); McDowell (1992a, 1992b); Miles and Crush (1993); Mitchell and Draper (1981); Moss (1993); Nast (1994); Pickles (1988); Pile (1991); Rose (1993); Schoenberger (1991); and Smith (1984).
- 9 For example, Keith (1992); Miles and Crush (1993); Moss (1993); Nast (1994); Rose (1993); and Smith (1988).
- 10 We concur with Lincoln and Guba that validity (credibility) is quite different from representativeness (transferability) in the sense that the former is about representations of experiences while the latter concerns the applicability of the findings to other contexts.
- 11 For example, Buttimer (1974); Donovan (1986); Eyles (1985); and McDowell (1994).
- 12 The 'sending context' is one in which an original study has been undertaken while the 'receiving context' is one to which the sending context findings may apply. Such an application must be determined empirically.
- 13 Droogleever Fortuijn and Karsten do quite the opposite. See their quotation earlier in this paper.
- 14 See Bell and Newby (1977); Bell and Roberts (1984); Eyles (1988b); and Roberts (1984).

References

- Athens L 1984 Scientific criteria for evaluating qualitative studies in Denzin N ed. *Studies in symbolic interaction volume 5* Jai Press, London
- Barnes T and Duncan J 1992 *Writing worlds* Routledge, London
- Baxter J Eyles J and Willms D 1992 The Hagersville tire fire: interpreting risk through a qualitative research design *Qualitative Health Research 2* 2 208-37
- Bell C and Newby H 1977 *Doing sociological research* Allen Unwin, London

- Bell C and Roberts N** 1984 *Social researching* RKP, London
- Bonnett A** 1992 Anti-racism in 'white' areas: the example of Tyneside *Antipode* 24 1 1-15
- Borland K** 1991 That's not what I said: interpretive conflict in oral narrative research in **Gluck S B and Patai D** eds *Women's words: the feminist practice of oral history* Routledge, New York
- Bridge G** 1994 Gentrification, class and residence: a re-appraisal *Environment and Planning D: Society and Space* 12 31-51
- Brown M** 1995 Ironies of distance: an ongoing critique of the geographies of AIDS *Environment and Planning D: Society and Space* 13 159-83
- Bryman A** 1988 *Quantity and quality in social research* Unwin-Hyman, London
- Burgess R** 1984 *Field research* Allen Unwin, London
- Burns N** 1988 Standards for qualitative research *Nursing Science Quarterly* 2 1 44-52
- Buttimer A** 1974 *Values in geography* Association of American Geographers, Washington
- Clark M Gregory K J and Gurnell A M** 1987 *Horizons in physical geography* Barnes and Noble, Totowa, NJ
- Cooper A** 1994 Negotiating dilemmas of landscape, place and Christian commitment in a Suffolk parish *Transactions of the Institute of British Geographers* 19 202-12
- Cooper A** 1995 Adolescent dilemmas of landscape, place and religious experience in a Suffolk parish *Environment and Planning D: Society and Space* 13 349-63
- Cornwell J** 1988 A case-study approach to lay health beliefs: reconsidering the research process in **Eyles J and Smith D** eds *Qualitative methods in human geography* Basil Blackwell, Oxford 219-32
- Cosgrove D and Daniels S** 1988 *The iconography of landscape* Cambridge University Press, Cambridge
- Denzin N** 1978 *The research act* McGraw-Hill, New York
- Donovan J** 1986 *You don't buy sickness, it just comes* Gower, Aldershot
- Donovan J** 1988 'When you're ill you've gotta carry it': health and illness in the lives of black people in **Eyles J and Smith D** eds *Qualitative methods in human geography* Polity Press, Cambridge
- Droogleever Fortuijn J and Karsten L** 1989 Daily activity patterns of working parents in the Netherlands *Area* 21 4 265-76
- Dukes W** 1965 N=1 *Psychological Bulletin* 64 74-9
- Dyck I** 1989 Integrating home and wage workplace: women's daily lives in a Canadian suburb *The Canadian Geographer* 33 4 329-41
- Erickson F** 1973 What makes school ethnography 'ethnographic'? *Anthropology and Education Quarterly* 4 2 10-99
- Eyles J** 1985 *Senses of place* Silverbrook Press, Warrington
- Eyles J** 1988a Interpreting the geographical world in **Eyles J and Smith D** eds *Qualitative methods in human geography* Polity Press, Cambridge
- Eyles J** 1988b *Research in human geography* Blackwell, Oxford
- Eyles J and Donovan J** 1986 Making sense of sickness and care: an ethnography of health in a West Midlands town *Transactions of the Institute of British Geographers* 11 4 15-27
- Eyles J and Donovan J** 1990 *The social effects of health policy* Avebury Press, Aldershot
- Eyles J and Perri E** 1993 Life history as method: an Italian-Canadian family in an industrial city *The Canadian Geographer* 37 2 104-19
- Eyles J and Smith D** 1988 *Qualitative methods in human geography* Polity Press, Cambridge
- Eyles J Taylor S M Baxter J Sider D Willms D** 1993 The social construction of risk in a rural community: responses of local residents to the 1990 Hagersville (Ontario) tire fire *Risk Analysis* 13 3 281-90
- Fernandez Kelly M P** 1994 Towanda's triumph: social and cultural capital in the transition to adulthood in the urban ghetto *International Journal of Urban and Regional Research* 18 1 88-111
- Fielding N and Fielding L** 1986 *Linking data* Sage, Newbury Park
- Fish S** 1979 *Is there a text in this class?* Harvard University Press, Cambridge
- Frost P and Stablein R** 1992 *Doing exemplary research* Sage, Newbury Park
- Geertz C** 1973 *The interpretation of culture: selected essays* Basic Books, New York
- Gregson N and Lowe M** 1995 'Home'-making: on the spatiality of daily social reproduction in contemporary middle-class Britain *Transactions of the Institute of British Geographers* 20 224-35
- Halpern E** 1983 Auditing naturalistic inquiries: the development and application of a model Unpubl. PhD dissertation, Indiana University, Bloomington, IN
- Hammersley M** 1992 *What's wrong with ethnography* Routledge, London
- Harrison C and Burgess J** 1994 Social constructions of nature: a case study of conflicts over the development of Rainham Marshes *Transactions of the Institute of British Geographers* 19 291-310
- Herod A** 1991 Local political practice in response to a manufacturing plant closure: how geography complicates class analysis *Antipode* 23 4 385-402
- Hewitt K** 1994 'When the great planes came and made ashes of our city': towards an oral geography of the disasters of war *Antipode* 26 1 1-34
- Jackson P** 1985 Urban ethnography *Progress in Human Geography* 9 2 159-76
- Katz C** 1991 Sow what you know: the struggle for social reproduction in rural Sudan *Annals of the Association of American Geographers* 81 3 488-514
- Keith M** 1992 Angry writing: (re)presenting the unethical world of the ethnographer *Environment and Planning D: Society and Space* 10 551-68
- Kidder L** 1981 Qualitative research and quasi-experimental frameworks in **Brewer B and Collins B**

- eds *Scientific inquiry and the social sciences* Jossey-Bass, San Francisco
- King G Keohane R and Verba S** 1994 *Designing social inquiry: scientific inference in qualitative research* Princeton University Press, Princeton
- Kirk J and Miller M** 1986 *Reliability and validity in qualitative research* Sage, Newbury Park, CA
- Knaf K** 1994 Promoting academic integrity in qualitative research in **Morse J** ed. *Critical issues in qualitative research methods* Sage, Thousand Oaks, CA
- Knaf K and Breitmayer B** 1989 Triangulation in qualitative research: issues of conceptual clarity and purpose in **Morse J** ed. *Qualitative nursing research: a contemporary dialogue* Aspen, Rockville, MD 193–203
- Krefting L** 1990 Rigor in qualitative research: the assessment of trustworthiness *The American Journal of Occupational Therapy* 45 3 214–22
- Kuzel A** 1992 Sampling in qualitative inquiry in **Crabtree B and Miller W** eds *Doing qualitative research* Sage, Newbury Park
- Lamont M** 1984 How to become a dominant French philosopher *American Journal of Sociology* 93 584–622
- Leckie G** 1993 Female farmers in Canada and gender relations of a restructuring agricultural system *The Canadian Geographer* 37 3 212–30
- LeCompte M and Goetz J** 1982 Problems of reliability and validity in ethnographic research *Review of Educational Research* 52 1 31–60
- Ley D** 1974 *The black inner city as frontier outpost* Association of American Geographers, Washington
- Ley D and Samuels M** eds 1978 *Humanistic geography* Maaroufa Press, Chicago
- Lieberson S** 1992 Small N's and big conclusions in **Ragin C and Becker H** eds *What is a case?* Cambridge University Press, New York
- Lincoln Y and Guba E** 1981 Strategies for insuring the dependability (reliability) of naturalistic studies Paper presented at the joint annual meeting of the Evaluation Network and the Evaluation Research Society, Austin, TX
- Lincoln Y and Guba E** 1985 *Naturalistic inquiry* Sage, Beverly Hills, CA
- Livingstone D** 1992 *The geographical tradition* Oxford University Press, Cambridge
- Lowe M and Short J** 1990 Progressive human geography *Progress in Human Geography* 14 1 1–11
- Mackenzie F** 1992 'The worse it got, the more we laughed': a discourse of resistance among farmers of eastern Ontario *Environment and Planning D: Society and Space* 10 691–713
- McDowell L** 1992a Valid games? A response to Erica Schoenberger *Professional Geographer* 44 2 212–15
- McDowell L** 1992b Doing gender: feminism, feminists and research methods in human geography *Transactions of the Institute of British Geographers* 17 399–416
- McDowell L** 1994 Social justice, organizational culture and workplace democracy: cultural imperialism in the city of London *Urban Geography* 15 7 661–80
- McDowell L and Court G** 1994 Performing work: bodily representations in merchant banks *Environment and Planning D: Society and Space* 12 727–50
- Merton R** 1968 *Social theory and social structure* Free Press, New York
- Miles M and Crush J** 1993 Personal narratives as interactive texts: collecting and interpreting life histories *Professional Geographer* 45 1 84–94
- Mitchell B and Draper D** 1982 *Relevance of ethics in geography* Longman, London
- Morgan G** 1983 *Beyond method* Sage, Newbury Park, CA
- Moser C and Kalton G** 1981 *Survey methods in social investigation* Heineman, London
- Moss P** 1993 Focus: feminism as method *The Canadian Geographer* 37 1 48–61
- Nast H** 1994 Women in the field: critical feminist methodologies and theoretical perspectives *Professional Geographer* 46 1 54–66
- Newton J** 1995 An assessment of coping with environmental hazards in northern aboriginal communities *The Canadian Geographer* 39 2 112–20
- Patton M** 1990 *Qualitative evaluation and research methods* Sage, Newbury Park, CA
- Pelto P and Pelto G** 1978 *Anthropological research: the structure of inquiry* Cambridge University Press, Cambridge
- Pickles J** 1988 From fact-world to life-world in **Eyles J and Smith D** eds *Qualitative methods in human geography* Polity Press, Cambridge
- Pile S** 1991 Practising interpretative geography *Transactions of the Institute of British Geographers* 16 458–69
- Plummer K** 1983 *Documents of life* Allen Unwin, London
- Porteous J** 1988 Topocide: the annihilation of place in **Eyles J and Smith D** eds *Qualitative methods in human geography* Basil Blackwell, Oxford 75–93
- Ristein Hasselkus B** 1991 Qualitative research: not another orthodoxy *The Occupational Therapy Journal of Research* 11 1 3–7
- Roberts H** 1984 *Doing feminist research* RKP, London
- Rollinson P** 1990 The everyday geography of poor elderly hotel tenants in Chicago *Geografiska Annaler* 72B 2–3 47–57
- Rosaldo R** 1989 *Culture and truth* Beacon Press, Boston
- Rose D** 1993 Feminism, method and methods in human geography: an idiosyncratic overview *Canadian Geographer* 37 57–61
- Rose G** 1982 *Deciphering sociological research* MacMillan, London
- Rowe S and Wolch J** 1990 Social networks in time and space: homeless women in skid row Los Angeles *Annals of the Association of American Geographers* 80 2 184–204
- Rutherford T** 1995 'Control the ones you can': production restructuring selection and training in Kitchener Region Manufacturing 1987–1992 *The Canadian Geographer* 39 1 30–45

- Sandelowski M** 1986 The problem of rigor in qualitative research *Advances in Nursing Science* 8 3 27–37
- Schatzman L and Strauss A** 1973 *Problems in participant observation* Prentice-Hall, Englewood Cliffs, NJ
- Schoenberger E** 1991 The corporate interview as a research method in economic geography *Professional Geographer* 43 2 180–9
- Schutz A** 1962 Commonsense and scientific interpretation of human action in **Natson M** ed. *Collected papers volume 1* Nijhoff, The Hague
- Schwartz H and Jacobs J** 1979 *Qualitative methods* Free Press, New York
- Scriven M** 1971 Objectivity and subjectivity in educational research in **Thomas L** ed. *Philosophical redirection of educational research. The seventy-first yearbook of the National Society for the Study of Education* University of Chicago Press, Chicago
- Shute J and Knight D** 1995 Obtaining an understanding of environmental knowledge: Wendaban Stewardship Authority *The Canadian Geographer* 39 2 101–11
- Silverman D** 1993 *Interpreting qualitative data* Sage, London
- Smith S** 1984 Practising humanistic geography *Annals of the Association of American Geographers* 74 3 353–74
- Smith S** 1988 Constructing local knowledge: the analysis of self in everyday life in **Eyles J and Smith D** eds *Qualitative methods in human geography* Basil Blackwell, Oxford 17–38
- Smith S** 1994 Qualitative methods in **Johnston R** ed. *The dictionary of human geography* Blackwell, Oxford
- Spradley J** 1979 *The ethnographic interview* Holt Reinhart and Winston, New York
- Streiner D and Norman G** 1989 *Health measurement scales: a practical guide to their development and use* Oxford University Press, New York
- Teich A and Frankel M** 1992 *Good science and responsible scientists* Association of American Geographers, Washington
- Valentine G** 1989 The geography of women's fear *Area* 21 4 385–90
- Valentine G** 1993 (Hetero)sexing space: lesbian perceptions and experiences of everyday spaces *Environment and Planning D: Society and Space* 11 395–413
- Valentine G** 1995 Out and about: geographies of lesbian landscapes *International Journal of Urban and Regional Research* 19 1 96–111
- Vidich P Bensenman J and Stein M** 1964 *Reflections on community studies* Free Press, New York
- Wax R** 1971 *Doing fieldwork: warnings and advice* University of Chicago Press, Chicago
- Willms D Best J Taylor D Gilbert J Wilson D Lindsay E and Singer J** 1990 A systematic approach for using qualitative methods in primary prevention research *Medical Anthropology Quarterly* 4391–409
- Wilson D** 1993 Everyday life, spatiality and inner city disinvestment in a US city *International Journal of Urban and Regional Research* 17 4 576–94
- Winchester H and Costello L** 1995 Living on the street: social organization and gender relations of Australian street kids *Environment and Planning D: Society and Space* 13 329–48