

LSE Research Online

[Ernestina Coast](#), Natalie Mondain and Clementine Rossier Qualitative research in demography: quality, presentation and assessment

Conference Item [eg. keynote lecture, etc.]

Original citation:

Coast, Ernestina and Mondain, Natalie and Rossier, Clementine (2009) *Qualitative research in demography: quality, presentation and assessment*. In: XXVI IUSSP International Population Conference, 27th September - 2nd October, 2009, Marrakech, Morocco. (Unpublished)

This version available at: <http://eprints.lse.ac.uk/36788/>

Available in LSE Research Online: July 2011

© 2009 The Authors

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

Qualitative research in demography: quality, presentation and assessment¹

E. Coast²

N. Mondain³

C. Rossier⁴

L. Bernardi⁵

S. Randall⁶

1. Context: qualitative approaches in demography

There has been an increase in the number published studies, and ongoing research programmes, that use qualitative approaches (in full or in part) in recent years in the broad field of demography (Coast, 2003; Randall & Koppenhaver, 2004). Randall & Koppenhaver (2004) found 24 articles for 1991-1995 and 43 articles for 1996-2000 meeting criteria of qualitative research in demographic literature⁷. In an update, we found 54 articles meeting these criteria for 2001-2005. Changes have occurred in recent years in demography, mainly related to the need for more explanatory (rather than descriptive and predictive) frameworks, leading both to the improvement of quantitative methods as well as the emergence of qualitative approaches (Riley and McCarthy, 2003).

This rise in qualitative contributions to research endeavours has been mirrored in other, traditionally quantitative, disciplines (Daly, 2007). This paper acknowledges that the simple division of research into qualitative and/or quantitative is not particularly helpful, not least because they have associations with research paradigms with which they are not necessarily linked, but that discussion is beyond the scope of this paper. Since the 1980s, population studies scientists have increasingly used qualitative methods. The methods used have themselves shifted over time, from ethnographic approaches (Caldwell micro-approaches book ref.), towards increasing reliance on focus groups and in-depth interviews (Randall and Koppenhaver 2004). Qualitative approaches are used in demography at multiple points in the

¹ Paper for presentation at IUSSP 2009, Marrakech, 27th September - 2nd October

² Senior Lecturer in Population Studies, London School of Economics e.coast@lse.ac.uk

³ Assistant Professor, University of Ottawa Nathalie.Mondain@uottawa.ca

⁴ Institut National d'Études Démographiques (INED) clementine.rossier@ined.fr

⁵ Associate Professor, University of Lausanne laura.bernardi@unil.ch

⁶ Reader, University College London s.randall@ucl.ac.uk

⁷ Demography, Perspectives in Sexual and Reproductive Health, International Family Planning Perspectives, International Migration Review, Population -English edition, Population Studies, Population and Development Review, Studies in Family Planning. All these journals are on JSTOR, except for International Migration Review. Key words used to search abstracts: anthropology / gical / gist, ethnography / phic / pher, qualitative, focus groups, in-depth / semi-structured interviews, participant observation

production of knowledge, including: design and testing of quantitative questionnaires; to understand unexpected survey results; and, to grasp sensitive issues, perceptions, "cultural contexts", and other elements of the social world which are difficult to measure quantitatively (Bozon 2006).

This use of qualitative methods raised a number of criticisms on the part of anthropologists (Greengage 1997, Kreutzer & Fricke 1997, Base & Abby 1998, Coast 2003). While demographers borrowed the tools of sociology and anthropology, they often left out these disciplines' theories (i.e. they did not necessarily theorize demographic behaviours as socially embedded individual actions), nor did they necessarily embrace a non-positivist epistemology. This disconnection is problematic since methods, theories and epistemologies are linked (Riley & McCarthy 2003). Anthropological demography and historical demography tried in the last decades to reconcile these diverging perspectives by developing unified interpretative frameworks and methodology for population studies (Bernardi and Hutter 2007)

Within demography, qualitative approaches have tended to focus on certain topics, particularly those that are considered "sensitive". Of course, the choice of method is dependent on whether it is the best approach to answering a question, and within demography qualitative approaches have tended to be used in the study of family and partnership dynamic, fertility and family planning, and lifestage transitions e.g.: the transition to adulthood. Furthermore there is a plethora of qualitative research around proscribed behaviour such as unsanctioned sexual activity, illegal, illicit and 'undesirable' behaviour (e.g. adolescent and extra marital sexual activity, abortions etc). These issues tend to be poorly captured by surveys and where data on attitudes and actual behaviour is needed especially for guiding subsequent interventions. The problems posed by HIV/AIDS research and interventions have driven much of the movement towards qualitative research methods.

Within demography, qualitative approaches incorporate a richness of information that cannot be captured using only quantitative data (Randall & Koppenhaver, 2004; De Loenzien & Yana, 2005). As such, qualitative methods have increasingly drawn the attention of demographers and led to the emergence (or recognition) of a new field: anthropological demography (Kreutzer & Fricke, 1997; Bledsoe, 1998; Greengage, 1997) even if it must be made clear that the qualitative methods used by demographers are *not* anthropological (Randall & Koppenhaver, 2004). In demography, the most frequently used qualitative methods are focus group discussions (Barker & Rich. 1992; Calves, 2000; Gueye, M. et al, 2001; Castle et al, 1999), in-depth individual semi-structured interviews (LeGrand et al., 2003; Randall & LeGrand, 2003; Mondain & Delaunay, 2006; Mondain et al., 2007); and, life histories (Randall and Mondain, 2009). Combinations of methods are also used, attempting data triangulation (LeGrand et al. 2003; Randall & LeGrand, 2003), together with other "systematic" methods such as pile sorts. The choice of

such methods is likely to be driven by the fact that demographers are highly concerned by statistical representativeness and as such have been reluctant to use qualitative methods involving non-random selection of small(er) numbers of participants. Therefore, when acknowledging the advantages of using qualitative methods, demographers have 'naturally' turned towards those methods allowing them to gather a greater number of participants. Methods privileged by 'pure' anthropologists, such as participant observation, are less (well) used.

There is a wide body of work on how to collect, analyse and present qualitative data in the social sciences in general (Marshall & Rossman, 2006; Huberman & Miles, 2002; Patton, 2002; Denzin & Lincoln, 2000; Strauss & Corbin, 1990) and demography in particular [Obermeyer, 1997; Axinn & Pearce, 2006]. The body of work that deals with assessing the quality of qualitative evidence is much less well developed, and, to date, not found within the broad discipline of demography (for an exception see Matthews).

Mixed methods, the combined use of quantitative and qualitative methods, are increasingly widely used in social science in general (Tashakkori & Teddlie, 1998; Onwuegbuzie & Teddlie, 2002; Onwuegbuzie & Leech, 2005). In demography, broadly, examples remain scarce and in most cases, when both methods are used, authors have difficulties to articulate the statistical and qualitative results with each other and thus describe them separately (Mondain et al., 2004; Mondain et al, 2007)⁸. Approaches for assessing the quality of mixed-methods research in general are very poorly developed (Bryman, Sale & Brazil, 2004; Tashakkori & Teddlie, 2003), and are not explored explicitly in this paper.

Researchers need to produce, identify, assess, aggregate, interpret and disseminate the highest quality evidence (Curran et al, 2007), whilst acknowledging that quality is, in and of itself, an abstract concept (Boaz & Ashby, 2003). Let us consider the routine tasks of a demographer that require assessing both these. In their own research, an evaluation of data quality is an integral part of the whole process and will influence what analyses are subsequently undertaken. However the demographer must also assess both the data and analyses of others: reading published research; and, reviewing journal articles, funding proposals, and student theses. Some of this evidence may draw on qualitative data. Demographers have an arsenal of techniques for assessing the quality of quantitative data, but they are unlikely to have been trained in methods of qualitative data collection or analysis, and even less likely to have been trained to assess their quality. At the heart of this issue is that evidence produced by a qualitative study is likely to look quite different

⁸ For example, workshop on the articulation of quantitative and qualitative research in population studies was convened at the Institut Supérieur des Sciences de la Population, University of Ouagadougou 18-20th October, Ouagadougou, Burkina Faso

from that produced by a quantitative study. For a novice qualitative researcher (who might be a highly experienced quantitative researcher), the methodological minutiae of qualitative research might be overwhelming (Daly et al, 2007). Experienced qualitative researchers know, or think they know, good qualitative research when they see it although they approach qualitative evidence with variable levels of experience, skills and self-confidence. So how can the demographer without such qualitative experience know "high quality" when they see it, for the equating of peer review publication with high quality does not always necessarily hold (Grayson, 2002)?

For any study, regardless of ontology or epistemology, there are standard processes for doing research, such as matching method to research question(s) and ensuring ethical conduct. Our focus in this paper is on the production and reporting of qualitative research in demography, drawing on two perspectives: views from primary data collection; and, experiences of reading and reviewing published studies. We begin with a brief review of how quality is assessed, drawing on frameworks from other disciplines.

2. What is quality? Should it be assessed? How might it be assessed?

Qualitative research encompasses many different traditions and methods, and the entry points to appraise quality are many: research design; methodological approach; data collection; analysis; and, insight and interpretation. There are extensive concerns about how quality or rigour or thoroughness might be made manifest. These concerns are being driven in part by an agenda for inclusion of qualitative research in systematic reviews alongside those traditionally established for quantitative research.

Criteria for assessing the quality of quantitative research are well-developed, and there is much debate about whether criteria developed for quantitative evidence (for example, reliability, validity (both internal and external) and replicability) might also be used for qualitative evidence (Hope & Waterman, 2003; Bryman, 2001; Popay et al, 1998; Rolfe, 2006). Other authors argue that the very language of "scientific" endeavour is inappropriate for qualitative research. Sandelowski, for example, argues that the criterion of "trustworthiness" should replace notions of "validity". Criteria for judging the quality of qualitative evidence are highly contested. Indeed, for some, the very notion of developing such criteria undermine the endeavour of qualitative research; the very reflexivity and flexibility that characterizes qualitative research might be seen as at odds with the development of criteria for the evaluation of its quality. Or, should concerns about the quality of qualitative research be made on its own terms, with the case being made for a reformulation of notions of quality being quite different to those used for quantitative research? Dixon-Woods et al (2004) argue that the diversity or "near anarchy" in the range of qualitative approaches makes it difficult to achieve consensus as to what makes for a flawed piece of research.

In the same way that qualitative research is underpinned by a wide variety of epistemological and methodological approaches, so too is the agenda related to its quality appraisal. Many disciplines have developed criteria to assess the quality of qualitative studies, notably in the area of health-related research. Debates about what criteria are appropriate for assessing the quality of qualitative research are longstanding (Holloway & Wheeler, 1996; Perakyla, 1997; Seale, 1999). Some learned societies and journals now produce guidelines for the review of explicitly qualitative research⁹, and the Evidence-based Medicine Working Group has gone so far as to identify an "ideal qualitative study" (Guyatt & Rennie, 2001).

Much of the research about research and evidence quality has focused on methodological quality, not least because this is relatively straightforward to assess (Boaz & Ashby, 2003). Spencer et al (2003) reviewed 29 frameworks for assessing the quality of qualitative research, the majority of which were developed in the fields of medical or health-related research. They identified heterogeneity in terms of framework purpose (e.g.: assessment of articles, proposal review), applicability to different methods, and their length, format and coverage. The purpose of their review was to develop a framework for assessing quality of qualitative research, which resulted in the identification of 4 central principles¹⁰, articulated by 18 "Appraisal questions"¹¹ dealing with research design, findings, sample, data collection, analysis, reporting, reflexivity and neutrality. Each of these 18 questions has a series of linked quality indicators.

Dixon-Woods et al (2004) identified in excess of 100 sets of proposals for quality in qualitative research. They suggest that there is little emerging

⁹ Medical Sociology Group of the British Sociological Association; Canadian Journal of Public Health

¹⁰ Contributory in advancing wider knowledge or understanding about policy, practice, theory or a particular substantive field; Defensible in design by providing a research strategy that can address the evaluative questions posed; Rigorous in conduct through the systematic and transparent collection, analysis and interpretation of qualitative data; credible in claim through offering well-founded and plausible arguments about the significance of the evidence generated.

¹¹ How credible are the findings? How has knowledge/understanding been extended by the research? How well does the evaluation address its original claims and purpose? Scope for drawing wider inference - how well is this explained? How clear is the basis for evaluative appraisal? How defensible is the research design? How well defended is the sample design/target selection of cases / documents? Sample composition/ case inclusion - how well is the eventual coverage described? How well was the data collection carried out? How well has the approach to, and formulation of, the analyses been conveyed? Contexts of data sources - how well are they retained and portrayed? How well has the diversity of perspective and content been explored? How well has detail, depth and complexity (ie "richness") of the data been conveyed? How clear are the links between data, interpretation and conclusions i.e.: how well can the route to any conclusions be seen? How clear and coherent is the reporting? How clear are the assumptions/ theoretical perspectives/ values that have shaped the form and output of the evaluation? What evidence is there of attention to ethical issues? How adequately has the research process been documented?

consensus either as to whether quality assessment of qualitative research is achievable or even desirable. They suggest that what is critical in the development of any quality criteria is the need to distinguish “fatal flaws” from “minor errors”. More recently, Walsh & Downe (2006) reviewed 8 frameworks (including those that themselves incorporated reviews of other frameworks e.g.: Spencer et al, 2003), synthesised and removed ‘redundant’ criteria, and produced a list of 12 essential criteria¹², identified by 53 specific prompts¹³. In a recent review of frameworks to assess the quality of qualitative

¹² Clear statement of, and rationale for, research question/aims/purposes; Study thoroughly contextualised by existing literature; Method/design apparent, and consistent with research intent; Data collection strategy apparent and appropriate; Sample and sampling method appropriate; Analytic approach appropriate; Context described and taken account of in interpretation; Clear audit trail given; Data used to support interpretation; Researcher reflexivity demonstrated; Demonstration of sensitivity to ethical concerns; Relevance and transferability evident

¹³ Clarity of focus demonstrated; Explicit purpose given, such as descriptive/explanatory intent, theory building, hypothesis testing; Link between research and existing knowledge demonstrated; Evidence of systematic approach to literature review, location of literature to contextualise the findings, or both; Rationale given for use of qualitative design; Discussion of epistemological/ontological grounding; Rationale explored for specific qualitative method (e.g. ethnography, grounded theory, phenomenology); Discussion of why particular method chosen is most appropriate/sensitive/relevant for research question/aims; Setting appropriate; Were data collection methods appropriate for type of data required and for specific qualitative method? Were they likely to capture the complexity/diversity of experience and illuminate context in sufficient detail? Was triangulation of data sources used if appropriate? Selection criteria detailed, and description of how sampling was undertaken; Justification for sampling strategy given; Thickness of description likely to be achieved from sampling; Any disparity between planned and actual sample explained; Approach made explicit (e.g. Thematic distillation, constant comparative method, grounded theory); Was it appropriate for the qualitative method chosen? Was data managed by software package or by hand and why? Discussion of how coding systems/conceptual frameworks evolved; How was context of data retained during analysis; Evidence that the subjective meanings of participants were Portrayed; Evidence of more than one researcher involved in stages if appropriate to epistemological/theoretical stance; Did research participants have any involvement in analysis (e.g. member checking); Evidence provided that data reached saturation or discussion/rationale if it did not; Evidence that deviant data was sought, or discussion/rationale if it was not; Description of social/physical and interpersonal contexts of data collection; Evidence that researcher spent time ‘dwelling with the data’, interrogating it for competing/alternative explanations of phenomena; Sufficient discussion of research processes such that others can follow ‘decision trail’; Extensive use of field notes entries/verbatim interview quotes in discussion of findings; Clear exposition of how interpretation led to conclusions; Discussion of relationship between researcher and participants during fieldwork; Demonstration of researcher’s influence on stages of research process; Evidence of self-awareness/insight; Documentation of effects of the research on researcher; Evidence of how problems/complications met were dealt with; Ethical committee approval granted; Clear commitment to integrity, honesty, transparency, equality and mutual respect in relationships with participants; Evidence of fair dealing with all research participants; Recording of dilemmas met and how resolved in relation to ethical issues; Documentation of how autonomy, consent, confidentiality, anonymity were managed; Sufficient evidence for typicality specificity to be assessed; Analysis interwoven with existing theories and other relevant explanatory literature drawn from similar settings and studies; Discussion of how explanatory propositions/emergent theory may fit other contexts; Limitations/weaknesses of study clearly outlined; Clearly resonates with other knowledge and experience; Results/conclusions

evidence for inclusion in systematic reviews, Dixon-Woods et al (2007) concluded that while the use of frameworks tended to produce less agreement between assessors, they did sensitise them to methodological issues.

What does it mean to try to put into practice, both in terms of primary data collection, and also in terms of reading and reviewing the work of others, the assessment of quality in qualitative research? The next section involves two attempts at reviewing what it means to put into practice the assessment of quality. The first draws on primary data collection experiences of demographers, illuminating the sorts of researcher reflexivity around data collection, analysis and presentation that many of the frameworks for quality assessment refer to. The second part demonstrates the sorts of issues inherent in trying to assess or read for quality when examining published literature in demography.

3.1 Dealing with quality: perspectives from the demographic field

At every stage of qualitative research a key dimension is the inductive and iterative nature of the research process. Words and speech are fundamental data source and thus it is essential that respondents are willing to talk. Part of the research process is therefore identifying the best ways to induce respondents to talk openly or debate the issues under study. However, the ways people talk, what they are willing to talk about and with whom and under what circumstances are integral not only to data collection, but also to analysis and interpretation. Good qualitative research needs to allow for the new and unexpected to emerge, be identified - hopefully by both field workers and researchers if they are not one and the same - and then be developed. In the same way that qualitative data coding should be inductive and guided by the data, so too should data collection. This need for responsiveness and flexibility is present throughout the research process, not only, but particularly for qualitative research in demography dealing with powerful but and personal issues in people's lives: death, sexuality, reproduction, marriage.

Under such conditions, how should qualitative demographic research be designed? And what factors and risks should be considered during this planning process? For each of the following dimensions - choosing the appropriate qualitative instruments; number and sampling of respondents; selection and training of interviewers; the fieldwork; data management and analysis - we briefly describe the main issues and use examples of our own work. The examples we use are mainly taken from studies conducted in sub-Saharan Africa or with African populations (Appendix 1).

obviously supported by evidence; Interpretation plausible and 'makes sense'; Provides new insights and increases understanding; Significance for current policy and practice outlined; Assessment of value/empowerment for participants; Outlines further directions for investigation; Comment on whether aims/purposes of research were achieved

3.1.1: Which qualitative methods to choose?

Method choice is context sensitive and the inappropriateness of certain methods may only emerge with hindsight after fieldwork. In the 1999 Senegal research project our purpose was to combine different qualitative methods to investigate fertility decision making strategies. As part of a comparative study with Zimbabwe the original intention had been to use identical methods in both countries. The Senegal research was to repeat the approach used in Zimbabwe where the whole team spent two weeks in each research site using diverse methods such as free-listing, in-depth interviews, photo-stories and focus groups. Interview guides were directed towards the demographic research hypotheses around links between child mortality and reproductive decisions. But with collaborators trained in contrasting academic disciplines (demography in Zimbabwe, anthropology in Senegal), different expertise of the students allocated to the fieldwork and a strong belief amongst the Senegalese team that a more indirect approach was best in the Senegalese context, the Senegal approach allocated pairs of fieldworkers to spend two months in each of the three research sites (Dakar, a small town and a village) with a phase of observation and immersion before initiating in-depth semi-structured interviews and focus groups.

The methods were not equally successful in each site: focus group discussions worked well with urban women but were inappropriate in the village where potential participants all knew each other well and would have been uncomfortable addressing the study topics in front of kin and neighbours. Informal group discussions were effective in the village when young people gathered in the researcher's room in the evenings to chat. Focus groups with urban men were not a success: men disliked the informality of a method which encouraged them to debate and discuss amongst themselves and wanted a formal situation whereby the facilitator asked a question and each answered in turn. Furthermore men would not challenge or question accepted social norms in public, especially those around religious interpretations of appropriate reproductive behaviour. In private in-depth interviews a few men were more open to debating such issues. Whereas in Senegal respondents seemed more inclined to be frank and honest in private in-depth interviews, in Zimbabwe the interviewers believed respondents were far more likely to tell the truth in the more public focus groups. An essential stage in qualitative research is a conscious reflection on the quality of the data produced by different methods: cultural traditions and norms and gender roles may have a substantial impact on the success or failure of individual approaches.

3.1.2: Sample size and strategies to identify respondents

Qualitative research does not involve big numbers - close attention must be paid to the selection or sampling of respondents - it is essential to document these choices and to reflect on particular biases that may have been generated. A snowballing approach to respondent recruitment is often used which can be very effective if a homogenous sample is needed with people with

similar characteristics or attitudes. However it becomes inappropriate if a wider understanding is required from a diverse socio-economic range of respondents.

Such diversity was sought in recent work on the effects of out-migration on family dynamics in a small town in Senegal (Randall and Mondain, 2009; Mondain and Randall, 2009). In the absence of a sampling frame we used the six neighbourhoods in the town as clusters. On the first day the interviewers went out with pre-selected characteristics of respondent (by age and sex); starting from the neighbourhood chief's house they counted 10 houses and at the 10th house asked if a respondent corresponding to the criteria lived there and could be interviewed then or later; if not they continued to the next house. Subsequent respondents were identified by counting a further 10 houses from the preceding interview. This achieved a far more heterogeneous sample than would have been achieved through snowballing although there was probably a bias towards unemployed individuals who were at home in the day. If several individuals of the required characteristics were resident there was a tendency to select the person present rather than the absent one. However in a town where there is little formal employment and few fixed working hours this was probably not a major bias but would have been a major problem in a large city.

If sampling frames are available they can help in selection of respondents; however, the small target numbers of respondents in most qualitative studies often require quotas to ensure adequate socio-demographic diversity of the participants. Representativeness is not the key issue in most qualitative research however: what is much more important is a willingness to participate in the study. Given that often the sample size will be less than 50 respondents, the one or two individuals representing a particular age/sex/education category cannot and should not be seen as representing that group: overall, qualitative data is aiming to understand perceptions, processes, constraints, dilemmas and uncertainties and should not be used to infer to the wider population of similar individuals. Since a successful social interaction between researcher and respondent is an essential dimension to good qualitative data, important issues in sample selection are (a) identifying respondents who are willing to talk and generating an environment where they are able to do so freely and (b) reflecting consciously on possible biases in the data, where possible taking steps to address these.

The issue of selecting key informants is complex. One possibility is to decide on appropriate key informants *a priori*, based on discussions with local colleagues. In the African context, key informants are usually the local leaders, political or religious, social and health workers, teachers, etc. Qualitative approaches leave the door open for others to emerge as key informants during fieldwork and the identification of key informants should be an iterative process throughout, with interviews identifying new, influential characters in the

community who can be followed up. In the particular (but typically demographic) case of Demographic Surveillance Systems (DSS) the permanent interviewers working in these sites often become key informants as well (Mondain & Bologo, 2007).

3.1.3: Selection and training of interviewers

Given that qualitative research is largely about exploring understanding, perceptions and dilemmas, communication in an appropriate language is incredibly important. Good mastery of the local language is essential not only for basic communication but also for the success of the whole qualitative research endeavour. This requires good local interviewers. Herein lies a dilemma, especially if working in rural areas or with poorly educated urban dwellers. Either one selects research assistants with primary or unfinished secondary education who are best able to develop excellent rapport with respondents with whom they are likely to be socio-economically and educationally close, or one selects university students or graduates who are better able to understand the intellectual dimensions of the study and may even want to integrate the work into their own dissertations, but who may not always be as good at developing rapport. There are arguments for and against both strategies. Furthermore, whichever strategy is chosen and however good the training, there are some individuals who are just excellent at inducing respondents to talk and others, who because of their innate personality, find it much harder. Good training and team building (see below) can go some way to improving the quality of data from less gifted interviewers.

A key issue in qualitative research is translation and ensuring that issues in translation are at the forefront of interviewers' minds. Key concepts around demographic research dealing with birth, death and life transitions are often difficult to translate, and yet most qualitative research in Africa ultimately depends on translation both for analysis and publication. In the 2007 Senegal research project a critical dimension of the data collection process was identifying and discussing local core expressions which had very specific meanings. For example the Wolof term '*mougn*', mostly used by women and referring to their need for being patient and obedient in the context of their daily lives and relations with in-laws and husbands could not be translated by one French/English word (namely 'patience') as this would have led to the loss of the complex meaning of '*mougn*'. Regular interactions between researchers and interviewers, who simultaneously become key informants, are vital; good transcripts of qualitative interviews will often retain words in the local language annotated with long detailed explanations of meaning.

Furthermore, producing qualitative data implies specific skills in the art of discussion and building rapport. Matthews (2005) discussing the issue of studying marriage emphasizes the fact that researchers should consider "the interaction not as an interview but as an episode of participant observation." By this she refers to the need to consider the specificities of local

communication and modes of interaction¹⁴; those may indeed vary significantly between different socio-cultural contexts, and depend not just on the language spoken but also to belonging to certain ethnic, social or caste groups.

For example, in the research undertaken in the Sereer community in 1999, interviews were conducted jointly by the principal researcher and two interpreters (male and female to match with the sex of the respondent); it became obvious that as soon as a praise-singer or blacksmith (caste) was interviewed, interpreters showed a lack of interest and would not make the same efforts to generate good rapport. Because they were local but not themselves caste, their relationship with the respondent was largely determined by the acceptable social relations in their daily lives between themselves and a caste person from their community. Had interpreters been outsiders, with more education and less embedded in the community, these problems would not have occurred; on the other hand, it was made clear that most participants from this area would have not talked openly to an outsider (Gokah, 2006). Pursuing this dilemma of outsider/insider in the community, the quality of the interviews can also be jeopardized when respondents are aware that the interviewer is asking questions on topics he should know about. This excerpt from an interview conducted with a 43 year-old man in French by the researcher and translated into the local language by the interpreter demonstrates some of the complexities around selection of interviewers: should they belong to the community or be external while at the same time mastering the language?

Q [in Sereer] : She asks why did your family prefer your first wife for your marriage ? What were their criteria?

A [in Sereer] : Yes, ... and you know that. In the past, people would say: go and see the daughter of this man because she is well brought up, her father has such and such qualities, her mother is as such. (...). It is for this reason that our elders made a rigorous selection of the girl to marry. I guess you remember this old selective method of the girl to marry. (Senegal, Sereer community 1999)

In order to maximize the quality of qualitative data it is essential to focus on understanding the goals of the research during the training of interviewers. Here again there are trade offs between well-educated students from outside and local interviewers with less education. Interviewing requires skill at probing and identifying 'pockets' of potential information left open by informants and windows of opportunity to pursue potentially interesting avenues which is only possible if the interviewers can see this potential because they understand the issues. In selecting interviewers, different criteria must be considered: academic training, former fieldwork experience (not always an advantage if different research methods are being used), and study-specific criteria such as age, sex, marital status, ethnicity. Training

¹⁴ Studying this was the aim of a project conducted in five different DSS to explore the problems related to standardized inquiries in various socio-cultural contexts (Mondain & Bologo, 2007).

interviewers in qualitative methods is particularly challenging as it is related to the art of communicating, itself culturally specific. We need to consider the extent to which external research teams (often from the North) are able to provide adequate training in interviewing in a context that is foreign to them. Certain general guidelines can be discussed in common (such as the art of probing, allowing time for silences, etc.) but in the end, interviewers are those who know how to explain, ask and create a climate of mutual trust - and some are inevitably better than others.

3.1.4: Fieldwork

The typical demographic qualitative inquiry is based on an interview guide generating fluid situations in terms of adapting the questions and conducting the interviews. It is essential for the principal researchers to maintain close contact with interviewers both in order to identify key contextual dimensions while analyzing the data later on and to ensure that everyone is aware of new issues which emerge as fieldwork progresses. Regular field meetings are an essential component of the fieldwork and generate team spirit and good collaboration. .

Although qualitative research training usually emphasizes the need to avoid leading questions and those with yes/no answers it is also important to find ways of getting interviewers to encourage informants to focus on specific incidents rather than their more general 'feelings' (Matthews, 2005) which can end up being very vague and indeterminate - with little idea of how they impinge on actual life. This was a challenge for the 2007 fieldwork in Senegal where we wanted to establish the differences between marriages and family life within migrant households versus households with no migrants. The interviewers, knowing our interest in migration, found it difficult to avoid direct questions which just generated opinions about migrant households totally ungrounded in any specific events or experience. Good qualitative data on demographic issues usually needs to be well grounded in personal experience.

Systematic transcription after each interview is best undertaken by the interviewers themselves as rapidly as possible after the interview. This allows them to include good contextual information as well as non-verbal communication which can be critical in interpreting what people say and how they say it.

Q Could the fact of losing a child push a husband to impose another pregnancy on his wife?

A (in a very serious tone of voice) What are you saying? But it's not linked. The death of a child and wishing for a pregnancy, it's not linked. One makes a child with God's blessing. If He takes it back, you can only wait until He gives it back to you...I can see no possible calculation.

Dakar

1999

- I Uhuh. An if you have children yourself are you going to do the same thing [foster out your child]?*
- R (very quickly) Ah no, no. That never, never, never. Never foster out my children. I would not even foster my children to my own sister. I am going to bring up my children myself.*
- Small town Senegal 2007**

In both these cases the non-verbal communication adds to the intensity of the statements. Such additions to transcripts are only likely whilst the transcriber still remembers the details of the interview. Interviewers should also be encouraged to report the quality of the interview and whether they felt the respondent was telling the truth, was happy to talk or was disinterested, without being made to feel that a poor interview necessarily reflects badly upon themselves. Other contextual information is very important and contributes to the quality of interpretation.

The quality of the data management during the fieldwork is a key preliminary step to good data quality and therefore analysis. We give three examples, one where the data management was closely followed up during the fieldwork, one where mistakes were made, and one where power relations within the team made it difficult to obtain good quality transcriptions.

During the 2007 fieldwork in Senegal, the two principal researchers were present in the field, supervising the interviewers who had been hired and trained by one of them. Every day, each interviewer undertook one interview in order to translate and transcribe the same day. Transcription had to be completed before conducting further interviews. The transcript was then read by the researchers generating comments on the way the interview was conducted, where probes had been useful and also for questions on issues they were not familiar with (especially cultural dimensions) and specific words used. Every evening a team debriefing involved discussion of the day's work and comments on previous days' transcripts; there were frequent suggestions about good questions to ask or issues that should be pursued. This continuous process of discussion, reading and immediate follow up of the interviews made the further data management and coding easier and more efficient and was also an effective way of providing training to the weaker and less confident team members.

In 2005-6 research was conducted in Montreal (Canada) on the transition to adulthood of first generation young African immigrants using a team of four trained interviewers supervised by a PhD student. For various reasons a different person transcribed the interviews, leading to long delays between interview and researchers reading the transcripts, rendering it impossible to pursue gaps or lack of clarity. Without regular and timely follow up there was no opportunity for comments on interview quality and no improvements over the course of the project.

In the 1999 Sereer marriage behaviour project interviews were conducted by the principal researcher with two interpreters who were local residents who spoke French but did not write it well enough to translate and transcribe interviews which was done by another. Here there was an issue of transcription quality which reflected relations between different team members: the transcriber was older and better educated than the two interpreters working with the principal researcher, who at that time, was a young female PhD student. Power relationships based on age, gender and education level rendered it difficult to persuade the transcriber to remain faithful to the recording because he argued the level of Sereer language was poor and therefore a direct translation into French could not be understood by the researcher. As a result he 're-wrote' the interviews into an academic French style totally dissonant with local peasant forms of expression.

3.1.5: Presentation of results

Presenting demographic qualitative research in a way which allows the reader to assess the quality of the research must include detailed description and reflection on methods. Who are the respondents: how were they selected? What are the potential biases in selection? The inherent qualities and characteristics of respondents allow the authors to discuss the potential for generalizations from their data and its limits.

In most cases, qualitative demographic research analysis is backed up using quotes to illustrate key analytic points, with contextual socio-demographic information about the respondent. The core question here is: how much data should be included and how should it be selected? It is important to note that the data appears through the form of excerpts and those constitute *evidence*, *not examples*; usually they are chosen from among a set of other similar excerpts by the authors although rationale for selection of quotes must be justified in the text with indications about whether the point being made was heard frequently or is a rare exception.

Several questions in presenting the data emerge:

- Should qualitative data be quantified? Something that is easy with software such as N6. In other words should statistics be generated about the numbers of respondents who mentioned certain aspects or amount of text on a certain topic? Such statistics give a false picture for various reasons: the fluid nature of an in-depth interview along with following up issues of interest to the respondent mean that not all questions are systematically asked of everyone. Furthermore silence is not always a sign of absence of importance (Randall & Koppenhaver 2004). On the other hand indications of scale are helpful with typical 'fluid' descriptions such as 'many respondents said ...'; 'most women mentioned ...', etc. ?

- How long should the quotes be? If too long the paper may not be publishable and the analytical dimension will be undeveloped; This is critical for articles but can be got round in books where very long quotes may be presented (see Cicourel 1974, Brand 2001)
- Should there be quotes for each point raised or can discourses be summarized instead of quoted?
- Should the question which has led to the answer be systematically included in the quote? This is an important issue because including context allows the reader to assess data quality and identify leading questions, but too much context increases the word count which may be problematic for journal articles.
- Should the quotes be included in the written paragraphs or isolated in the text? Although this is more an issue of style; including quotes in the written paragraphs may appear as a fluid way to incorporate the data in the text but requires that the excerpts are small soundbites taken out of context.
- Should each quote be commented on and analysed or should they just serve as illustrations allowing the reader to perform their own analysis.?
- Qualitative research often includes different types of supporting information: field notes, photos, informal discussions, etc. To what extent should these be reported on and used in the paper, thus breaking with the traditions in demographic publications based on data gathered systematically? This issue is important as our ability to analyse what respondents say depends on this 'parallel' work of presence, informal discussions, observation, etc. How do we include these dimensions into the results and conclusions?

Most of these issues remain up for debate but may be important in trying to publish qualitative research, especially in traditional demographic journals.

In order to assess the quality of qualitative research there are various dimensions: the appropriateness of the methods used; identification and selection of respondents; the organization of fieldwork; and the way the analysis was conducted. The presentation of quotes is the visible part of the iceberg of the research process and the choices made in their selection needs to be documented. It is clear that all this detail cannot always be included in a paper's methods section because of space constraints. On the other hand these dimensions are key in the assessment of the quality of the data and should at least be addressed while reviewing papers; in other words, reviewers should be made aware of these aspects and not hesitate to require authors to justify (not in the paper necessarily) their research design in more details. Website references can be a useful tool here where details of research methods can be

posted. Furthermore there are debates around the free access to the whole qualitative database and researchers with confidence in the quality of their data should find ways of making this available once suitably anonymised.

3.2 Assessing quality: published demographic literature

This section reviews a sample of articles using qualitative methods (both singly and alongside quantitative methods) published between 2001 and 2005 in major demographic journals¹⁵, assessing them for the quality of methodological reporting.

3.2.1 Method

We selected a subset of articles to perform this analysis. One author did an exploratory analysis for 13 articles chosen randomly (without paying attention to the journal title). She read each article entirely and handled each article as if she was doing an article review for a journal. From this analysis, she developed a "rating tool". She stopped reading new articles when she reached saturation and her rating tool remained stable. She completed the rating tool for each article, and analyzed the results of the rating item by item (see below). The results are summarised separately for operation-research articles

¹⁵ To select our sample of articles, we replicated the search of Randall and Koppenhaver (2004) for the 2001-2005 period. As these authors did, we searched the following journals: Demography, Perspectives in Sexual and Reproductive Health, International Family Planning Perspectives, International Migration Review, Population -English edition, Population Studies, Population and Development Review, Studies in Family Planning. All these journals are on JSTOR, except for International Migration Review. As they did we used the following key words to search the abstracts: anthropology / gical / gist, ethnography / phic / pher, qualitative, focus groups, in-depth / semi-structured interviews, participant observation. We then experimented with different keywords in the same set of journals and time period. We introduced the following keywords: sociology / gical / gist, mixed methods, and did not find any additional papers using qualitative methods (we did not consider theoretical or review articles whose abstracts mentioned one of these words). To be consistent, we eliminated 5 articles which were reviews of qualitative studies or referring only to anthropological theory, without use of qualitative data. Also, noticing variations in the spelling of some of the keywords of interest, and the use of "interviews" as a short-hand for qualitative interviews, we introduced the following key words in the search: focus-group, focus-groups, semistructured interviews, interviews. With these new key words, we found 21 additional articles using qualitative data in the set of journals for the 2001-2005 period. Altogether, we thus identified 70 articles for the present analysis. Focusing only on articles available in JSTOR (that is, eliminating International Migration Review), we worked with 61 articles.

Our sampling criteria maximize ease of access (of concern for the Ouagadougou-based author). The drawback to the JSTOR database is that the articles selected are not among the most recent ones, but the phenomenon of interest is not likely to change rapidly. Also, the JSTOR database does not contain every "demographic" peer-reviewed journal. Finally, JSTOR over-represents journals dealing with family planning. Out of the 61 articles, 16 were published in either Demography, Population, Population Studies or Population and Development Review, and the remaining 45 in International Family Planning Perspectives and Perspectives on Sexual and Reproductive Health. To minimize this problem, we aimed at selecting for the analysis the same number of articles from each journal.

(tests or design of reproductive health interventions) and for fundamental-research articles (all other studies), because these two types of research have very different methodological needs. Another author then applied this grid to 13 additional articles (chosen to even out the number of articles by journal title); this second analysis modified only marginally the results of the first analysis.

Each article is first graded for its overall methodological strength (evaluating only the qualitative part). To assess this parameter, we ask whether the research question has been answered with the methodology used. The answer is either "poor", "medium" or "good" (poor is taken here as the equivalent of a "reject" diagnosis when reviewing an article submitted to a journal, medium is taken as a "revise and resubmit" answer, and good is the equivalent of "accepted"). Each article is then graded for the completeness of its methodological description, along six items, scoring 0 or 0.5 or 1 according to the degree of completeness. The six items are:

- 1) Rationale for the qualitative data collection tools chosen mentioned
- 2) Presence of elements which enable the reader to verify the quality of the material collected: information on interviewers, their training, information on data format, language and translation, on interview conditions; assessment of the quality of the displayed qualitative data
- 3) Description of the choice of site, sample recruitment, sample size, and rationale for site/ sampling criteria and size
- 4) Description of the analytical method used
- 5) Display of results, and possibility for the reader to verify the demonstration
- 6) Cross-validation of the results.

Development of this grid was informed by the authors' years of reviewing qualitative studies submitted to peer-reviewed journals.

3.2.2 *Results*

Out of the 26 articles selected (from a total of 61), two made only minor use of qualitative data, and we discarded them from the analysis. Of the remaining 24 articles, five were operation-research articles and 19 were fundamental research studies (Appendix 2).

3.2.2.1 *Qualitative data and analysis in operation-research*

The reviewed action research studies (5) have overall a fairly appropriate use of qualitative methods: 3 are rated as methodologically "good" and 2 as "medium".

I. Data collection method

Focus groups (FG) are used in all these studies, often combined with in-depth interviews (IDI), and in one case ethnographic notes, and in another, free listing techniques. However, the reason for the choice of data collection tool is never discussed

II. Quality of the collected data

Most of these articles provide information on the FG composition and on the characteristics of the moderator, on the interviewer who conducted the IDI, on the format of the data (taped and transcribed), language and translation. Short quotes are used in the display of data (except in one case see below), which do not allow the reader to assess personally the quality of the data, for example, to what extent responses are induced by the interviewer.

III. Choice of site and respondents

The sampling criteria are usually explained (the full range of diversity among clients and providers is usually aimed at), and possible impacts of recruitment specificities on results are discussed. The size of the sample however is never discussed, but the clinical design (a "before" "after" design) usually renders the question of number irrelevant.

IV. Analytical method used

All these studies use thematic (content) analysis, and usually described it as such. This analytical approach is adapted to the goal of these studies, which is descriptive. The exception is the "medium" rated article, which does not describe its analytical method. In that article, the quantitative results are placed in the "results" section and the qualitative results are introduced only in the discussion (no quotes), giving a different status to the two bodies of data and rendering it impossible to verify the quality of the qualitative data and demonstration. We return below to the issue of researchers treating qualitative data, that is the content of individuals' discourses, as "the final truth" or as "the scientific answer" to their question.

V. Display of results

Qualitative results are usually illustrated with numerous short quotes (often not contextualized beyond the source: provider or client). Usually one quote is provided for every point made (every result), so that readers can verify that the analysis of the interventions' failure, success or planning is sound. Often results are translated into concrete policy recommendation.

VI. Cross-validation of analysis and results

Four studies contrast the points of view of providers and clients on the intervention, and 2 out of 5 also use quantitative data to rate programmatic success or failure and contrast these results to the qualitative findings. In one study the contrasted groups are differentiated by ethnic background. In one study, cross-validation of thematic coding is performed, and in another one, cross validation is done through a sequential type of data collection. There is at least a minimal check on the validity of the results in every operation-research study reviewed, and often more than one check.

3.2.2.2 *Qualitative data and analysis in research*

If the use of qualitative data and analysis is overall satisfactory in operation-research published in demographic journals, more methodological problems appear when researchers use a qualitative approach to study social phenomena in the same discipline. Out of the 19 research articles reviewed, only 11 are rated as "good", 5 as "medium", and 3 as "poor". Studies rated as "good" or "medium" vary substantially in the quality of the methodological description ("good" articles have sometimes very short methodological description, because of the use of short-hand such as "grounded theory", "standard ethnography", "life histories"; the methodological description may be spread throughout the text). However, it is very significant that all studies rated as "poor" are lacking methodological information.

I. Data collection method

None of the 19 studies explains their choice of method (FG, IDI, participant observation). However, these are exactly the kind of papers where you would expect a thorough discussion of the tools. Since different types of tools are adapted to different types of research question, these choices need to be made explicit.

II. The quality of the collected data

Most article report on IDI data format, which are always taped and transcribed and sometimes translated. Few articles (but two out of the three "good" articles) report on who the interviewers are and their training and interview conditions; this information is crucial to identify possible interviewer bias (for example, in one case, the interviewer is a White male researcher and the respondents are non-White young males). One test of the quality of IDI data is first-hand access to the data themselves. This entails displaying long quotes for IDI (entire interviews or interviews extracts would be ideal, but not possible in an article format). Only 2 out of the 19 articles display long quotes. In one case, the article is rated as "good" and the data shown is of excellent quality; in the other case (a study rated as "medium"), the data shown is of very poor quality, although this information could not be deduced from the description of data quality. Poor quality IDI data is to be inferred when answers are short and interviewer's questions or probes are numerous (the likelihood of interviewer's bias is then very high).

III. Choice of site and respondents

Site(s) and respondents choice are linked to the stage of the research (exploratory, theory building, theory testing). In exploratory research, one would choose a typical place and respondents (both engaging in and avoiding the practice of interest). In theory-building studies, one would look for maximum diversity given the phenomenon of interest. In theory testing,

one needs to have the comparison groups necessary according to the hypothesized factors at work. All studies describe the choice of sites and sampling criteria. However, they do not usually relate their options to the type of research conducted.

Sample size is related to the previous point: in exploratory research, the in-depth interrogation of a few respondents or in-depth observation of a small site is relevant. In theory-building or testing studies, maximal diversity has to be reached, which implies a relatively large number of sites and observations or interviews. Ideally, the size of the sample is not set in advance, but is defined during fieldwork, when the researchers "reach saturation", that is, does not meet new situations when conducting new interviews or observations. For planning and budgeting reasons, it is often difficult to follow the principle of saturation, but a relatively large number of cases can be assumed to enable saturation. Saturation is reached earlier when the sample is limited to one group, and more interviews are necessary to reach saturation in several targeted groups (and thus typically in theory-testing studies). None of the 19 articles discusses sample size or saturation, although all "good" and two of the "medium" studies feature a large enough sample size to infer saturation purposes. One "poor" and one "medium" study suffer from an ill-defined study purpose, mixing an exploratory design (a typical place, typical respondents, few respondents, many observations on the respondents) with a theory-building design (foster as much diversity as possible by diversifying sites and recruitment criteria, and a single interview).

IV. Analytical method used

The analytical method used is always described very succinctly: most articles mention that they use "coding", one mentions coding and "grounded theory", one mentions "typology building", and one mentions coding and a comparative study design, one mention the method of "emphasis of response". The 4 latter studies are rated as "good". "Coding" is a vague notion, and insufficient to describe qualitative data analysis in fundamental research. The analysis can involve thematic coding in a first, descriptive analytical stage, but needs to go beyond it, since the idea is to understand the structuring dimensions of the phenomenon of interest. As mentioned already, the method used to get at the underlying social factors of a phenomenon is *comparative analysis*. Theory testing studies imply a comparative study design from the start, and narratives coded into themes are thus compared across different groups. Thematic coding is not sufficient to explain this procedure: a more exact wording would be "thematic coding" and then "comparison of themes across pre-defined groups". Theory-building studies also involve comparisons. The groups of comparison are identified through data analysis (typology building or grounded theory): the cases or situations are compared one by one and grouped, and contrasts between groups help the researcher understand what dimensions

differentiate the groups, that is, what social factors structure the practice or behaviour. "Coding" is not enough to describe this analytical procedure, although it is part of it; narratives are coded, and then cases are compared across themes, and then grouped, and the groupings further analyzed. As already mentioned, even exploratory type research involves comparisons: researchers compare minimally between situations of norm compliance and deviant situations to uncover the main social mechanisms underlying shared representations and practices (see the importance of the study of deviancy in anthropology); here, given the smaller number of cases involved, a formal coding is not always necessary.

The majority of the studies relying on thematic coding use a qualitative data coding software. However, using a qualitative software is not related to the final grade received by the paper. Articles rated as "poor" use a software, and articles rated as "good" do not use a software, and the reverse is true too. (As a side note: one of the "medium" rated study using a software states that the software processes the data and does the coding, which is wrong: these software enable the researcher to code a text into the codes of his or her choice, and to retrieve easily the texts coded under a given category).

V. Display of results

Most studies display multiple quotes (usually short quotes, which makes it difficult to assess data quality); and the quotes are usually situated (at least some characteristics, but some go into details into each case). Both the content of the narrative and the characteristics and life history of the respondent allow for the reader to check on the argumentation, and the more details we have on both, the easiest it is for the reader to judge of the validity of the results. Because of the constraints of the article format however, only two articles provide large amounts of suited narratives (the two articles are graded as "good"). A few studies prefer to summarize the results without quoting "data". They can as well and as satisfactorily answer the research question, but the reader has less opportunity to judge the validity of the results.

VI. Cross-validation of results

Most "good" or "medium" studies cross-validated coding or contrast qualitative and quantitative results. Studies with no cross-validation strategy are rated as "poor" or "medium".

3.2.3 Discussion

Qualitative studies of poor quality usually have incomplete methodological descriptions, suggesting that studies of lesser quality are conducted by researchers who are less aware of the methodological standards of qualitative research, and thus less able to describe and defend them. Some excellent qualitative studies display complete but very short methodological information,

which is fine among knowledgeable readers, but does not help promote good standards for qualitative research in a quantitative field. Editors, professors, tutors, funders and other research standards setters could increase the quality of qualitative research in demography by requesting complete and detailed methodological description in the works they review.

Qualitative studies in the field of demography could also gain in quality by paying more attention to issues that are important to quantitative studies in the field: data quality and cross-validation. Key to qualitative data quality is the interviewer or observer her/himself, so detailed description of the interviewer - interviewee interactions and discussion of possible biases are to be promoted. Similarly, systematic display of longer quotes (along with greater awareness of the signs of poor qualitative data quality) will improve quality checks. Editors need to be aware of the implication of short article length on qualitative studies, and should be encouraged to allow for longer texts¹⁶. Cross-validation of the analysis, and not only cross-validation of coding, encourages reliability; multiple modes of cross-validation should be encouraged (existing results, other data sources, confronting the narratives of respondents with diverging interests, etc.)

But the main problem with qualitative research published in demography (aside from operation-research) is that the methods of qualitative analysis (grounded theory, typology building, comparative analysis) are poorly reported, and appear to be poorly understood. The development of qualitative curricula and teaching material may help here, especially since researchers who perform these analyses well do not explain how they proceed, perhaps because their readership is usually made of other specialists. Many of these authors have developed these techniques after long immersions in another discipline: sociology, gender studies, psychology, anthropology.

4. Discussion

When we read or review research using qualitative approaches, we will inevitably apply criteria as to whether the research is successful or not. In the same way that an experienced quantitative researcher can distinguish between a good and a poor study, so can an experienced qualitative researcher. For a quantitative researcher, or a novice qualitative researcher the evidence presented by a piece of qualitative research is likely to look very different from that produced by, say, secondary analysis of a large-scale quantitative dataset.

Extended "checklists" will be unlikely to gain (universal) endorsement from those trained as ethnographers, but might be attractive to demographers who

¹⁶ Some journals, especially in the medical sciences, permit longer maximum word lengths for qualitative as opposed to quantitative articles.

want to better understand how credible the qualitative data are. This category not only includes reviewers, but also includes those who have not collected their own data but have subcontracted data collection to others, which is often done in multi-method research projects. Our aim here is not to achieve consensus -unlikely and ultimately undesirable given the breadth of qualitative epistemology - but to identify "points of entry" for the non-specialist. It is of value to have a framework for appraising the quality of qualitative research, so long as it is used with care. By drawing together the state of the art in terms of frameworks for quality assessment with experiences of primary data collection and research reviewing, we demonstrate how common-sense approaches can improve the quality of qualitative research and its reviewing. For many qualitative researchers, concerns about the "criteriology" of assessing quality are well-founded if criteria become highly procedural, mechanistic and prescriptive. What is at issue here is thinking about informed judgement of quality, both when research is being done and when it is being read by others. Many of the frameworks use open-ended questions to prompt judgement, and take into account that quality standards will inevitably be shaped by research question(s) and whether they have been answered.

There are, however, important practical issues (and solutions) raised by these assessment of quality. For example, transparency in reporting is linked to issues of under- and non-specification, especially when research reporting is constrained by journal word counts. Transparent research can be presented in a way that can be both appraised and (re-)used by others. Authors should be able to make innovative use of internet-based repositories of their research instruments and/ or data. At the very least, authors can make available copies of their research instruments.

Assessing quality needs some expertise in the conduct and use of qualitative research, necessitating improved support, education and guidance for non-experts. Insight and interpretation are key aspects of qualitative research, and are the most difficult to appraise. But any research needs to have, and justify sound methods and defensible conclusions. By focusing on both the production and consumption of qualitative research in demography we hope to shed light on what is involved in being a good researcher in general. When we read, or review, research using qualitative approaches, it is inevitable that we should apply criteria for deciding what a good study is and what is not. By highlighting what is involved in producing and reviewing qualitative research in demography we contribute to an articulation of what is often implicit - the decision about what is good and what is not.

Bibliography

- Allison, Paul D., 1984, *Event History Analysis. Regression for Longitudinal Event Data*, Sage University Paper No 46.
- Barker, Gary K et Susan Rich, 1992, « Influences on Adolescent Sexuality in Nigeria and Kenya: Findings from Recent Focus Groups Discussions », *Studies in Family Planning*, 23,3: 199-210.
- Basu A. M. And P. Aaby (1998). *The methods and uses of anthropological demography*. Oxford : Clarendon Press.
- Bernardi L. and I. Hutter (2007). "Anthropological Demography", *Demographic Research* 17(18): 541-566.
- Bledsoe C, F.Banja, A.G.Hill, 1998. Reproductive Mishaps and Western Contraception: An African Challenge to fertility theory. *PDR* 24(1)
- Boaz, A. and D. Ashby (2003). Fit for purpose? Assessing research quality for evidence-based policy and practice. ESRC UK Centre for evidence based policy and practice Working Paper: 18.
- Bonvalet C. and Z. Andreyev (2003). « The Local Family Circle", *Population (English Edition)*, 58(1), 9-42.
- Bozon M. (2006). "L'apport des méthodes qualitatives en démographie ». Chapter 136 in : / [*Démographie : analyse et synthèse. 8, Observation, méthodes auxiliaires, enseignement et recherche*](#) . - Paris : INED, p. 439-463.
- Brand Saskia (2001) *Mediating means and fate*, Leiden, Brill
- Bryman, A., S. Becker, et al. (2008). *Quality Criteria for Quantitative, Qualitative and Mixed Methods Research: A View from Social Policy.* 11: 261-276.
- Calves, Anne-E. 2000. "Premarital childbearing in urban Cameroon: Paternal recognition, child care and financial support". *Journal of Comparative Family Studies* 30(1): 443-452.
- Castle S., M.Konate, P.Ulin, S.Martin, 1999. A qualitative study of clandestine contraceptive use in urban Mali *Studies in Family Planning* 30 (3)
- Cicourel A (1974) *Theory and Method in a study of Argentine fertility* New York, Wiley
- Cleves, Mario A., William W. Gould et Roberto G. Gutierrez, 2002, *An Introduction to Survival Analysis using STATA*, Stata Press.
- Coast, E. (2003). "An evaluation of demographer's use of ethnographies". *Population Studies* 57(3), 337-346
- Coast, E., Hampshire K., Randall S. (2007) "Disciplining anthropological demography", *Demographic Research* 16(16): 493-518.
- De Loenzien Myriam, Simon Yana David, *Les approches qualitatives en démographie. Théories et applications*, 2005
- Dixon-Woods, M., R. L. Shaw, et al. (2004). "The problem of appraising qualitative research." *Qual Saf Health Care* 13: 223-225.
- Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B, Bonas S, Booth A and Jones D (2007) 'Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods' in *J Health Serv Res Policy vol 12 no 1 Jan 2007*

- Gagnon, A. & N. Mondain, 2008. "Démographie et santé mondiale : un carrefour indispensable ». In: V. Ridde & K. Mohindra (eds) *Développer et pratiquer la recherche et l'enseignement en santé mondiale au Québec*, Actes du colloque Développer et pratiquer la recherche et l'enseignement en santé mondiale au Québec Présenté dans le cadre du 75e Congrès de l'Acfas, à l'Université du Québec à Trois-Rivières, le 11 mai 2007
- Gokah, T. 2006. The Naive Researcher: Doing Social Research in Africa, *Int. J. Social Research Methodology*, Vol. 9, No.1, pp. 61-73
- Greenhalgh, S. (1997) "Methods and meanings: Reflections on disciplinary difference". *Population and Development Review*, 23(4), 819-824.
- Gueye, Mouhamadou, Sarah Castle et Mamadou Kani-Konate, 2001, « Timing of First Intercourse Among Malian Adolescents: Implications for Contraceptive Use », *International Family Planning Perspectives*, 27,2: 56-62.
- Holloway & Wheeler (1996) *Qualitative research for nurses* Oxford: Blackwell Science
- Kertzer, David I et Tom Fricke, 1997, *Anthropological Demography. Toward a New Synthesis*, The University of Chicago Press
- LeGrand L., T. Koppenhaver, N. Mondain and S. Randall (2003). "Reassessing the Insurance Effect: A Qualitative Analysis of Fertility Behavior in Senegal and Zimbabwe", *Population and Development Review*, 29(3), 375-403
- Kertzer, D., & Fricke, T. (1997) *Anthropological Demography*. University of Chicago Press: and other problems", *Demographic Research* 11(3): 57-94. Chicago.
- Kirshenbaum S. B., A. E. Hirky, J. Correale, R. B. Goldstein, M. O. Johnson, M. J. Rotheram-Borus and A. A. Ehrhardt (2004). "'Throwing the Dice": Pregnancy Decision-Making among HIV-Positive Women in Four U.S. Cities", *Perspectives on Sexual and Reproductive Health*, 36(3), 106-113
- Matthews, S. 2005. "Crafting Qualitative Research Articles on Marriages and Families", *Journal of Marriage and Family* 67: 799-808
- Mondain, N., LeGrand, T. & Delaunay, V., 2004. L'évolution de la polygamie en milieu rural sénégalais : institution en crise ou en mutation ? *Cahiers Québécois de Démographie*, 33(2) : 271-308
- Mondain, N., T. LeGrand et P. Sabourin. 2007. "Changing Patterns in Men's First Marriage among the Sereer in Rural Senegal", *Journal of Comparative Family Studies*, 38(4): 627-644
- Mondain N. et Delaunay V. 2006. « La vie avant le mariage : les grossesses pré-nuptiales chez les Sereer Siin au Sénégal. » In : *Enfants d'aujourd'hui : diversité des contextes, pluralité des parcours* : tome 2. Paris (FRA) ; Paris : AIDELF ; PUF, p. 799-814
- Mondain, N., S. Lardoux, S. Bignami and A. Gagnon, 2008. "Transition to adulthood among African immigrant adolescents in Western countries: evidence from a qualitative study in Montreal, Canada", paper presented at the EPC Conference, Barcelona, July 2008

Mondain, N., S. Randall and A. Diagne, 2009. "Assessing the effects of out-migration on those left behind in Senegal: local family dynamics between change and continuity", paper accepted for oral presentation at the IUSSP Conference in Marrakech, Sept-Oct 2009

Mondain, N. & E. Bologo, 2007. "Ensuring the sustainability of DSS by making scientific results available to lay persons: case studies in Senegal and Burkina Faso", paper presented at the UAPS conference, Arusha, December 2007.

Onwuegbuzie, Anthony J. & Nancy L. Leech, 2005. On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies, *Int. J. Social Research Methodology*, Vol. 8, No. 5, pp. 375-387

Popay, J. (2006). The systematic review of qualitative findings. The 17th International Nursing Research Conference Focusing on Evidence-Based Practice.

Randall, S. and Koppenhaver T. (2004) "Qualitative data in demography: The sound of silence and other problems", *Demographic Research*, 11(3): 57-94.

Randall S. et T. LeGrand, 2003. "Reproductive Strategies and Decisions in Senegal: the role of child mortality" *Population-E*, 58(6): 687-716

Randall, S., N. Mondain and A. Diagne, 2009. "Absent parents in modern Senegal: consequences for child wellbeing", paper accepted for oral presentation at the IUSSP Conference in Marrakech, Sept-Oct 2009.

Riley N et J. McCarthy. 2003. "The essentials of demography", in *Demography in the Age of the Postmodern*, Chap.3, Cambridge University Press, Cambridge, p.35-60.

Rolfe, G. (2006). "Validity, trustworthiness and rigour: quality and the idea of qualitative research." Methodological Issues in Nursing Research: 304-310.

Sandelowski, M., S. Docherty, et al. (1997). "Qualitative metasynthesis: issues and techniques." Research in Nursing 20: 365-371.

Seale, C. (1999) *The quality of qualitative research* Sage

Spencer, L., R. J., et al. (2003). Quality in qualitative evidence: a framework for assessing research evidence: a quality framework. London, Cabinet Office Government Chief Social Researcher's Office.

Tashakkori, A., Teddlie, C., *Mixed Methodology : Combining Qualitative and Quantitative Approaches*, London, Sage, 1998.

Walsh D and Downe S (2006) 'Appraising the quality of qualitative research' in *Midwifery* vol 22

Appendix 1

Year	Title of project	Methods and team	Publications
1999	Rock-Project: relation between the decline in child mortality and fertility decision making in Senegal and Zimbabwe	Individual and couple in-depth semi-structured interviews Focus groups. Team: man/woman pairs in the 3 sites (Dakar, small town and village); international team of principal researchers	LeGrand T., Koppenhaver T., Mondain N., et Randall, S., 2003. Randall & LeGrand, 2003 Randall & Koppenhaver, 2004
1999	PhD project: changes in marriage behaviour in rural Senegal: the case of the Sereer	Individual in-depth semi-structured interviews Focus group discussions (only exploratory) Individual structured interviews Team: the principal researcher and 2 interpreters, one male, one female	Mondain et al, 2004 Mondain & Delaunay, 2006 Mondain, et al, 2007.
2005-2006	Projet “Jeunes en transition” in Montreal, Canada: first generation African immigrant youth transition to adulthood	Focus group discussions Individual in-depth semi-structured interviews Team: 2 men and 2 women, one of them supervising the project for her PhD thesis; one principal researcher	Gagnon & Mondain, 2008 Mondain et al, 2008

2007	Projet: «Les effets de la migration sur les dynamiques familiales locales : le cas d'une petite ville au Sénégal »	Individual in-depth semi-structured interviews Life histories Team: 2 men and 2 women; 2 principal researchers	Randall et al, 2009 Mondain et al, 2009
2008+	Observatoire de Population d'Ouagadougou	Preliminary qualitative phase characterizing residential districts prior to establishment of DSS. Rapid appraisal, observations, key informant interviews, informal discussions, both individual and group	None

Appendix 2

1	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Zulu 2001 <i>Demography</i> qualitative + quantitative Malawi Post partum abstinence	FG: to elicit norms IDI with individuals: understand relevance of these norms to behaviour IDI of “special informants”: what norms they teach	No information on who collected the data and how Local language, transcribed and translated into English at the same time (data recorded: not stated, but implied)	Three areas of the country, ethnically diverse, with different post-partum length Random list of individuals in the survey village (each= common sampling frame), first selection for 22 FG (10 participants), we do not know how, the rest of the list for IDI 22 FG, and 61 in-depth interviews No rationale for why that number, but number high	Use of a software (Nudist). After a first reading, development of first broad codes of norms about abstinence and their rationale. Then, development of finer codes when coding line by line in Nudist. Goal: diversity of norms within each area, and contrasts between the three areas.	Quotes are used (= “clearest articulations of statements and views expressed”) Short quotes, many quotes	Comparison with quantitative survey data No cross-validation of codes	MEDIUM Argues that macro quali studies are not enough, micro quali studies are necessary to make causal inferences (earlier work showed longer abstinence in patrilinear societies). Problem: he explain why abstinence is longer in the Northern region (the end of abstinence is linked to return of menstruation there), but no discussion of why that would be in patrilinear societies and not in others. The explanation remains at the individual level
Completeness	1	0,5	1	1	0,5	1	5

of methodological description							
-------------------------------------	--	--	--	--	--	--	--

2	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Otoide et al. 2001 IFPP Nigeria Why do young people use abortion and not contraception?	20 FG with women aged 15-24 (149 women in total) no	Team of researchers conducted the FG. Nothing on their characteristics (especially sex) In English and pidgin English FG taped, and then transcribed. Notes taken during the FG. No need for translation (local investigators)	Selected to represent the diversity of youth in one city (geographic location within city, occupation of youth, ..). The sexual and contraceptive practices of youth in the city are said to be the same as in the rest of the country. Nothing on why 20 FG, but presumable, same criteria (have enough diversity)	No method of analysis mentioned	Some short quotes (in pidgin with translation or english) as illustrations, within the text, not contextualized	No	POOR
Completeness of methodological description	0	0,5	1	0	0,5	0	2

3	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Legrand et al. 2003 PDR</p> <p>Test of the insurance effect on fertility behaviors</p> <p>Zimbabwe and Senegal</p>	<p>Individuals and couples IDI (each spouse separately)</p> <p>Focus groups</p> <p>FG shown to be of little use among the Wolof (do not like to discuss these issues in public), but useful with the Shona (not the same problem)</p> <p>Less couple interviews and focus groups in Senegal too maximise diversity because lots of heterogeneity</p> <p>Participant observation in Senegalese villages</p> <p>Gender (couple) perspective not used here</p>	<p>Same sex interviewers; description of interviewers training and supervision</p> <p>Because field staff was different, variations across countries in the topics most covered</p> <p>Discussions with field staff to understand what topics respondents did not like, silences are part of the data to understand norms “the sound of silence”</p>	<p>The urban and rural sites were chosen to represent diverse socio eco contexts, within the two countries’ largest ethnic group area</p> <p>IDI Random choice within set categories to maximise diversity.</p> <p>FG: not random choice, but respect pre-set categories of participants</p> <p>72 IDI, 24 FG, 37 couple interviews in Zimbabwe</p> <p>122 IDI, 14 FG, 9 couple interviews in Senegal</p> <p>Numbers not justified, but large enough to insure saturation, sampling strategy to maximise diversity</p>	<p>First coding from the two country teams put in common into one, codes on pre-defined questions or themes and emerging theme codes.</p> <p>NUD IST software.</p> <p>Explanatory analysis: comparison across rural and urban and Z and S to understand social-eco change and reproductive attitudes. Comparative design not explained in the methodological appendix</p>	<p>Numerous quotes, relatively long, situated</p>	<p>cross-validation on coding items, analysis performed in each country and the discussion on the results</p>	<p>GOOD</p>
<p>Completeness of methodological description</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>6</p>

4	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Schuler et al. 2001 IFFP Qualitative Bangladesh Evaluates the impact of a shift in national FP program	IDI Some focus groups Some clinic observations During several years to observe change, retrospective data prior change Some respondents interviewed several times Unstructured ethnographic notes No rationale, but all tools used!	Nothing on interviewers Transcripts. No mention of language or translation	Three rural and two urban sites; in each site sub-areas close and further from the clinique Selected through key informants and hospital records, with set criteria to maximize diversity Several hundred transcripts (IDI and observations in cliniques), about 125 used for this article	Thematic coding using ethnographic software SPdata No further explanation	Lots of cases described to illustrate points, with always a quote within the description of the case	Providers and clients views are contrasted, discourses and observations are contrasted	GOOD
Completeness of methodological description	1	0,5	1	0,5	1	1	5

5	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Kirshenbaum et al. 2004 PSRH Abortion decision making among HIV positive women USA	56 IDI No rationale for the choice, but appropriate to the purpose	open-ended questions, probes, different themes, exact wording of questions on reproduction, 2-3 hours long Information on interviewers Interviews audio taped and transcribed	4 cities Diverse recruitment points: clinics, associations, adds., Discussion of the implications of the age of the sample (relatively old compared to similar studies) on the result No rationale for the number, but large	Method based on grounded theory. Two steps: a first primary coding was done. Then a team of 8 researchers worked on 16 interviews to refine coding, cross-validation of coding procedures. These researchers coded 38 additional interviews (in pairs), until saturation was reached on analysis. Then, researchers worked in pairs more in depth on themes touching reproduction; new finer coding, cross-validation of coding. Explanatory analysis: results compared across groups of women, according to the time of diagnosis and life cycle stage	Many cases described, with a short quotes to illustrate the case	Of coding (and of analysis? Not clear)	GOOD
Completeness of methodological description	1	1	1	1	1	1	6

6	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Gueye et al. 2001 IFPP Quantitative and qualitative Adolescent sexuality Mali	30 FG No rationale for this choice	No description of interview guide Nothing on data collection procedures, on moderators FG in local languages taped, translated into French and transcribed	Male and female FG, urban and rural Purposive sample, screening for educational and marital status Nothing on choice of number	Just mentions: "Coding and analysis using The Ethnograph software"	Describes with short unsituated quotes the reasons to have sex early	With quantitative results discuss the biases of focus group (hypothetical cases are discussed, not personal cases), to explain discrepancies with quanti data (where women declare having early sex out of love, this reason does not appear in the FG)	POOR Consider respondents' discourse as being the truth, not the reflection of an underlying social order. The link between adolescents reason to have sex early and macro social change is just stated, not demonstrated.
Completeness of methodological description	0	0,5	0,5	0	0,5	1	2,5

7	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Vanlandingham et al. 2002, IPPF</p> <p>Tailand, male premarital sexuality</p>	<p>IDI with 10 young men with lots of friends</p> <p>Complex and intimate topic exploratory analysis</p>	<p>Guide elaborated during 5 pre-test interviews. Open ended.</p> <p>By lead author with Thai research assistant, in secluded private area</p> <p>The impact of a Western interviewer is not discussed.</p> <p>No identifying elements, interviews recorded.</p> <p>Transcribed into Thai and then translated into English.</p>	<p>Selected from diverse entry points (by one Thai researcher and Thai research assistant), through personal contacts with one level of unknown between them</p> <p>Number not discussed</p>	<p>Coded and analyzed using The Ethnograph software. Two types of coding: pre-formed codes (guidelines) and emerging codes.</p> <p>Use of results matrix (respondent x themes)</p> <p>The goals is on different topics to give the dominant view as well as the diversity of view points.</p>	<p>Numerous typical exchanges (between interviewer and respondent); long quotes unables to rate data quality (which is poor)</p> <p>Use of Thai concept (in local language in the text)</p>	<p>Cross-validation of coding (the two authors coded until they agreed, the first author coded the rest).</p>	<p>MEDIUM</p> <p>Shows that young men have premarital sex less often with prostitutes and more with girlfriends</p> <p>The social context is one of AIDS, but also urbanization and lowered parental control (women having pre marital sex): link with the context only stated, and the macro features also just stated, but well done (lots of nuanced context and references).</p> <p>Good information, and some good methodological features (double coding, etc.) but big problems: very small sample, and very directed interviews</p>
Completeness of methodological description	1	1	0,5	0,5	1	1	5

8	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Adeokun et al. 2002 IPPF Dual protection intervention, Nigeria	24 focus 18 IDI with providers No rationale	No description of guide, nothing on format, interview conditions, interviewers	24 focus groups (4 with clients, 4 with providers, the rest in the community, different sex, age and socio-eco status to be representative) Nothing on IDI	nothing	Results: only quantitative evidence. Qualitative results brought up in the discussion, no quotes, explain the reasons for the quantitative results (some improvement but not that many) The qualitative data explains the relative lack success of the program (time constraints for providers, changing staff and poor training on the job, and most of all, men's reluctance to use condoms with their wives, even female condoms).	Yes, with elaborated quantitative data Provider and clients' views are contrasted	MEDIUM
Completeness of methodological description	1	0	0,5	0	0	1	2,5

9	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Eastman et al. 2005 PSRH</p> <p>Design of a parental sexual education program at work</p> <p>US</p>	<p>FG with parents (3) and adolescents (6)</p> <p>IDI with employers</p> <p>No rationale, but choice appropriate</p>	<p>Description of guide, by trained facilitators, matched by gender to the group</p> <p>Interview and FG conditions described</p> <p>Transcripts</p>	<p>In different work environments, diversity of employees seeked, groups stratified by gender, child's age and gender.</p> <p>On a voluntary base; this bias' implication on the results is discussed</p> <p>Adolescents chosen in one school with mixed socio economic and ethnic students, groups grade and gender specific</p> <p>Varied employers</p> <p>No rationale for numbers (but exploratory stage research)</p>	<p>Three researchers defined major topics by reading the transcripts; for each source of data, three researchers then found the text relevant to each topic (written on an index card), and classified them into themes within topics;</p>	<p>Lots of short quotes in the text (source not detailed, only adolescents, parent or employer)</p>	<p>cross-validation of categorization by discussion between researcher. Creation of a code book. Test of the coding by having a new researcher recode the transcripts with the code book: Coehn's kappas to grade interrater reliability.</p>	<p>GOOD</p>
<p>Completeness of methodological</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>1</p>	<p>6</p>

description							
-------------	--	--	--	--	--	--	--

10	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Katz et al. 2002 IFPP IUD use El Salvador	IDI with providers FG with clients	Description of guide FG animated by a team of two people, a moderator and a record keeper IDI and FG in Spanish, taped and transcribed into Spanish	From the list of public health facility with FP, random selection of 15 urban and 15 rural clinics; one provider selected randomly in each clinic FG: randomly selected 6 urban clinics and 4 rural clinics; groups by method used (sterilization, re-supply method, IUD) No rationale for numbers, but high	Reading of first transcripts, coding scheme designed, then coding of the text of each transcript with a software (DTSearch)	Short quotes, stand alone, many, no details on respondent (just provider or client) Explains well that the problem are a negative view of this method by women, and the fact that provider do not talk about this method and encourage use, because they do not feel confident in inserting / removing it (lack of experience), and also prescription writing implies less effort.	Quantitative data (mystery clients) compared to qualitative data Clients and providers perspectives compared	GOOD
Completeness of methodological description	1	1	1	1	1	1	6

11	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Hardee et al. 2004 IPPF</p> <p>Link between one-child policy and improved women's status</p> <p>China, rural</p>	<p>FG</p> <p>No rationale</p>	<p>Guide described (in results section!)</p> <p>Moderated by staff of official FP research center in China, met in various places, lasted 1 to 2 hours</p> <p>Not always possible to match moderator with the group (sex and age), which would have been preferable</p> <p>FG secretly observed by officials!</p> <p>Focus group discussion taped (supposedly), transcribed into Chinese, then translated to</p>	<p>Two rural counties each in three provinces, different levels of eco development and FP success, within each county, seven township randomly selected for FG, local leaders helped selected participants, which included a bias (favourable towards state FP policy), but not possible other ways; diverse opinions and some negative ones show that bias not total.</p> <p>Diverse groups of men and women, 5 to 12 participants per group, altogether 32 women and 232 men (! Mistake), no number of FG, but apparently high.</p>	<p>Reading the transcripts and identifying theme and representative quotes.</p> <p>In the results, we see that results are contrasted by generation, to get at social change, but not explained in the analytical method</p>	<p>Short quotes, many; respondents described</p>	<p>Quantitative data for contraception</p>	<p>MEDIUM</p> <p>Concludes from the data to entrenched gender inequalities (sex ratio, worse mortality for women, many inequalities), but to some improvements: decrease of son preference, more day in family decision, sometimes equal opportunities at work, possibility to earn money.</p> <p>Do not work on the link between improved status and FP, which is just stated. (do not use the links made by women themselves). Improved status may be a consequence of economic growth alone.</p>

		English					
Completeness of methodological description	0	1	1	0,5	0,5	0,5	3,5

12	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Bonvalet 2003 Population Why do people live near their extended family? France	IDI Bias of the narrative (real life is made of compromise) is discussed	Very short description: family and residential histories, life histories, but from the quotes, we see that the interviews were very open Nothing on interviewer and interview conditions No indication on format (but there must have been taped and transcribed; the quotes show very careful transcription)	99 in-depth re-interviews among national survey respondents, of which 37 lived near extended family used for this analysis No explanation of selection of respondents within survey respondents (presumably: random) Nothing on number, but reasonable given purpose	Construction of a typology of families living near extended family; not explained beyond that. Different types are explained in terms of social background; ownership of family business, making up for difficult childhood.	Use of case to illustrate each type, with lots of quotes, very long. Enough quotes to verify the analysis	With quantitative data and other research on the same topic	GOOD The information is very short, but sufficient to educated readers (life histories, typologies);
Completeness of methodological description	1	0,5	1	1	1	1	5,5

L1	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Agadjanian 2001 Population Studies qualitative + quantitative Mozambique Contraceptive Use	FG: expert interviews, individual interviews, and participant observation in specific events General purpose: understanding the mechanisms through which church attendance affect contraceptive use (specific aims are not specified and results from each method are presented together)	No information on who collected the data and how besides “research team” (one guess the authors did for the use of the personal pronouns “I” when describing the data collection “I asked...”) No information about the language or the transcriptions (one guesses Portuguese from the use of words in the quotes) (data recorded: not stated, but implied by the long quotes)	Affiliation to the main religious congregations of the Maputo suburb belt No information about the number of interviews or FG No reference to previous publications by the author in which the study would be better described	Nothing about methods explicitly The facto the author explores an hypothesis contrasting data from differnt religious communities and in differnt settings	A few quotes (not always specified whether from interviews of FG) are used to illustrate findings which are mainly summarized	Interpretation of qualitative and quantitative part of the study is not merged explicitly, but the conclusion interpret survey data using the findings of the qualitative study implicitly	GOOD No argument for quail study at all: It is just presented as an obvious component to understand the different interaction and communication going on within different congregations. The research shows that religiosity is relevant not per se’ but for it offers to women in particular opportunities for social interaction and therefore exposure to modern behaviour like contraceptive use

Completeness of methodological description	0.5	0,5	0	0	0,5	0.5	2.0
--	-----	-----	---	---	-----	-----	-----

L2	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Perelli – Harris 2005 PDR qualitative + quantitative Ukraine Fertility	FG and individual interviews: understanding the reasons for early childbearing among women No differentiation of the reasons to use one or the other method	No information on who collected the data, but presumably research assistants working in local language No information on language, transcription	Different regions of Ukraine and various age and education 22 FG (15 with women and 7 with men with 8-10 people) No information on the number of individual interviews Recruitment : Flyers distributed at metro stations	Nothing about methods of analyses The author search for themes explaining early childbearing with little differentiation by social group (she works	findings are summarized , no use of quotes	the conclusions interpret survey data using the findings of the qualitative study in a convincing way	GOOD The qualitative study is used to interpret puzzling demographic trends like early childbearing and low fertility. Conclusions show that early childbearing and low fertility are consistent with cultural beliefs about health, gender, as well as economic uncertainties and practices of intergenerational support
Completeness of methodological description	1	0	0.5 !	0	0.5	1	3

L3	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Schroder-Butterfill and Kreager</p> <p>PDR 2005</p> <p>Childlessness in old age and vulnerability of the elderly</p> <p>East Java Indonesia</p>	<p>12 month participant observation</p> <p>Semi-structured interviews with all elderly over 60 in one village (203)</p> <p>Aim: collect life histories of all elderly over 60 in the village and to compare ego and alter information in relation to pattern of support</p>	<p>Detailed information and justification for interviewing men and women separately and for having at least two rounds, multiple perspectives and triangulation with observation and daily contacts (sensitive topics like adoption, bad relationships etc..)</p> <p>Childlessness is a stigma and a reason to divorce.</p> <p>Children with no contact to parents were not mentioned in the first interview</p>	<p>Village was one of the three extensive rural communities in which field sites were conducted and in which levels of childlessness vary (choice on east Java because highest prevalence of elderly childless)</p> <p>203 IDI in Kidul, Second round IDI with 45% of the same sample and interviews with one related person of the elderly for 50% of the sample</p> <p>Census of the elderly of the village (9 missing justified)</p>	<p>It is not explicitly mentioned but it is clear from the way in which results are presented that summary tables guided the choice of case studies to be studied in depth</p> <p>The authors distinguish levels of support across social parity and number of marriages.</p>	<p>Individual case studies used as illustration</p>	<p>Observation complete and correct interview data (where stronger is the effect of social desirability and sensitive topics)</p> <p>A random sample of households has been given a questionnaire on intra household economy and exchanges and on elderly care and health so that a typology of economic differences locally relevant could be developed</p>	<p>GOOD</p> <p>A Great example of village study in which methods are combined at all stage: of the research design and data collection is illustrated and justified in a detailed fashion</p>
Completeness of methodological description	1	1	1	0.5	1	1	5.5

L4	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Johnson hanks 2003</p> <p>Population (english edition)</p> <p>qualitative + quantitative</p> <p>Cameroon</p> <p>Education and fertility</p>	<p>Ethnographic material</p> <p>General purpose: understanding the relation between education and childbearing preferences</p> <p>comparing norms on premarital sex concept of self and independence, individuality</p>	Data collected by author through three stage fieldwork (described in another paper referenced)	2 communities where maximum contrast of school participation, marriage delay and fertility	“standard anthropological practice” = “establish the generative principle of action from disparate cases and examples” main trend exploratory	A few quotes (not always specified whether from interviews of FG) are used to illustrate findings which are mainly summarized	Interpretation of qualitative and quantitative part of the study is not merged explicitly, but the conclusion interpret survey data using the findings of the qualitative study implicitly	<p>GOOD</p> <p>The account of the socially meaningful processes related to education and fertility is convincing and pertinent to the initial question</p> <p>The examination of the concepts of selfhood, individuality independence and marriage are examined in context and related to reproductive practices</p>
Completeness of methodological description	0.5	0,5	0.5	1	0,5	0.5	3.5

L5	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Ville et al 2005 qualitative + quantitative Population (english edition) France Identity and migration	Exploratory semi-structured interviews to capture forms of identities and themes composing them	This information is in a footnote and does not say anything about data collection	“Varied and relatively complex life trajectories” 22 interviews (11 men and 11 women aged 23 to 92)	Not mentioned main trends exploratory	Extensive quotes from the interviews	No	GOOD The specific aim here was to construct a questionnaire which would incorporate a valid appraisal of identity. This is well explained
Completeness of methodological description	1	0,5	0.5 (because I think that a word on how these variation is achieved woud not have harmed comprehension)	0	1	0	3

L7	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
<p>Finer et al 2005</p> <p>Perspectives on Sexual and Reproductive Health</p> <p>qualitative + quantitative</p> <p>US</p> <p>abortion</p>	<p>In depth interviews (does not mention why interviews and not FG)</p>	<p>38 in-depth interviews with women in four sites</p> <p>Content parallel to the one of the structured survey - interviews during medical visit – English language and cash compensation – in depth followed structured interviews</p>	<p>Hospital based sample selected for variation in access to and reimbursement for abortion – all abortion patients could participate</p> <p>Reflection on the descriptive of the sample and its age bias and reasons for it</p>	<p>Transcription, correction and top down coding – validity check by cross authors reliability of the coding (software aid N6)</p> <p>Problem: the authors mistakenly think they can get at the factors of abortions by asking women why they abort, instead of analyzing the situations which lead women to abort, and concluding on the reasons to abort from the latter analysis (either by focusing on main factors present in all situations, or by getting at certain factors by contrasting different groups of abortion situations)</p> <p>In fact, such an analysis would have needed to compare women who had an unplanned pregnancy and aborted and those who did not...</p>	<p>Summary by authors of what respondents said and short illustrative quotes</p>	<p>Constant comparison of coded interviews and results from survey data analysis</p>	<p>MEDIUM</p> <p>The qualitative is used for illustration purposes only – no major theory or exploration is carried out or cross validated with survey data</p> <p>Indeed, no real analysis since they look for the answer in the wrong place! POOR? (I think it can stay with medium given that even it is just illustrative it is relatively decently carried out. >But if you feel strong about POOR than change it)</p>

Completeness of methodological description	0	1	1	0.5	0,5	1	4
--	---	---	---	-----	-----	---	---

L8	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Donaldson 2002 Population Studies India Contraception and population Policy	Published and unpublished reports on India pop policy (target free system) Interviews with senior Indian and foreign officials and pop specialists	No	No	No Exploratory main trend with some care in distinguishing perspective of different actors	No a part some very short quotes,	Interviews and reports material are read in parallel	GOOD Contrast between the official view of the Indian government (policy is successful) and the variety of implementations across the country states. Health personnel perceive resistance to certain type contraception partially due to poor quality of services.
Completeness of methodological description	1	0	0	0	0,5	0.5	2.0

L9	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Luke and Watkins 2002 PDR Developing countries (5 cases : Ghana, Bangladesh, Malawi, Senegal, Jordan) Population Policy	Secondary data Interviews with national elites collected (Policy Project) Additional interviews collected independently in Malawi Range of reactions to Cairo	Project interviews of the Implementation of reproductive health policies and programs in LDC Critical appreciation of the normative consensus effect (USAID, funds and Cairo) 110 interviews	5 case studies (4 out of original 8 of PP) selected for regional variation in health services and in the Cairo program implementation Respondents selected by the Policy Projects staff (NGO, government and donors, academics, consultants)	Read and code transcriptions Systematic comparison of categories developed through thematic coding and analytical grids Great discussion of the biases (i.e. USAID funding)	Extensive quotes and summary	No (the paper is entirely based on the reported reactions of interviewees though critically commented by authors)	GOOD
Completeness of methodological description	1	1	1	1	1	0	5.0

L9	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Marcell et al 2003 USA Reproductive health Perspectives on Sexual and Reproductive Health	FG to gain better understanding of reproductive health of young males to develop interventions to promote responsible sexual behaviour and care services + Free listing activities Group interviews in preparation of the FG guideline were interviews with male students of health classes	Flyers/ lunch breaks Parental consent 15 dollars compensation	32 boys in 3 * 2 FG in two schools in S Francisco Variation in ethnic background Self selection of respondents (good discussion of limits of self selection)	Standard content analysis for FG Grid and Emphasis of response method	Quotes	Three methods in a sequence should have improved validity and reliability of the FG	GOOD Complex framework to appreciate reproductive health topics in respondents lives Concrete results in terms of policy for developing health care services for male adolescents
Completeness of methodological description	1	1	1	1	0.5	0.5	5

L10	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Nie 2005 PDR Shanghai Population policy (responses to one child policy)	(Semi-structured ?) Interviews	Interviews per mail and phone (more than once sometimes)	47 interviews (2/3 women) 20-79 years old Mixed income and employment situation but mostly college educated Snowballing from 4 relatives and family friends known for > 5 years	No mention main trends in the exploration of the reception of the policy	Quotes	No	POOR It is difficult to judge to which extent what has been presented as results are the results of social desirability, acceptance or internalization. The problem is here the limit of IDI as an observation tool, participant observation, or systematic repeat interviews may be more useful Discussion is poor
Completeness of methodological description	0.5	0.5	0.5	0	0.5	0	2

L11	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Buckley et al 2004 Studies in Family Planning Uzbekistan Reproductive and Sexual health	FG discussion to understand reasons for worsening sexual health despite health and reproductive programs	In Russian and Uzbek Cash honorarium Transcribed/double translation	Mothers with children <3 , unmarried women and single men 18-24 secondary educated Selected from housing blocks Does not say number of FG	Not said but it looks like main stream content analysis	quotes	DHS tables	MEDIUM Interpret DHS results FG are not suited to go beyond a description of dominant norms; the research question needed individual level behavioural observations to understand access or non access to sexual health programmes and its link with health conditions(I so not share this entirely: i believe that in this case they had a lower level ambition that is to explore attitudes towards health and reproductive programs so FG could be enough)
Completeness of methodological description	0.5	1	1	0	0.5	0.5	3.5

L12	Tools, and rationale for choosing these tools	Description of data collection conditions and interviewers, language, data format	Selection criteria Number of IDI / FG	Method of analysis	Evidence, demonstration	Cross-validation	Overall methodological quality
Mills and Bertrand 2005 Studies in Family Planning Ghana Health / obstretic care	FG because of “abundance of information about beliefs opinions and perceptions”	Very good description of the research setting Community contact for recruitment Details about FG sessions Local language	10 chiefdoms were the basis for 18 purposively groups 18 FG Groups described in table	Transcription and translation Initial coding with Atlas ti	Quotes Discussion of biases (social desirability)	Reaction to the dissemination seminar	GOOD The multiple perspectives from all type of actors involved in the health obstretic care is useful to identify reasons for behavioral patterns
Completeness of methodological description	0.5	1	1	1	0.5	0.5	4.5