



# Rationale and procedures for nesting semi-structured interviews in surveys or censuses

**Enid Schatz**

University of Missouri

*Demographers who use survey data and census data from health and demographic surveillance areas can gain substantially from expanding their repertoire of methods to make use of qualitative methods. Similarly, those who conduct and analyse data primarily from semi-structured interviews or focus groups can benefit from information provided by survey research. This paper presents a systematic mixed-methods model—data-linked nested studies—for sampling respondents for semi-structured interviews from survey or census lists. The paper outlines how to conduct these types of study, and their technical and analytical advantages. It highlights the benefits of building on a strong foundation, the ability to compare samples, and the expansion of the range of evidence for, or against, the validity of the substantive findings. Case studies from two data-linked nested projects—in Malawi and South Africa—are used to describe in detail the nested-study approach.*

**Keywords:** mixed methods; qualitative methods; survey; health and demographic surveillance systems; Africa; South Africa; Malawi; HIV/AIDS; ageing

[Submitted November 2009; Final version accepted May 2011]

For demographers who use surveys or censuses from health and demographic surveillance sites (HDSS), there are real advantages in nesting studies that use qualitative methods within those projects. Over the last 20 years the use of such mixed-methods research has grown rapidly (Creswell 2009). In much of the current deployment of the procedures, the collection of survey or census data serves as the core method and is supplemented by the use of qualitative data (Bryman 2006). A majority of these studies do not make use of the same study samples for both approaches, but in this paper I make a case for a mixed-methods model that I describe as ‘data-linked nesting’. It comprises the use of samples from survey or census populations for supplemental projects that make use of qualitative methods, such as semi-structured interviews or focus groups.

A number of papers and studies specify the benefits of mixed-methods approaches for demographic research (e.g., Sieber 1973; Caldwell 1985; Knodel 1997; Obermeyer 1997; Pearce 2002; Randall and Koppenhaver 2004; Axinn and Pearce 2007). Other studies actually demonstrate the value of using several methods in a particular analysis (e.g.,

Short et al. 2002; Watkins 2004; Seeley et al. 2008a; Stewart et al. 2008; Adams and Trinitapoli 2009). However, few texts provide those researchers who primarily collect and analyse survey and census data with concrete explanations of how they can, and why they should, enhance their projects by incorporating qualitative methods (Leahey 2007).

Critics of mixed-methods approaches often argue that they are flawed by the conflicting paradigms in play (for discussions of the ‘paradigm wars’ see Rossman and Wilson 1985; Sale and Brazil 2004; Schwandt 2006). However, a significant group of researchers have moved beyond this stage (Creswell 2009) and view mixed-methods research as a key research approach that is justified by its results (Johnson et al. 2007). The legitimacy it has acquired over the past decade is shown by the pressure to conduct it now exerted by funding agencies (Lamont and White n.d.; Small n.d.).

While the parallel collection of quantitative and qualitative data is not new (Stycos 1981; Knodel 1997; Bryman 2006), this growing area of social demography presents an opportunity to broaden the study and understanding of social and demographic phenomena (Hill 1997). Nested studies, like

other mixed-methods studies, provide opportunities to corroborate and elaborate findings and to initiate new investigations (Rossman and Wilson 1985). The technique of data-linked nesting described below goes beyond combining two parallel data sources on the same topic. The method entails selecting respondents for semi-structured interviews (or for other qualitative methods of data collection) from among survey/census respondents to allow the construction of simultaneous micro and macro perspectives on a topic (Elliott 2008).

The paper begins with an introduction to the methods used in two data-linked nested projects: one embedded in a panel of surveys in rural Malawi and the other in a HDSS in rural South Africa. Next, examples from the case studies are used to highlight the benefits of building on an existing research infrastructure and making use of survey/census data for sampling, and the particular analytic advantages of the availability of linked data sources used in combination. In the final sections, I discuss challenges of nested studies and draw conclusions.

### **Case studies**

The illustrative examples of data-linked nesting described here made direct use of existing survey/census lists and data to select households or individuals for an in-depth investigation of a particular topic. (Examples of other studies that use this approach in low-income countries include: Fricke 1994; Casterline et al. 1997; Pearce 2002; Adato et al. 2007; Adato 2008; Madhavan et al. 2008; Seeley et al. 2008b; Adams and Trinitapoli 2009). In each study outlined below, the survey/census preceded the collection of semi-structured interviews, thus providing a sampling frame and the infrastructure for the nested project. The first project paired the following two studies: (i) the Malawi Diffusion and Ideational Change Project (MDICP), a multi-wave study in rural Malawi surveying married women of reproductive age and their husbands about issues related to family planning and AIDS; (ii) the Gender Context Study, which focused on measurement issues related to women's status and gender dynamics (Schatz 2002, 2003, 2005). [See [http://www.malawi.pop.upenn.edu/Level%203/Malawi/level3\\_malawi\\_main.html](http://www.malawi.pop.upenn.edu/Level%203/Malawi/level3_malawi_main.html) for more information on this project, and to access other papers written using the MDICP data.] The second study linked census data from the Agincourt Health and Demographic Surveillance System (Agincourt) with the Gogo [Grandmother] Project

that explored older women's roles, responsibilities, and kin relationships in rural South Africa (Schatz 2007, 2009; Schatz and Ogunmefun 2007; Ogunmefun and Schatz 2009). [For more information about the Agincourt Health and Demographic Surveillance Site, nested projects, and papers written using Agincourt data, see <http://web.wits.ac.za/Academic/Health/PublicHealth/Agincourt/> and <http://www.agincourt.co.za/DataSection/index.htm>.]

### *MDICP/Gender Context Study*

In 2000, I conducted the Gender Context Study in one northern and one southern district of rural Malawi. The Gender Context Study re-interviewed 50 MDICP couples to investigate the following unexpected first-wave survey results: women in the matrilineal south appeared to be less autonomous than women in the patrilineal north. Women's autonomy was measured in the survey through a series of questions on freedom of movement, the acceptability of divorce, and family planning decision-making (see Table 1). During the 10 weeks of fieldwork, local interviewers conducted two semi-structured interviews with 91 respondents (50 women and 41 men)—half in the southern and half in the northern MDICP research sites.

Owing to the nested nature of the study, open-ended responses to questions about women's autonomy mirroring the survey questions could be compared directly with closed survey responses. By yielding additional data and comparisons that provided a more complete and complex understanding of the settings, gender dynamics, and family life (Schatz 2002, 2003, 2005), the nested study was able to account for the unexpected statistical results.

### *Agincourt/Gogo Project*

In 2004 I oversaw the Gogo [Grandmother] Project in several villages in the Agincourt field site, where an annual census has been administered in 21 villages since 1992. The Agincourt research site in the rural north-east of South Africa is overseen by the Rural Public Health and Health Transitions Research Unit, which is attached to the South African Medical Research Council and University of the Witwatersrand School of Public Health. The nested study focused on the roles, responsibilities, and relationships of 60 older women (aged 60 and over) sampled from the Agincourt census in an area where AIDS is endemic.

**Table 1** Comparison of Gender Context Study sub-sample and MDICP 1998 sample, Malawi 1998

	Gender Context Study sub-sample	MDCIP 1998 survey sample
<b>Demographic characteristics</b>		
Mean age	33.5	31.3
Mean number children ever born	5.0	4.2
Mean years schooling	4.6	4.1
<b>Marital status (%)</b>		
Married	84	86
Divorced/separated	12	11
Widowed	4	3
<b>Reproductive health variables (%)</b>		
Ever-used contraception (modern methods)	38	34
Unmet need (of those who want no more children)	71 (21)	73 (302)
Worry about HIV/AIDS (% very worried)	77	75
<b>Women's status variables (%)</b>		
<i>Women can divorce husband if:</i>		
He provides no financial support	13	27
He beats her	56	67
He is unfaithful	60	64
She suspects he has AIDS	19	22
He won't allow family planning (FP)	21	22
<i>Can go without permission:</i>		
Market	17	17
Health centre	23	19
<i>Family planning decision-making</i>		
If husband doesn't want FP, find way	63	58
Can get way if wants to space	52	55
Can get way if wants to stop	38	35
Would use secretly if had to	56	57
<i>N</i>	48	994

Source: Malawi Diffusion and Ideational Change Project Wave 1 Survey (1998).

The Gogo Project fieldwork took place over 4 months. Owing to the strong research infrastructure of the site, I was able use telephone and e-mail communication with the project manager to supervise the second half of the nested study from a distance. Three local interviewers each conducted three semi-structured interviews with 20 older women, fully translating and transcribing each interview. In this study, rather than responding to particular statistical findings, the goal was to uncover information not available in the census data about older women's perceptions of the impact of AIDS on their lives.

### The benefits of data-linked nesting as a mixed-methods procedure

As in many other mixed-methods procedures, nested studies benefit from existing data that provide a

profile of the study community (Hill 1997; Pearce 2002; Short et al. 2002; Adato et al. 2007). Information from the previous research can be used to delineate topics, groups, or individuals for more in-depth study (e.g., Fricke 1994; Adams and Trinitapoli 2009; Pearce and Deaton 2011). The embedded nature of data-linked nesting also affords technical advantages, which include the ability to build on existing research infrastructure and to generate a variety of types of sub-samples from survey/census lists and data. The analytical advantages include being able to examine data quality through substantive contradictions or corroboration in the findings drawn from the two samples; multiple data sources can provide explanations and uncover meanings. While often findings from both sources are 'true', triangulation might reveal that questions are understood differently, or different methods may lead respondents to answer in different ways. In data-linked nested projects, I move beyond simply

using two data sources on the same topic; rather, I select respondents for semi-structured interviews from among the survey/census respondents, positioning the two data sources in a tighter relationship. Examples from my case studies in Malawi and South Africa highlight the benefits of data-linked nesting.

### *The benefits of a strong foundation*

A data-linked nested design directly connects the collection and analysis of multiple data sources, thus benefiting studies designed to examine a given topic through both quantitative and qualitative data. The infrastructure provided by working in an existing study area can assist with entrée into the community, provide access to experienced field staff, and connect the researcher with an existing wealth of knowledge and data. While all of these are helpful in parallel designed mixed-methods projects, they are in many ways essential to nested designs.

*Community entrée.* By nesting the Gender Context Study (Malawi) and Gogo Project (South Africa) within existing study areas, I had access to valuable information as well as an introduction to the communities. In Malawi, the interviewers and I had previously worked on the MDICP survey and were thus familiar with the villages, and, importantly, with the headmen who had to give their permission before we could work within the communities. In South Africa, the association with Agincourt gave legitimacy to the project and facilitated entrée into the communities, and ensured that the households included in the study had not recently been part of other nested projects.

*Interviewers.* Once accepted into the community, the existing research infrastructure can further assist by providing access to trained or experienced interviewers with knowledge of the area and its surroundings, and sometimes respondents (Pearce 2002). For each study, I trained and supervised a small team of local interviewers who conducted repeated semi-structured interviews, thus increasing rapport between interviewer and subject. While some of the local interviewers had previously been trained as survey enumerators, they easily changed from closed to open-ended interviewing. However, it is often necessary for enumerators to undergo extensive re-training in how to conduct in-depth interviews. The Gogo Project interviewers lived within the larger Agincourt community and their knowl-

edge of the place and customs assisted in finding respondents and building rapport with them. Additionally, the interviewers in each project gained experience and skills that led to employment on subsequent surveys and other nested projects.

*Knowledge transfer.* Working in a site with an existing community of researchers provides a number of benefits. In Agincourt, detailed GIS village maps made it easier to locate respondents efficiently. At each site, formal and informal conversations with local researchers and staff informed the project. It was therefore possible to avoid unnecessary repetition, to discuss emerging results, and to compare findings with other research at the site. And, finally, I was able to access existing information on individuals and households in each sub-sample, thus reducing the need to ask standard demographic background questions, and lessening the burden on respondents.

*Strategic sampling.* By tapping into existing data, it was possible to strengthen the quality of the nested projects by drawing purposive samples that made use of census data or responses to actual survey questions. As others had done in their nested projects (Seeley et al. 1995; Mensch et al. 1999; Adato 2008), I drew on extensive information about individuals, couples, and households to select particular individuals or types of individual for each smaller, more in-depth nested study. Given the extensive lists available through an existing census/survey, it is possible, and perhaps tempting for 'quantitatively-oriented' researchers, to use simple random sampling or stratified random sampling. While this addresses some of the potential selection biases arising from convenience or quota sampling (sampling methods used for many studies that make use of qualitative methods), researchers should be cautious about using random samples for very small qualitative projects. Doing so might actually create a more biased sample than one selected using specific criteria to capture those with the best knowledge of the topic being studied (Adato 2008; Small 2009).

More advantageous than random sampling is the ability in data-linked nested studies to use previously collected census or survey data to select individuals with attributes beyond straightforward demographic characteristics like age and sex. In Malawi, using lists and responses from the MDICP 1998 survey to select respondents for the Gender Context Study, I purposively selected respondents for heterogeneity in age (under 25/between 25–35/35+), marriage



type (married/divorced/widowed, monogamous/polygynous), lineage (matrilineal/patrilineal), and residence pattern (matrilocal/patrilocal). This purposive sample provided a wide range of perspectives on women and men's roles in, and local customs related to, marriage and divorce, as well as gendered attitudes about family planning and HIV/AIDS. In addition, insights into these topics from various points of view clarified how gender as a cultural construct shaped local opinions about such issues as contraceptive use and worry about AIDS. The advantages of creating such a sample from a survey list are the ease and speed with which the nested sub-sample can be generated, and the ability to have a sufficient number of alternative respondents before entering the field. Rather than spending time finding individuals that meet the criteria, the survey data provide an available searchable list. Further uses of purposive sampling have allowed researchers to compare groups who answered particular survey questions differently (e.g., Casterline et al. 1997 on contraceptive use) and to compare those in intervention vs. control groups in a larger study (e.g., Adato 2008 on cash transfer programmes).

The access to information about all members of the target population also can counterbalance the tendency to overlook (or exaggerate) certain constituencies; researchers are able to target those with characteristics that are infrequent in the population, and avoid overstating the prevalence of anomalous cases (Sieber 1973). Thus, data-linked nested studies offer a way of looking beyond averages to outliers, as well as providing the means of situating outliers in relation to their setting (Pearce 2002).

An interest in how AIDS affects older South African women's lives steered the Gogo Project. However, data from Agincourt on household-level mortality and cause of death (Kahn et al. 2000; Kahn 2006) showed that between 2001 and 2003 only 3.1 per cent of households with an older woman had experienced an AIDS-related adult death, and another 11.3 per cent an adult death from another cause. To capture a sufficient number of 'affected' cases, and to maximize the opportunity to compare across sub-sample households, the Agincourt Unit agreed to provide a non-proportional stratified random sample—one that did not represent the actual proportion of households affected by an adult death (AIDS-related or other cause) in the population. The sub-sample was divided into three equal strata of 20 households, 60 in total, each with a woman over age 60. The three strata comprised: (i) households with an adult AIDS-related death in the 2001–03 period; (ii) households with a

non-AIDS-related adult death in this period; and (iii) households with no death during this period. The analytical advantages of this sub-sample will be discussed below.

Other nested studies have shown the benefits of targeting outliers or difficult-to-find populations by making use of survey or other existing data. For example, Pearce (2002) used this approach to focus on anomalous cases that did not conform to the general statistical relationship between religion and fertility preferences in Nepal. Bledsoe (1994) pursued the stories of just a few women in Gambia who were using contraception after a miscarriage, an unexpected practice in a high-fertility area, to focus on and explain the phenomenon. Seeley et al. (1995, 2008a, b) used knowledge about household structure and composition from a previous survey in Uganda to oversample the under-represented female-headed households in a study about poverty, ageing, and AIDS.

The Malawi Religion Project (adams and Trinitapoli 2009) and the National Study of Youth and Religion in the United States (Pearce and Deaton 2011) used another strategic sampling procedure. Each project included a large number of semi-structured interviews with a quota or a stratified random sub-sample from the even larger survey. While not intended to be representative in their analysis, they provided sufficient variance across cases to make it possible to address numerous research questions that emerged in the course of the project. When a project has a specific focus on a 'class of individuals', strategic sampling as described above for the Gogo Project is beneficial. However, interviews with a large number and general swathe of participants permits a variety of (unplanned) sub-analyses to be conducted at a later date.

#### *Analytical advantages of nested analysis*

In addition to the technical advantages outlined above, data-linked nesting allows the integration of data sources to provide both subtler and more representative portraits of the phenomena and individuals under study than is possible with a single method, or even parallel data collection. Rossman and Wilson (1985) outline three ways that mixed-methods research might improve conclusions. First, results from one source can be used to *corroborate* the results obtained from another. Secondly, results from one data source can be used to *elaborate* the findings of another, often strengthening the overall conclusions. Thirdly, conflicting results from

different sources can *initiate* further investigations to uncover new interpretations. Below, I discuss how a data-linked nested study—with respondents drawn from and traceable back to a survey or census—can exploit these three analytical advantages.

*Corroboration.* One of the main aims of mixed-methods research is to contribute to inductive inference, the ‘process of creating meaningful and consistent explanations, meanings, conceptual frameworks, and/or theories’ (Tashakkori and Teddlie 2003, p.709). In nested studies this entails examining interview data alongside existing (or subsequently collected) survey/census data. One could argue that inference is even more robust in nested studies because these allow the direct comparison of results from nested sub-samples with results from the population from which the sub-samples were drawn—to assess similarities and differences. Nested studies provide an opportunity to assess whether sub-samples have distributions similar to those of the general (survey/census) population, not just on demographic characteristics, but also on pertinent variables. Some researchers may find analyses of qualitative data more convincing if they are able to see how similar or different members of a sub-sample are from others in the same strata in the larger population.

Although the sub-sample selected for the Gender Context Study in Malawi was not intended to be representative, Table 1 shows that women in the sub-sample are similar to women in the larger MDICP sample in demographic characteristics and responses on the MDICP 1998 survey. [For a comparable table for men in the sub-sample and MDICP sample, see Schatz 2002.] The nested-study respondents were on average slightly older, more educated, and had given birth to approximately one more child than women in the larger survey sample. Table 1 also presents respondents’ aggregated survey responses to vari-

ables related to contraceptive use, worry about AIDS, and questions on women’s autonomy. These were all issues central to both the nested study and the survey. Again, the two samples differ only slightly, even when the question refers to rather complex issues, such as respondents’ worry about future HIV infection. Also similar are the responses to questions meant to address women’s autonomy, as shown in the last section of Table 1.

The similarities between the two samples suggest that it is reasonable to generalize from the narratives of the Gender Context Study to the MDICP population. In some respects—for example, on the subject of divorce on which the sub-sample’s survey responses suggest less autonomy for women than do those of the larger survey population—the comparison allows for an even stronger conclusion about women’s autonomy. In the semi-structured interviews, women provide detailed reasons why women can leave their spouses (Schatz 2002). Since these women’s responses suggested that they had less autonomy than those who gave corresponding responses to the fixed-choice survey questions, one can assume that women whose responses in the survey suggested that they had *more* autonomy would have had these reasons and more. Thus, the narratives related to divorce of the sub-sample respondents probably applied to the majority of individuals, even those whose responses to the survey suggested that they had less autonomy.

Table 2 shows census statistics from 2003 for households covered by Agincourt and the Gogo Project for issues related to household vulnerability: household structure (e.g., mean size and mean number of children) and household composition (e.g., households with at least one child, at least one fostered child, or at least one orphan). Given the disproportionate number of households in the Gogo Project sub-sample with a recent death, it is

**Table 2** Comparison of Agincourt and Gogo Project households, South Africa 2003

sub-sample	Agincourt households with 60+ woman	Gogo Project
Mean household size	7.0 (1–40)	8.6 (1–24)
Mean number of children under 15 in household	2.3 (0–18)	3.6 (0–13)
Households with at least one child under 15 (%)	76.2	86.7
Households with at least one fostered child	26.5	33.3
Households with at least one maternal orphan	8.3	20.0
Households with at least one AIDS-related adult death (2001–03)	3.1	33.3
Households with at least one non-AIDS-related adult death (2001–03)	11.3	33.3
Total <i>N</i>	2,671	60

Source: Agincourt Health and Demographic Surveillance System 2003 Census Data.

not surprising that these households appear more vulnerable than the population at large. The mean household size and mean number of children in the household are larger, and a greater percentage of the sub-sample households include at least one fostered or orphaned child aged under 15. In the 2003 census, households with a death are on average larger, have more children, and are more likely to have a child in the household (not shown in Table 2). Moreover, a greater percentage of households where an AIDS-related death occurred than in all other household types, in both the sub-sample and the larger population, had at least one maternal orphan (not shown in Table 2).

While the sampling strategy of the Gogo Project places limits on the extent to which inferences can be drawn about the larger census population, it also provides compensating analytical advantages (Pearce 2002; Small 2009). If the Gogo Project's sampling strategy had been to randomly select Agincourt households with a woman over the age of 60, the sampled households with a death, especially an AIDS-related death, would probably have been very few in number. The effect would have been to undermine an important conclusion of the Gogo Project: that differences between the strata in terms of both experience with caregiving and use of pensions were smaller than expected from the literature (Schatz and Ogunmefun 2007). The non-proportional sampling used permitted comparisons across strata that confirmed that there were few differences across these strata of households in older women's pension usage, caregiving responsibilities, and well-being.

Other studies have used various forms of nesting to assess similarities and differences between the population and sub-samples. For example, Messersmith et al. (2000) show the value of this type of comparison in their paper on condom use in Nigeria. The authors contrast demographic characteristics, responses to questions about sexual activity and sexually transmitted infections (STI), as well as number of lifetime sexual partners between their population-based survey, sub-samples of high- and low-risk respondents, and a sample of sex-worker clients (not survey respondents). While neither sub-sample was representative of the survey population, the comparison provided information about the differing likelihood of having reported having an STI, which cautioned against generalization and qualified the interpretation of results from each sample. Casterline et al. (1997) explored contraceptive use in the Philippines through semi-structured interviews with a nested sub-sample of survey

participants, both to corroborate survey findings and to help explain them. The interview data showed that individuals' discordant survey responses about fertility preferences and contraceptive practice were not a result of poor survey measurement of either concept. Instead, findings from the sub-sample suggested that the discrepancies in survey responses were largely determined by the strength of wives' and husbands' fertility preferences, and fear of side effects from the use of family planning methods.

*Elaboration.* A particularly valuable advantage of data-linked nesting is the opportunity it provides for improving data quality by using interview data to elaborate on concepts captured in both studies (Adato 2008). For example, respondents in the nested studies can be asked a question that was asked in the survey, followed by questions on the way they heard and understood the question—as in a cognitive interview (Willis 2005). This procedure may lead to a conclusion that the numerical responses were quite biased, or it may simply help shape the interpretation of the variable for analysis. Some differences by mode of interview are to be expected. Survey respondents are required to give short pre-coded answers, whereas in-depth interview respondents are permitted to talk at length. Moreover, since the interviews are almost always recorded and the coding done later, analysis can be based on the respondents' own language and perceptions of the topic.

In Malawi, I was able to use the Gender Context Study to elaborate on how context influenced survey responses related to women's autonomy (Schatz 2003). One clear example was that the freedom-of-movement questions were understood and answered differently in each region in which the survey was fielded owing to differences in the location of the local market and health centre in relation to the villages. In the southern (matrilineal) research area, the market and health centre were on a main road outside the study villages, whereas in the northern (patrilineal) area, they were within the study villages. Thus, unsurprisingly, women in the south had less freedom of movement, and were more likely to need their husband's permission to go to these places (Schatz 2003). The dangers of a main road—the possibility of meeting strangers and the difficulty of monitoring women on the main road, not the danger of the market or health centre itself—were captured by these questions in the south. Understanding this altered the interpretation of the regional differences in aggregate survey results for this variable.

Elaboration took a different form in the South African study. There were no data from Agincourt on caregiving, but the AIDS literature suggested that older women in households with an AIDS-related death would have increased burdens of caregiving for adult children and later for orphaned grandchildren. Contrary to these expectations, however, the nested study found few differences across these different types of household (e.g., those with and without a recent AIDS-related death). Instead, one of the primary findings of the study was the interconnectivity of households: nearly all households and individuals in the study area were affected by AIDS through their kin networks, even when a recent AIDS-related death had not occurred within a household (Schatz 2007; Schatz and Ogunmefun 2007). Respondents both in households with and without a recent AIDS-related death spoke of caring for those who were sick or for children orphaned by AIDS within the respondents' homes as well as in the homes of their kin. Older women also contributed financially to health care and funeral costs beyond their own households. It was only by having sufficient households in each category, made possible by the available census data, that it was practical to question the validity of widely accepted interpretations of the impact of AIDS-related deaths on older women and their households.

Substantial contradictions between findings from different sources can lead to important evaluations of data quality, as well as illuminating ways in which data should (or should not be) interpreted. For example, Poulin (2010) compared survey data with qualitative data collected shortly afterwards in Malawi, and found that a significant proportion of adolescent respondents (12 per cent of males and 30 per cent of females) who had denied ever having had sexual intercourse in the survey, subsequently revealed a sexual relationship in an in-depth interview. Such results clearly suggest caution in the use of the measure used in the survey. Similarly, in a study in China, Short et al. (2002) asked the same question—how many hours a woman spent on childcare in the previous week—in a survey, in an in-depth interview, and in a focus group. They found that interview and focus-group respondents reported estimated hours in a *typical* week, rather than childcare hours in the *previous* week. This finding cautioned the researchers in making strong claims about time spent in childcare using the survey data, since an unknown number of respondents had not interpreted the question in the way intended by the survey designers. The interview and focus-group data were also valuable in illuminating what women 'counted'

and did not count as childcare, which informed the use of the 'childcare variable' in regression analyses, as well as assumptions about causal ordering in the relationship between childcare and the participation of women in the labour force.

*Initiation.* The potential for making use of contradictory results to initiate further investigation is particularly clear when nesting a study that entails the collection of longitudinal data. The nested study allows for a more iterative process between the stages of work and the different methods of data collection (Adato 2008). Survey and census data can be used to set up the nested project in the ways outlined above, and a nested project using qualitative methods can help inform future survey instruments, as well as point researchers to new topics for examination at the population level.

Studies implementing qualitative methods can help clarify what individuals might have been thinking about when they answered a particular closed-ended survey question, and lead to reframing the topic or rewriting the question(s) in subsequent surveys. The MDICP survey asked women if they could leave their husbands if he 'did not support his family financially' and respondents in the Gender Context Study were asked about acceptable reasons for divorce in a way that allowed for explanation *and* qualification of their responses. Only about a quarter of the women in the MDICP sample and 13 per cent in the sub-sample reported in the 1998 survey that it was permissible to leave a man who does not support his family financially. Yet in the Gender Context Study, nearly all respondents spontaneously mentioned this reason for divorce during the in-depth interviews. In these interviews, however, women generally reported that *not* supporting one's family financially was not a divorceable offence, but squandering the family's money was. Men who actively sought an income but found none were excused. However, if a woman thought her husband was wasting money on beer and girlfriends, she deemed this behaviour inappropriate and said the man deserved to be divorced. Thus, the fixed-choice responses on the survey hid important subtleties about what 'supporting one's family' meant to respondents and what particular circumstances individuals were considering when they thought about what might lead to divorce. Technically, a man would not be supporting his family financially in either of these cases, but women saw one scenario as excusable, while the other was not. This distinction appears to have made a significant difference in



the way women heard and answered the question. Had the research team tested these questions using cognitive interviews or another form of semi-structured interviewing before conducting the survey, more specific survey questions could have been developed. For example, the questions might have asked not only whether it was permissible to leave one's wife, but under what circumstances. In most mixed-methods approaches, qualitative data can improve the quality of survey instruments by highlighting what particulars need to be included to make a question meaningful and precise in a particular context. Here, asking a sub-sample of respondents the questions asked in the survey provided further insight into why a certain survey question needed to be asked differently.

In the project in South Africa, an important result of the research was that it became apparent that many older women who were caring for young children were eligible for, but were not able to access child grants (a welfare grant provided by the South African government). Other studies conducted in the study area around this time found that many children cared for by their own mothers also did not have the necessary documentation (a birth certificate) to access these grants. As a result of these findings, the Agincourt unit included a Social Grants module as part of the 2005 census to investigate the number of eligible households not accessing the grants. This information, shared with local government, led to the provision by the government of a mobile unit to help caregivers—grandmothers, mothers, and others—of children aged 14 or younger to obtain a birth certificate and submit an application for a child grant (Twine et al. 2007). In this case, the iterative process led to further data collection and then a concrete positive outcome for households similar to those from the Gogo Project.

Other nested projects in South Africa, such as the work by Adato and colleagues and that of Townsend, Madhavan, and colleagues offer additional important examples of the initiation of new ventures. In a qualitative follow-up study to a large household survey, Adato et al. (2007) found that a large number of individuals who reported being 'unemployed' in a survey actually were engaged in economic activities. By uncovering the nature of work activities missed by household surveys, this study helped clarify how future labour force surveys could capture a wider range of work and employment. Another example of a project that identified new areas for investigation is provided by the nested mixed-methods study undertaken by Townsend, Madhavan, and colleagues

on the role of non-custodial fathers in the lives of their children in rural South Africa (Townsend et al. 2006; Madhavan et al. 2008). By highlighting continued and significant relationships between non-custodial fathers and their children, their ethnographic study revealed the need for future surveys to measure both inter-household and intra-household networks in order to better understand the determinants of children's well-being.

### **Problems posed by data-linked nested studies**

Keeping the quality of all data collection high is essential to creating a data-linked nested project that is more than the sum of its parts. The merits of decisions made about the nested project that are based on existing survey/census data depend on the quality of those data. For example, if survey questions produce misleading responses or if there are biases in the sample selection, these defects will also affect the selection of individuals or the analysis of data, or both, for the nested sub-sample.

Using survey or census lists to generate random samples is not the best way to capture under-represented or difficult-to-define populations; in such cases, purposive samples are more appropriate. For example, the use of existing survey/census data might not be the best way to target respondents for projects focused on sensitive issues that individuals are likely to over-report (e.g., income) or under-report (e.g., number of sexual partners), or to identify hidden or difficult-to-define populations. In addition, if a simple or stratified random sampling technique is used, it is less likely that under-represented groups will be captured in a sub-sample.

Decisions made about the size of the nested sample have important ramifications. While statistical power guides decisions about the size of the survey sample, one common approach with individual interviews is to continue adding respondents to the sample until no new relevant information is likely to be provided by an additional respondent (Esterberg 2001). When nested projects are to be funded from the same pool of resources as the survey, the inability to predict the necessary size of the nested study may appear to 'quantitatively-oriented' observers to show indecisiveness rather than strategic wisdom. Alternatively, a decision to specify a set number of interview respondents before the start of fieldwork may seem short-sighted to a 'qualitatively-oriented' observer.

While there are clear advantages to selecting a sub-sample of a larger sample, ethical considerations

arise when returning to re-interview respondents to a previous study and making use of information drawn from existing data (Leahey 2007). During the consent process, most study respondents agree to allow researchers to use the information they will obtain to conduct analyses, but rarely are they asked for permission to return for another interview. To protect themselves, researchers who wish to conduct any type of follow-up study should (and this will be required by most review boards) include a question in the initial interview asking if the respondent would be willing to be invited to participate in future research. Because each of my studies described above were embedded in longitudinal research that was set up to collect data at multiple points in time, adding in the semi-structured interviews did not present this issue. As I became involved in each study area in collaboration with the principal investigators, I was given access to survey and census lists and samples that otherwise might have been more difficult to access (Leahey 2007).

Returning to households may lead respondents to ask why and how they were selected into a study. If sampling on a sensitive dependent variable—such as having a recent AIDS-related death in the household—it might be possible for the researcher to request an abridged sub-sample list, for example one that specifies only whether or not a death occurred, but not the cause. After completing fieldwork, the researcher can request a more detailed sub-sample list for use in analysis. This allows the research team to explain to respondents that their selection was based on a known characteristic (that a death occurred in the household), but not one that would stigmatize the individual or household.

Another consideration is respondent fatigue. While there are substantial benefits to conducting a nested study shortly after the survey/census, the time elapsed should be sufficient for households not to feel overburdened with requests to participate in studies. In Agincourt, the central administration regulates sample selection to ensure that households that have recently been included in intensive investigations (surveys, in-depth interviews, etc.) are excluded from the list of potential respondents before a sub-sample is selected. A short time lapse to follow-up can minimize changes in individual or household circumstances between data collections, but it adds to pressure on the limited time available to clean, prepare, and analyse survey/census data, and to produce the sub-sample. If the survey/census is used as a tool to locate particular individuals for the nested study, it may be necessary to return to the

field quickly to capture this population. However, if additional funding is needed to implement the second phase of data collection, the process of obtaining it may make it impractical to return quickly to the field.

Analysing large amounts of data of very different types can be a complex task. Defining in advance potentially illuminating comparisons can help make the most of mixed-methods analyses, as can the use of modern software packages to facilitate the analysis of data from nested studies. Bledsoe et al. (1999) provide an early example of how software, in their case Epi Info, can be used to maximize the integration of qualitative and quantitative data. A number of commercial and open-source software packages are available, including Nvivo, Atlas.ti, and Dedoose ([www.dedoose.com](http://www.dedoose.com)), a program designed for mixed-methods analysis. Survey responses or other supplemental demographic or census data for sub-sample respondents often can be imported and then directly compared with coded narrative data. Coded interview text then can be sorted by the response given on the survey or by designated categories from the census data. While the sample/sub-sample comparison may permit making inferences from small nested samples, generalizing the ideas of a single or small group of individuals from in-depth interview data—no matter how persuasive the story—to others with the same observable characteristics in the larger population remains problematic.

In nested studies there is a greater need than usual to be aware of potential confidentiality breaches because of the risks that arise from the combination of data sources (Leahey 2007). Ethnographers often protect respondents' individual identities by using a pseudonym for their geographic location. In contrast, survey research often discloses geographic locations, but protects individuals by presenting only aggregate analyses. Nested studies often present both geographic locations (from knowledge of where the survey/census took place) and disaggregated information. In data-linked projects, the risk of deductive disclosure may be even greater. Because of these risks, researchers need to be vigilant in not divulging information that would allow locals or outsiders to pinpoint the source—individual or household—of certain information.

## Conclusions

This paper has presented a model and case studies of a data-linked nested procedure that is an

improvement on the more common separate, parallel mixed-methods design. The procedure draws on different sources of data to corroborate and elaborate findings and to initiate new avenues of research (Rossman and Wilson 1985). It reflects a growing view of mixed methods as a unique and valuable approach to research, providing researchers with added value from each qualitative and quantitative method used (Johnson et al. 2007; Creswell 2009). While demographers have a long history of bringing different types of data together (Stycos 1981; Caldwell 1985; Knodel 1997; Pearce 2002; Axinn and Pearce 2007), and the methods of anthropological demography have increasingly been used with survey and census methods (Fricke 1994; Hill 1997; Bledsoe et al. 1999), the use of nested sub-samples takes an important next step by linking data from surveys/censuses directly with that from semi-structured interviews.

The studies described above demonstrate improvements in method by using existing data to generate sub-samples for semi-structured interviews: a purposive sub-sample to maximize variation in marriage and age in Malawi, and a non-proportional random sample, which oversampled households with an AIDS-related death in South Africa. Analytical benefits emerge in direct comparisons of characteristics of the survey/census population with its sub-sample and in comparisons of substantive findings; these assessments assess similarities and differences across data sources in each study. In Malawi, the respondents in the two samples had similar demographic characteristics and attitudes about women's autonomy, but the interview data enriched the interpretation of responses to the freedom-of-movement survey questions, and initiated ideas about how to reframe future survey questions about divorce attitudes. The South African nested study highlighted the importance of within-community comparisons provided by non-representative sub-samples, emphasized the unanticipated interconnectivity of older women's households in the HIV/AIDS era, and suggested the need for the initiation of a welfare-grants intervention. Each study provided the opportunity to link the specific words of individuals with previously collected and aggregated survey/census responses. The studies also revealed the challenges of this approach, including new ethical and confidentiality considerations, and the problems of bringing into communication large amounts of data from a variety of sources.

The studies, supported by other similar work, show that data-linked nesting yields developments

in method and analysis that can add to the value of different types of data collected for one investigation. Demographers and population scientists interested in social phenomena—particularly those who can draw sub-samples from survey respondents or populations that are the subject of a socio-demographic surveillance census—will gain insight into individual cases from a combined micro and macro perspective.

## Notes

- 1 Enid Schatz is at the University of Missouri, 420 Lewis Hall, Columbia, MO 65211, USA and University of Colorado-Boulder, and MRC/Wits Rural Public Health and Health Transition Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. E-mail: schatzej@health.missouri.edu
- 2 This work would not have been possible without the support and willingness of two research organizations to allow me to work at their study sites, namely, the Malawi Diffusion and Ideational Change Project and the MRC/Wits Rural Health and Health Transitions Research Unit (Agincourt). The principal investigators of these sites, and the staff who work there, assisted me in many ways and on many points. This paper began with the presentation of an idea at the Wit-Brown-Colorado-APHRC Colloquium at the University of Colorado in spring 2009, and early versions of the paper were presented at meetings of the American Sociological Association, summer 2009, and the International Union for the Scientific Study of Population in autumn 2009. The panels, discussants, and audiences at these meetings helped advance my arguments and thinking about data-linked nesting. Helpful suggestions for improving the paper were made by Nicole Angotti, Joan Hermsen, Rebecca Dingo, Sangeetha Madhavan, Jane Menken, Susan Short, Susan Watkins, Jill Williams, and anonymous reviewers for this journal. This work was supported by an NICHD training grant (T32 HD-007242-18) to the Population Studies Center at the University of Pennsylvania, and by the following grants: grants by the Andrew W. Mellon Foundation to the University of Pennsylvania and the University of Colorado; a grant by the Wellcome Trust [085477/Z/08/Z] that provided core funding for Agincourt; and a grant by The William and Flora Hewlett Foundation (2005-6516 and 2006-8377) to the University of Colorado. The work was also supported by the University of Colorado Population Center and the University of Missouri Research Board.

## References

- adams, jimi and Jenny Trinitapoli. 2009. The Malawi Religion Project: data collection and selected analysis, *Demographic Research* 21(10): 255–288.
- Adato, Michelle, Francie Lund, and Phakama Mhlongo. 2007. Methodological innovations in research on the dynamics of poverty: a longitudinal study in Kwa Zulu-Natal, South Africa, *World Development* 35(2): 247–263.
- Adato, Mchelle. 2008. Combining survey and ethnographic methods to improve evaluation of conditional cash transfer, *International Journal of Multiple Research Approaches* 2(2): 222–236.
- Axinn, William and Lisa Pearce. 2007. *Mixed Methods Data Collection Strategies*. Cambridge: Cambridge University Press.
- Bledsoe, Caroline. 1994. *Contingent Lives: Fertility, Time and Aging in West Africa*. Chicago: University of Chicago Press.
- Bledsoe, Caroline, Jane Guyer, Kuate Defo Barthelemy, and Shobha Shagle. 1999. Anthropological vision and research logic: an Epi Info primer, Program of African Studies Working Paper No. 7, Northwestern University.
- Bryman, Alan. 2006. Integrating quantitative and qualitative research: how is it done, *Qualitative Research* 6(1): 97–113.
- Caldwell, John C. 1985. Strengths and limitations of the survey approach for measuring and understanding fertility change: alternative possibilities, in John G. Cleland and John Hobcraft (eds.), *Reproductive Change in Developing Countries. Insights from the World Fertility Survey*. Oxford: Oxford University Press, pp. 45–63.
- Casterline, John B., Aurora E. Perez, and Ann E. Biddlecom. 1997. Factors underlying unmet need for family planning in the Philippines, *Studies in Family Planning* 28(3): 173–191.
- Creswell, John W. 2009. Editorial: mapping the field of mixed methods research, *Journal of Mixed Methods Research* 3(4): 95–108.
- Elliott, Jane. 2008. The narrative potential of the British Birth Cohort Studies, *Qualitative Research* 8(3): 411–421.
- Esterberg, Kristin G. 2001. *Qualitative Methods in Social Research*, 1st edn. Boston: McGraw-Hill Humanities/Social Sciences/Languages.
- Fricke, Tom E. 1994. *Himalayan Households: Tamang Demography and Domestic Processes*. New York: Columbia University Press.
- Hill, Allan G. 1997. Truth lies in the eye of the beholder': the nature of evidence in demography and anthropology, in David I. Kertzer and Thomas Earl Fricke (eds.), *Anthropological Demography: Toward a New Synthesis*. Chicago: University of Chicago Press, pp. 223–247.
- Johnson, R. Burke, Anthony J. Onwuegbuzie, and Lisa A. Turner. 2007. Toward a definition of mixed methods research, *Journal of Mixed Methods Research* 1(2): 112–133.
- Kahn, Kathleen, Stephen Tollman, Michel Garenne, and John Gear. 2000. Validation and application of verbal autopsies in a rural area of South Africa, *Tropical Medicine and International Health* 5(11): 824–831.
- Kahn, Kathleen. 2006. Dying to make a fresh start: mortality and health transition in a new South Africa, Umeå University Medical Dissertations, Epidemiology and Public Health Sciences, Department of Public Health and Clinical Medicine, Umeå University.
- Knodel, John. 1997. A case for nonanthropological qualitative methods for demographers, *Population and Development Review* 23(4): 847–853.
- Lamont, Michèle and Patricia White. n.d. *Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. Washington, DC: National Science Foundation.
- Leahey, Erin. 2007. Convergence and confidentiality? Limits to the implementation of mixed methodology, *Social Science Research* 36(1): 149–158.
- Madhavan, Sangeetha, Nicholas Townsend, and Anita Garey. 2008. 'Absent breadwinners': father-child connections and paternal support in rural South Africa, *Journal of Southern African Studies* 34(3): 647–663.
- Mensch, Barbara S., Daniel Bagah, Wesley H. Clark, and Fred Binka. 1999. The changing nature of adolescence in the Kassena-Nankana District of northern Ghana, *Studies in Family Planning* 30(2): 95–111.
- Messersmith, Lisa J., Thomas T. Kane, Adetanwa I. Odebiyi, and Alfred A. Adewuyi. 2000. Who's at risk? Men's STD experience and condom use in southwest Nigeria, *Studies in Family Planning* 31(3): 203–216.
- Obermeyer, Carla Makhlouf. 1997. Qualitative methods: a key to a better understanding of demographic behavior? *Population and Development Review* 23(4): 813–818.
- Ogunmefun, Catherine and Enid J. Schatz. 2009. Caregivers' sacrifices: the opportunity costs of adult morbidity and mortality for female pensioners in rural South Africa, *Development Southern Africa* 26(1): 95–109.
- Pearce, Lisa D. 2002. Integrating survey and ethnographic methods for systematic anomalous case analysis, *Sociological Methodology* 32(1): 103–132.
- Pearce, Lisa and Melinda Deaton. 2011. *A Faith of their Own: Stability and Change in the Religiosity of America's Adolescents*. Oxford: Oxford University Press.
- Poulin, Michelle. 2010. Reporting on first sexual experience: survey interviews, in-depth interviews, and



- interviewer–respondent interaction, *Demographic Research* 22(11): 237–288.
- Randall, Sara and Todd Koppenhaver. 2004. Qualitative data in demography: the sound of silence and other problems, *Demographic Research* 11(3): 57–96.
- Rossmann, Gretchen B. and Bruce L. Wilson. 1985. Numbers and words: combining quantitative and qualitative methods in a single large-scale evaluation study, *Evaluation Research* 9(5): 627–643.
- Sale, Joanna E. M. and Kevin Brazil. 2004. A strategy to identify critical appraisal criteria for primary mixed-methods studies, *Quality & Quantity* 38(4): 351–365.
- Schatz, Enid J. 2002. Numbers and narratives: making sense of gender and context in rural Malawi, PhD dissertation, University of Pennsylvania, Graduate Group in Demography and Department of Sociology.
- Schatz, Enid J. 2003. Comparing, contextualizing, and conceptualizing: enhancing quantitative data on women's situation in rural Africa, *Demographic Research* S1(5): 143–174.
- Schatz, Enid J. 2005. 'Take your mat and go!': rural Malawian women's strategies in the HIV/AIDS era, *Culture, Health & Sexuality* 7(5): 479–492.
- Schatz, Enid. 2007. 'Taking care of my own blood': older women's relationships to their households in rural South Africa, *Scandinavian Journal of Public Health* 35(Suppl. 69): 147–154.
- Schatz, Enid and Catherine Ogunmefun. 2007. Caring and contributing: the role of older women in multi-generational households in the HIV/AIDS era, *World Development* 35(8): 1390–1403.
- Schatz, Enid J. 2009. Reframing vulnerability: Mozambican refugees' access to state-funded pensions in rural South Africa, *Journal of Cross-Cultural Gerontology* 24(3): 241–258.
- Schwandt, Thomas. 2006. Opposition redirected, *International Journal of Qualitative Studies in Education* 19(6): 803–810.
- Seeley, Janet A., Ellen B. Kajura, and Daan W. Mulder. 1995. Methods used to study household coping strategies in rural South West Uganda, *Health Policy Planning* 10(1): 79–88.
- Seeley, Janet, Samuel Biraro, Leigh Anne Shafer, Pamela Nasirumbi, Susan Foster, Jimmy Whitworth, and Heiner Grosskurth. 2008a. Using in-depth qualitative data to enhance our understanding of quantitative results regarding the impact of HIV and AIDs on households in rural Uganda, *Social Science & Medicine* 67(9): 1434–1446.
- Janet, Seeley, Brent Wolff, Elizabeth Kabunga, Grace Tumwekwase, and Heiner Grosskurth. 2008b. 'This is where we buried our sons': people of advanced age coping with the impact of the AIDS epidemic in a resource-poor setting in rural Uganda, *Ageing & Society* 29(1): 115–134.
- Short, Susan E., Feinian Chen, Barbara Entwisle, and Fengying Zhai. 2002. Maternal work and child care in China: a multi-method analysis, *Population and Development Review* 28(1): 31–57.
- Sieber, Sam D. 1973. The integration of fieldwork and survey methods, *American Journal of Sociology* 78(6): 1335–1359.
- Small, Mario Luis. 2009. How many cases do I need? *Ethnography* 10(1): 5–38.
- Small, Mario. n.d. Lost in translation: how not to make qualitative research more scientific, in Michèle Lamont and Patricia White (Eds), *Report from Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. Washington, DC: National Science Foundation.
- Stewart, Miriam, Edward Makwarimba, Alison Barnfather, Nicole Letourneau, and Anne Neufeld. 2008. Researching reducing health disparities: mixed methods approaches, *Social Science & Medicine* 66(6): 1406–1417.
- Stycos, J. Mayone. 1981. A critique of focus group and survey research: the machismo case, *Studies in Family Planning* 12(12): 450–456.
- Tashakkori, Abbas and Charles Teddlie. *Handbook of Mixed Methods in Social & Behavioural Research*. Thousand Oaks, CA: Sage.
- Townsend, Nicholas, Sangeetha Madhavan, and Anita Garey. 2006. Father presence in rural South Africa: historical changes and life-course patterns, *International Journal of Sociology of the Family* 32(2): 173–190.
- Twine, Rhian, Mark A. Collinson, Tara J. Polzer, and Kathleen Kahn. 2007. Evaluating access to a child-oriented poverty alleviation intervention in rural South Africa, *Scandinavian Journal of Public Health* 35(Suppl. 69): 118–127.
- Watkins, Susan Cotts. 2004. Navigating the AIDS epidemic in rural Malawi, *Population and Development Review* 30(4): 673–705.
- Willis, G. 2005. *Cognitive Interviewing*. Thousand Oaks, CA: Sage.