Published in final edited form as:

Am J Hum Biol. 2014 ; 26(5): 577–589. doi:10.1002/ajhb.22543.

Collecting women's reproductive histories

Cynthia M Beall

Case Western Reserve University, Anthropology Department, Cleveland, OH 44106

Paul W. Leslie

Department of Anthropology, and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3115

Abstract

The importance of women's reproductive histories for scientific questions mandates rigor in collecting data. Unfortunately, few studies say much about how histories were constructed and validated. The aim of this report, therefore, is to illustrate the elements of a rigorous system of data collection. It focuses particularly on potential sources of inaccuracy in collecting reproductive histories and on options for avoiding them and evaluating the results.

A few studies are exemplary in their description of methods of data collection and evaluation of data quality because they clearly address the main issues of ascertaining whether or not an event occurred and, if so, its timing. Fundamental variables such as chronological age, live birth or marriage may have different meanings in different cultures or communities. Techniques start with asking the appropriate people meaningful questions that they can and will answer, in suitable settings, about themselves and others. Good community relations and well-trained, aware interviewers who check and cross-check, are fundamental. A range of techniques estimate age, date events, and optimize the value of imperfect data.

Robust data collection procedures rely on skillful and knowledgeable interviewing. Reliability can be improved, evaluated and explained. Researchers can plan to implement robust data collection procedures and should assess their data for the scientific community to raise confidence in reproductive history data.

Keywords

Reproductive history; age estimation; demography; reliability; validity; cross-cultural comparison

INTRODUCTION

Reproduction is the coin of the evolutionary realm. Women's reproductive histories are central to addressing major scientific questions including detecting natural selection and other forces of evolution, and teasing apart the often complex relationships among fertility,

To whom the proof is to be sent, Cynthia M Beall, Case Western Reserve University, Anthropology Department, Cleveland, OH 44106, Telephone 216 368 2277, Fax 217 368 5334, cmb2@case.edu.

The authors have no conflict of interest to declare.

mortality, and environmental (physical, biotic, socio-cultural/economic) characteristics. Here are some examples. I. Lester Firschein reported that Black Carib mothers from what is now Belize, who were heterozygous Hb^A Hb^S reported nearly two more pregnancies and nearly 1.5 more surviving children than their Hb^AHb^A counterparts (Firschein 1961), thereby clearly demonstrating differential fertility of genotypes that is central to natural selection. Claudia Moreau and her colleagues found that Quebecois women at the front of a more than two-century long wave of population expansion have 20% higher effective fertility than women at the core of the expanding population (Moreau et al. 2011), thus contributing disproportionately to the gene pool. Mary Shenk and her colleagues measured total fertility and number of surviving children to discover that economic factors are the main drivers of fertility reduction in rural Pakistan, thus making an important contribution to understanding the course of the demographic transition (Shenk et al. 2013). Monique Borgerhoff Mulder investigated the tradeoffs between offspring quantity and quality among agropastoral Kipsigis in Kenya and learned that quality was more important to women (Borgerhoff Mulder 2001). The consequences for fertility of environmental factors such as high altitude or disease or family composition have been investigated (Hawkes 2003; Julian et al. 2009; Kjetland et al. 2010; Strassmann and Garrard 2011; Vitzthum and Wiley 2003; Wang et al. 2014) to improve understanding of evolutionary and health processes. Such questions often require data collection from a community sample rather than using data available from a census or large-scale survey.

It may seem that reproductive history is such a salient feature of a woman's life that collecting the information is straightforward. Actually, it is a complex task requiring knowledge, planning and skill. The aim of this report, therefore, is to consider some of the problems and pitfalls, and to illustrate the elements of a rigorous system of data collection, particularly from the standpoint of potential sources of inaccuracy in collecting reproductive histories in communities and techniques for avoiding them and assessing the results. We hope this report will help researchers to plan and undertake data collection and the scientific community in general to understand the process of collecting women's reproductive histories. We welcome information and insights from readers.

Reproductive history is defined as "An important aggregate factor in epidemiological studies of women's health. The concept usually includes the number and timing of pregnancies and their outcomes, the incidence of breast feeding, and may include age of menarche and menopause, regularity of menstruation, fertility, gynecological or obstetric problems, or contraceptive usage." [<http://www.ncbi.nlm.nih.gov/mesh/68017584> accessed November 29, 2013.] This bare bones definition belies the complexity of each element listed, as described below.

A PubMed search on 'women's reproductive histories' yielded 1957 references (November 29, 2013). The more than 100 publications per year since the phrase was introduced as a PubMed search term in 1997 might suggest that collecting such information follows well-established procedures. That is not the case, although the concept has been in use for generations. The search identified studies ranging from reports on knowledge, attitude and practice of birth control, to cancer risk factors, to *in vitro* fertilization, to community studies such as the ones discussed here. The typical report simply states that data were collected by

a questionnaire administered by research assistants or self-administered or taken from medical records. Only a small handful of studies discussed validating the data.

Some population scientists, surprisingly, have questioned the value of reproductive histories. For example, Cleland suggested “the collection of pregnancy or birth histories is no guarantee of higher-quality data than could have been obtained from alternative and less expensive survey strategies.” (Cleland 1996: 445). Some researchers have been sufficiently concerned that the close questioning needed to elicit detailed reproductive histories would be both time consuming and would force people to talk about embarrassing or disturbing events (perhaps compromising cooperation in other aspects of the research) that they decided to obtain demographic data in other ways (Coast 2000). However, the dominant view is that the substantial benefits of reproductive histories ensure their continued use. We agree, but see a need for greater attention to data quality.

There are excellent frameworks for studying fertility patterns. For example, the landmark publication by Kingsley Davis and Judith Blake (Davis and Blake 1956) introduced a conceptual framework for analyzing determinants of fertility by describing how cultural, economic, and institutional factors must act through a set of intermediate factors (affecting intercourse, conception, gestation and successful parturition) to result in a particular set of fertility rates and patterns, which are produced by aggregated reproductive histories. Subsequent scholars simplified their framework in order to facilitate quantification of the factors influencing fertility (Bongaarts 1978; Bongaarts 1987; Bongaarts and Potter 1983; Wood 1994) or modified it to focus more explicitly on determinants of fecundability (the probability of conception) (Campbell and Wood 1988).

While these conceptual frameworks focus on what to observe, less discussion focuses on how to observe. The frameworks establish the principle that reproductive history must be analyzed in environmental, sociocultural, and biological terms. However, they do not operationalize the process of collecting the reproductive histories themselves, or of establishing their reliability, for example their repeatability (similarity when re-measured) and accuracy (similarity to the true value).

BACKGROUND

Reproductive histories are collected to address a variety of research topics in human biology and related disciplines, yet researchers planning projects that entail collection of reproductive histories will not easily find advice on methods for constructing a reliable set of reproductive histories. Standard textbooks in demography and demographic methods (e.g. Siegel and Swanson 2004), usually written for or from the perspective of disciplines that tend to rely on large scale surveys (including censuses) and registries, typically say little about reproductive history data on the community level. Volumes on anthropological demography often do take the community perspective, yet do not provide methods instructions (e.g. Basu and Aaby 1998).

As a result, researchers must consult individual studies, including those reported in gray literature, doctoral dissertations or other difficult to locate sources. Even there, research

methods are often presented vaguely. Fairly frequently, some detail is given on the common and important problems of age estimation and dating of events but we see much less on assessing the validity of the data. Assertions of cross-checking are often made, accompanied by the authors' assurance of their confidence in the data, but with little or no indication of how this was done, with whom, and how the inevitable discrepancies were reconciled. Such confidence may be justified, but the assurances do little to convince the reader. More important for the present report, they are no substitute for a description of the field methods and validation techniques used. To the extent that this results from space limits in professional journals, the increasing use of links to supplementary material available on-line should alleviate such constraints. However, documenting the process of data collection and reporting is essential for both evaluating published research results and sharing effective techniques.

A few book-length reports of research that makes use of reproductive history data stand out as exemplary exceptions to the general picture of methodological silence. They include Nancy Howell's Demography of the Dobe !Kung (Howell 2000; Howell 1979), which has been deemed "perhaps the most important study ever of anthropological demography" (Pennington and Harpending 1993: 47). Another is the book by Renee Pennington and Henry Harpending who devote a chapter to *Field-work and methods* in The Structure of an African Pastoralist Community (Pennington and Harpending 1993). A third is Kim Hill and Magdalena Hurtado's Ache Life History. The Ecology and Demography of a Foraging People (1996) which has a chapter and several sections on field work and data quality. We believe it worthwhile to expand on these studies and to compile and present an overview of the methodological challenges and best practices associated with collecting reproductive histories. We draw on our own experience and that of others, both published and unpublished. The focus is on women's, rather than men's reproductive histories. As James Wood noted, it is easier to ascertain maternity than paternity and women are the 'rate limiting step' of fertility (Wood 1994: 7). However, the issues encountered in collecting men's reproductive histories overlap substantially with those discussed here.

METHODOLOGICAL PROBLEMS AND SOLUTIONS

General considerations

Awareness of potential problems is useful in itself, as it may forestall overconfidence and over-interpretation of data or results. Given that awareness, the questions of how to detect and how to mitigate problems arise.

One problem is the possibility of bias in sample selection. While this is more properly a study design issue, we discuss it briefly here because it could contribute to or exacerbate some of the methodological challenges described later. Selection bias arises when the probability of being included in the study sample is related to characteristics that are of interest in the study. For example, if the research question addresses changes in fertility patterns over time in a community where men and women move to retirement communities after their children marry, the sample may inadvertently exclude a proportion of older adults and have a bias toward younger ones and a shorter time span. In some societies women who are sterile or who have very few surviving children may be more likely to leave the

community; this will bias the collected reproductive histories toward those characteristic of women with higher fertility and/or lower child mortality.

Another source of sample bias takes the form of “censoring”, a situation where information on a given event cannot be collected from all study participants because not all have experienced it yet or because they experienced it and left the community (through death or emigration) before the study. Reproductive histories provide information about reproduction-relevant events from a woman’s birth to the time of data collection. If she has not yet married or had a first pregnancy or achieved menopause, then the timing of those events is unknown for her. In such a situation, the data are said to be right-censored or right-truncated (with time viewed as progressing from left to right). If the mean age at menarche is estimated from a sample of females who have reached menarche by the time of the study, right-censoring may bias the results by preferentially excluding individuals who will reach menarche at older ages (unless appropriate statistical techniques are used or women in the sample are all well beyond menarche, say 20 years or older). On the other hand, if maternal mortality is higher in higher fertility women, then left-censoring may lead to under-estimation of past fertility because their reproductive histories are preferentially excluded from data collection unless corrective measures are undertaken. One approach to obtain information on deceased women is the sisterhood method of asking men and women about their sisters including those who have died (Merdad et al. 2013; WHO Department of Reproductive Health and Research 1997).

A third source of sample bias occurs when the sample represents cohorts of people who lived through different times. The Ache study is a dramatic instance. It covered the time from 1890 to 1993 and included three contrasting periods: a forest hunter-gatherer period, a peaceful contact (with larger Paraguayan society) period, and a subsequent reservation period. The Ache situation is more marked than many, but researchers need to consider the possible influences of changes in cultural, economic and institutional factors on their samples and, perhaps, on their reporting. In recognition of this, Hill and Hurtado present three sets of results corresponding to the three historical periods.

Sample bias affects which women’s reproductive histories are included in an analysis. In addition, two general categories of methodological problems are encountered when compiling the individual reproductive histories comprising the sample:

- 1) ascertainment – did an event happen? Ascertainment problems may entail under-enumeration or (less commonly) over-enumeration that is random or that reflects systematic error.
- 2) timing – when did events happen? Timing problems may entail uncertain age or date estimation and inaccurate recall of past events as well as bias, for example toward greater accuracy of recent events.

Ascertainment problems

Ascertainment problems stem from underreporting and selectivity, two sources of error that in some cases aggravate one another. Underreporting may be intentional (born of reluctance to mention children who died or fear of witchcraft or other harm, for example) or

unintentional – simple recall error, inattentiveness, or lack of awareness (e.g., unrecognized fetal death or difficulty distinguishing between stillbirth and very early infant death). Selectivity can interact with underreporting. For example, in retrospective reproductive histories, memory bias can result in an incorrect picture of increasing fertility in recent times, as more recent events are more likely to be recalled accurately.

Inaccurate ascertainment may result from misunderstanding the question.

The standard census-type question of “How many children do you have?” exemplifies getting what you ask for, but not what you wanted, because of poor choice of vocabulary or phrasing.

Aside from the reluctance of people in some societies to count their children as reported by Pennington and Harpending (Pennington and Harpending 1993), the answer clearly depends on how “have” is understood by the respondent. Consider the differences among the questions: “How many children do you have?” vs. “How many children have you had?” vs. “How many births have you had?” Responses to the former may well omit offspring who died or were fostered/lent out (the parent no longer “has” them with them) and lead to under-reporting. On the other hand, these responses may include biological children, grandchildren, and adopted children. They may include fostered-in children and exclude fostered-out children, and in some cases might include borrowed or bought children or children under labor contract and lead to over-reporting. Ways of avoiding such confusions include asking “Who are the children to whom you have given birth?” rather than “Who are your children?” (Pennington and Harpending 1993:48). Asking about a woman’s sequence of pregnancies and the outcome of each not only helps skirt some of these problems but may also improve recall. If taking a genealogical approach, it may be better to ask “Who gave birth to you?” rather than “Who are your parents?”. This is because in some communities, such as on Ifaluk in Micronesia, the majority of individuals have both biological and adoptive parents (Turke 1988).

Timing problems

Poor estimation of ages and timing of vital events, arising from lack of written records and different systems of reckoning time and age in different cultures, is one of the most common problems faced by those collecting reproductive histories or studying demography or human population biology more generally.

The accurate measure of time since birth can be difficult to obtain. Some societies pay little attention, have few or no reliable written records, or may not use the Western calendar. Age (one’s own or that of others) may be simply unknown, or only roughly known, subject to recall bias, age-heaping on salient numbers such as 5 or 10 years, or systematic over- or under-estimation. Cultural or institutional factors may prompt intentional misreporting. For example, in societies where age confers status and respect, older individuals may exaggerate their age; in places where families wish to avoid mandates to send their daughters to school (thus retaining their labor at home), parents may purposefully misstate the ages of young girls. Age inaccuracies cascade throughout analyses. An incorrect chronological age may lead to an incorrect age at marriage, time to conception, age-specific fertility rate, and so on.

As discussed below, the means of coping with timing problems vary with the type of event, though there are some commonalities.

Timing events in the reproductive history

Age at first pregnancy.—Information on this event marking the onset of reproduction depends on factors including recognition of pregnancy. For example, !Kung women in Namibia suspect pregnancy after one missed menstrual cycle and acknowledge it after two (Howell, 2000). Very early pregnancy loss may not be readily recognized and is thus not ascertainable by recall.

Irregular menses complicate recognition of pregnancy or its loss. For example, Indonesian women report pregnancies lasting from six to 16 months (counting back to last recognized and recalled menstruation). In some cases, then, a report of one successful pregnancy may actually be a report of one lost and one successful (Vanessa Hildebrand, personal communication).

Subsequent pregnancies.—Interview schedules take many forms depending on the appropriate and most effective approach to asking information in the particular research setting. Indeed, a structured conversation may be more appropriate than an interview in some instances. Some interviews or conversations proceed from first to subsequent pregnancies with the intent that the temporal sequence prompts accurate recall. Others ask about living children, infants, newborns and then follow with all deaths of children, infants, newborns and then with all pregnancies (Vanessa Hildebrand, personal communication). A reviewer of this report noted that in some contexts it may make more sense to proceed from youngest living child backwards in time to the oldest. When the youngest living child is present during the interview, he or she can provide a starting point for the interviewer to ask if there have been any pregnancies since that child was born (and next to ask about any between the youngest child and the next youngest living child and so on). Yet another option is to move from the youngest directly to the oldest living child. The oldest can provide a starting point for ascertaining if there were any pregnancies before that living child. After obtaining information on events before the birth of the oldest and after the birth of the youngest living child, the interviewer can proceed to the time in between. It may take some pre-testing to tailor the approach to a particular research setting. Some suggestions for facilitating such tailoring are included in the section on *Enhancing Reliability* below.

To evaluate accuracy, it is helpful to have a sense of the ‘usual’ time between conceptions and then to follow up on unusually long or short gaps. For example, Tibetan women in many rural communities generally have children every two or three years in the absence of birth control (Childs et al. 2005; Goldstein et al. 2002). During a recent study, follow-up questions to women reporting a gap of three or more years in between pregnancies resulted in prompting recall of additional pregnancies by 10% of the women (Beall, Childs and Craig, unpublished), as well as reports of the use of birth control and plenty of reports of ‘no reason’. Similarly, a report of two births in one calendar year prompted follow-up and identified pairs of twins (and one less pregnancy) as well as occasional pairs of siblings born within a single calendar year.

Reproductive histories with suspiciously long unexplained gaps might be dropped from the sample, but that should be done only with great caution as these gaps might be produced by temporary sterility induced by certain infections (Howell 2000) or absence of sexual partner (e.g., for labor migration) or prolonged breastfeeding (Konner and Worthman 1980) or complications during pregnancy or birth (Filippi et al. 2000; Koblinsky et al. 2012). Methods such as hormone or ultrasound testing (Ober et al. 1999; Ober et al. 1983; Vitzthum et al. 2006) can provide data relevant to these and other possibilities such as early pregnancy loss, but don't help with retrospective data collection.

Current pregnancy.—If a current pregnancy were underway and recognized then, the woman may or may not be prepared to reveal that information. Occasionally, a woman has approached one of us (CB) privately after an interview to report a pregnancy of which her husband was still uninformed.

Outcome of pregnancies.—The accuracy of reports of abortion, miscarriage, stillbirth, or live birth depends partly on how the distinction is made among those outcomes, the acceptability of discussing such topics, memory and recall. Poor outcomes or culturally awkward outcomes may simply be omitted or forgotten. Occurrence outside of or before marriage or soon thereafter may be embarrassing and may lead to deliberately inaccurate information. Nancy Howell (2000) explains that !Kung women are not reluctant to discuss fetal wastage but consider it unimportant, so this is likely underreported. Additionally, it may be difficult to distinguish a miscarriage from a stillbirth or a stillbirth from a live birth. Renee Pennington and Henry Harpending asked “was the baby born breathing” (1993) to elicit the stillbirth-live birth distinction.

Infant and child mortality.—The amount and timing of post-natal mortality become important because early offspring mortality may influence a woman's reproductive history (see review in Shenk et al. 2013). Also, studies of reproductive success require information on the number of children surviving to sexual maturity. To simplify data collection, a particular study may choose survival to 5, 10, or 15 years of age as an indicator of survival to maturity (Strassmann and Gillespie 2003). After ascertaining the outcome of each pregnancy, the current status of each live birth must be asked, with all the attendant potential of biased recall or inaccurate ascertainment and contributing factors. In some settings the site of delivery is relevant (Pervin et al. 2012), such as giving birth at husband's home or parental home (Basu and Aaby 1998) or in a home versus a clinic (Pervin et al. 2012).

Marriage.—The definition of marriage can vary widely. Howell (2000: 227) provides a good breakdown of the dangers of ethnocentrism in defining marriage in reproductive history studies. It is important to understand the range of options, their relative statuses, and the vocabulary used to discuss them. Indeed, “marriage” in many societies is more a process than a discrete event (Meekers 1992), so date of marriage may be ambiguous, or there may be several dates associated with different stages in that process. Some find that asking about marriage after asking about pregnancy avoids signaling expectations about the sequence of events. If paternity is important to the study, then DNA testing is an option (Strassmann et al. 2012; Tonelli et al. 1990).

In addition, the type of marriage can vary widely with consequences for reproductive history. For instance, Tibetan women in polyandrous marriages are rarely widowed. Thus they have a longer exposure to the 'risk' of pregnancy (Beall and Goldstein, unpublished data). On the other hand, the polygyny-fertility hypothesis reasons that polygyny reduces fertility of individual wives by reducing that 'risk' (Winking et al. 2013).

Reproductive senescence.—Another aspect of reproductive histories is the end of reproduction due to reproductive senescence leading to pre-menopausal secondary sterility and then menopause. In most of the topics discussed so far, the issue is ascertainment of events. In the case of end of reproduction, the aim is to ascertain the absence of menstruation. PubMed defines menopause as “The last menstrual period. Permanent cessation of menses (MENSTRUATION) is usually defined after 6 to 12 months of AMENORRHEA in a woman over 45 years of age.” [accessed February 12, 2014, emphasis in the original]. Compound the difficulty of evaluating the length of the absence of an event with recall bias (perhaps easier for younger post-menopausal women to recall) and the absence of menopausal symptoms in some communities (Martin et al. 1993) and the challenges of collecting reliable data on the timing of reproductive senescence mount.

ASKING QUESTIONS

Obtaining accurate data on both the occurrence and timing of vital events depends on how questions are asked of informants, and the context in which they are asked. We thus consider some pitfalls and best practices related to eliciting answers. Many of these will be familiar because they pertain to interviews in general. Undeniably, this careful approach takes longer than recording answers uncritically.

Whom to ask?

Just who the best sources of information are depends jointly on people's knowledge and willingness to communicate that knowledge. As a rule, women know their own reproductive histories in greater detail than their husbands. For that reason, Pennington and Harpending (1993) focused their reproductive history interviews on women. They found that Herero women name their biological children while men name their dependents. They also found that men were reliable informants about their own children but were often not so useful concerning wives' or girlfriends' reproductive histories before marriage or courtship. This was certainly true of Turkana interviewed by Leslie, Fry, and others as well, but women were much worse than men at dating events – men were much more adept at relating births, marriages and other occurrences to an event calendar (see below), perhaps because they were in the habit of using the named years and seasons in their frequent deliberations about herding strategies under changing conditions (Leslie and Dyson-Hudson 1999). It was thus best in that case to combine information garnered from husbands and wives.

Willingness to talk with interviewers is also crucial. Pennington and Harpending asked mothers about their daughters because the young daughters were often quite bashful, resulting in unreliable interviews. This also avoided ascertainment problems (due to young women with no children being away at school). In fact, many studies rely on multiple sources for individual women's information (see below).

Appropriateness of topics and questions

Some topics, such as deaths of children, are sensitive to varying degrees in most societies. In some, there is a deep reluctance to talk about deceased family members, sometimes extending to taboos on mentioning their names. This can obviously have a profound effect on the quality of reproductive histories, especially in populations with substantial early mortality. In other cases, a genealogical term may be appropriately used for a deceased individual although it would be shocking to use his or her name. In still others, very young children may not have been named before death.

Other topics may be problematic only in some places or if broached in certain ways. Thus, Pennington and Harpending found it inappropriate to ask “How many children do you have?” because Herero are averse to being counted – things, including livestock, are counted, not people. The solution in this case was to ask women about their sequence of pregnancies and the outcome of each.

Informants’ willingness to provide accurate and detailed reproductive histories may vary greatly depending on when they are approached and the social context. Who else is around and listening to, or at least aware of, the interview? Who is interviewing whom? (See section on relationships and context, below.)

Even when topics are fully appropriate, survey questionnaires sometimes ask for information in terms that are ambiguous or not culturally relevant, resulting in data of poor quality (Levine and Scrimshaw 1983; Meekers 1992). Meanings of conception, pregnancy, or family, for example, vary. The notion of premarital reproduction clearly depends on the understanding of “marriage” -- there may be different types of unions and the transition from unmarried to married may not occur at a single identifiable point. Questions about biological paternity may be inappropriate in some cases, but in others they may simply be understood differently – as when there is belief in partible paternity (a child may have more than one biological father) (Hill and Hurtado 1996; Mesoudi and Laland 2007). Consequently determining the best ways to ask questions is an important aspect of improving reliability. A tradeoff is that greater cultural specificity may compromise comparability to other studies. However, the benefit of improved accuracy of the reproductive histories obtained will often outweigh the cost. These considerations are of course relevant to sound practice for interviews or surveys in general.

What to include and what to exclude

There are often tradeoffs between the data needed and the quantity and quality of data that can be obtained. They may arise from constraints on the time informants can devote to an interview or to fatigue from doing so, as well as from the sensitivity of various topics. Fatigue leads to inaccurate answers, and even omission of events in order to hasten completion of the interview. Consider what the interview may be interrupting or delaying and be prepared to proceed, postpone, or truncate an interview, perhaps completing it later. It is generally preferable to have less data if it is essential and of high quality than to have more extensive but unreliable data.

ENHANCING RELIABILITY

Relationship with study community and context of interviews

It is difficult to overstate the importance of establishing trusting relationships with specific informants and the community. Hill and Hurtado (1996) comment specifically on the value of conveying interest in the informants' lives and the accuracy of their personal histories and experiences. Researchers frequently report that trust born of extended, respectful contact with a community results in more reliable reproductive histories (e.g. Bollig 2006). Garenne (Garenne 1994) analyzed the quality of maternity histories of Senegalese women grouped by fieldworker and concluded "a major determinant of the quality of the data seems to be the relationship between the enumerator and the interviewee." This of course goes for the quality of information about related topics as well as the reproductive histories themselves.

Thus, politeness, respect, and cultural sensitivity are important from the outset of the study. Still, trust derived from long term relationships with the respondent and/or her family and the larger community may make a further difference to data quality. This means that, in some cases, reliable reproductive histories may be obtained only after such trust is gained. One of us (PL) collected and updated reproductive histories among Turkana in northwest Kenya over more than a decade. Reproductive histories collected during the first few field seasons revealed substantially fewer child deaths than did later interviews. Some of this was no doubt attributable to learning on the part of the interviewer, but the familiarity and deepening relationships with Turkana families developed over that time surely helped, as well. The !Kung and Ache studies undoubtedly benefitted in similar ways. Realistically, however, those studies extended over many years and are not typical.

Having a research assistant from the community may also help, when feasible. A local assistant may know many of the families well enough to be able to notice and flag omissions or inaccuracies. And informants may know that the assistant knows, and thus be less likely to omit events or answer carelessly. On the other hand, if people prefer not to report private information to their neighbors then a local assistant may be a hindrance.

But even with local assistants, characteristics of the interviewer may be crucial. The interviewer's gender may make a big difference (especially for given gender of informant). In many places, women may be more comfortable and forthcoming talking with another woman, although this may vary with the age and/or marital status of both the informant and the interviewer. Johnson describes how data quality in her work in Papua New Guinea was affected by differences between interviewer and respondent in status, sex, and language, and by the context of interviews (group or individual, who else was present) (Johnson 1987). For example, a woman may flatly refuse to respond to questions about menstruation, conception, or whether she is currently pregnant, when those questions are posed by a male researcher (and, indeed, male research assistants may resist asking women about those same topics). Conversely in some societies, men may feel less obligated to work thoughtfully with a woman (especially if she is younger than they are) than with another man, thereby compromising accuracy of results. Worse, they may perceive the female interviewer or interpreter as inappropriately performing a male role and resist answering at all. Note that gender matching is not always an issue -- Hill and Hurtado found that that sex of the

interviewer did not matter with the Ache. Aside from gender, careful pairing of interviewee and interviewer may be required if there are divisions in caste, religion, or other characteristics that may undermine a good interview. Thus, the attitudes and perceptions of both respondents and interviewers are relevant.

Whether the interview should take place privately with the woman herself or in the presence of other people, usually female neighbors or relatives, will vary from situation to situation. In some, a private interview with a stranger would be inappropriate or would arouse concern about the topic or intentions. At the same time, certain topics, such as miscarriages or abortions, may be sensitive to the extent that those events are not reported at all or are not reported in the presence of certain individuals, such as the mother-in-law. In other settings, the same topics may be freely discussed and reported in a group.

At times, it is productive to use multiple informants concurrently rather than sequentially. For example, Pennington and Harpending got much better information about years of birth and death of family members when there were group discussions among kin sitting in on interviews (Pennington and Harpending 1993 p. 45). Family members can and do chime in with information the primary informant does not know, and they correct dates or omissions that the interviewer might otherwise take at face value. The group may act as distributed memory, particularly to date events (“Our daughters were born the same year, and mine is x years old because she was born when such-and-such an event occurred”). This may not be the best approach in all cases, but small group interviews are often worth considering. In our experience, they often occur spontaneously in any case.

Testing and training

Care in the formulation and expression of questions, and to whom these are addressed, is crucial. The possible confusions associated with phrasing questions about numbers of children and marital status were noted earlier. It thus makes sense to test questionnaires or protocols for less formal interviews before implementation to help ensure that questions are understood as intended. Many of these preparatory procedures are common in social science research – for example, having a set of questions translated into the local language and then translated back (by a different translator) to help identify likely misunderstandings – and we don’t cover those here. It is also advisable to train those who will be collecting the reproductive histories to give them experience in interacting with members of the study community and in identifying problems and misunderstandings as they crop up.

A useful resource for training assistants to deal with a variety of interviewing situations is found in the manuals and reports from the Demographic and Health Survey. It is the current project in a series that has collected and analyzed population and health data throughout the world for nearly three decades (Corsi et al. 2012). The Interviewer’s Manual and suggestions for question phrasing and prompts during interviews may be useful (ICF International 2012) (also refer to <http://www.measuredhs.com/publications/publication-search.cfm?type=35>, accessed November 30, 2013).

A few measures to consider are listed below.

- Explain to interviewers the intent of each question and the need for accuracy.

- Discuss the possible misunderstandings by the enumerator as well as the interviewee and develop a set of prompts, follow-ups and double-checks to use during the interview
- Have interviewers collect reproductive histories from an informant whose information you already know well and see how well the test interview compares.
- Have an enumerator interview you or someone else who purposefully includes problematic responses (impossible or unlikely responses that should be probed -- e.g., long gaps between reported births, or two births within 6 months).
- Precoded questionnaires may have the unintended effect of limiting follow-up or probing. Therefore interviewers should be encouraged to write marginal notes when information does not correspond neatly with codes.

Cross-checking with other sources of information

Perhaps the most generally effective means of enhancing the reliability of reproductive histories is to find ways to validate them through cross-checking. These include obtaining a given individual's reproductive history from multiple informants (e.g., husbands or female relatives, in addition to the woman herself), and use of other data sources (e.g., genealogies, birth or baptismal certificates, horoscopes or marriage licenses). Bollig (2006) describes how he obtained reproductive histories from Pokot women in Kenya by asking about number of offspring and dates of vital events, but supplemented these data with checks from repeated household censuses over eight years. This helped ameliorate the common problem of eliciting the timing of events. Reporting the extent of concordance between interview data and these other sources can help to assess its completeness and accuracy.

Use of multiple informants is especially effective because individuals vary in their ability to recall details of past events and willingness to discuss details surrounding them. Husbands, co-wives, and other relatives can supplement or correct the reproductive histories, especially concerning topics that may be sensitive. Many Turkana women were quite willing to discuss deaths of their sisters' children but failed to mention births of their own children who had died (Leslie et al. 1988). On the other hand, the reproductive histories obtained from others may be less detailed or less accurate in other ways than one obtained from the woman herself. This points to the value of having reproductive histories from multiple sources.

Hospital or vital records may validate or supplement data obtained by interview (Wiley 2004). For example, comparing the characteristics of one's sample with existing census or vital statistics information can evaluate the extent to which the two have similar composition. Either or both may be biased, however considering the possibility may open some avenues for checking. Records may provide information on seasonal or annual variation in births or deaths that could confirm reported patterns of neonatal mortality, causes of death and illness.

Repeated interviews

When feasible, collecting reproductive histories more than once – perhaps in subsequent field seasons – from the same person can improve both ascertainment and reliability of dating. In addition to updating reproductive histories with more recent events, discrepancies between paired reproductive history interviews are easily noticed and (often) easily resolved (Leslie and Dyson-Hudson 1999; Leslie et al. 1999a; Strassmann and Gillespie 2002). Formal measurement agreement techniques may be applied during data analysis to estimate the average size and standard deviation of differences between original and repeat interviews of a sample of women (Bland and Altman 1986).

Repeated interviews also provide an opportunity to more actively probe for errors and to assess the likely quality of information. Hill and Hurtado (1996:90) did this by purposefully stating incorrectly events previously reported (e.g., a wrong cause of death). In addition to helping to confirm (or call into question) data previously obtained, the informant's ability and willingness to catch and correct the mistake signals his or her level of attention and commitment to accuracy.

Recall and cross-checking may be supplemented with observation. A prospective study design collects information through observation or periodic interviews going forward from the study's beginning. It has the advantage of enabling recording of reproductive events as (or soon after) they happen. However, such studies are rare because they incur substantial expense and time. At the long end of the prospective spectrum, Carole Ober and her colleagues reported on the heritability of reproductive fitness traits, including completed family size measured as the total number of live births and stillbirths obtained from interviews of Hutterite men and women in North Dakota, US, over 35 years between 1982 and 2007 and checked against religious records (Kosova et al. 2010). At the shorter end of the prospective spectrum, Kathleen O'Connor and her colleagues studied age-related differences in fecundability among 700 rural Bangladeshi women through interviewing and conducting urine tests twice weekly for eleven months (O'Connor et al. 1998).

It is worth noting that, in some societies, people may have more than one name or may change their names. The name used in addressing them or referring to them may change over time or may differ according to who is speaking. If that is the case, care must be taken not to assume that a new name arising in a follow-up reproductive history represents an offspring omitted from the original interview rather than the same person with two different names.

Degrees of reliability

Reproductive histories are used for a wide range of purposes, and not all purposes demand the same degree of completeness and precision in the reproductive histories. Histories that lack accurate birth dates of offspring may be quite adequate for estimating completed family size; histories with likely underreporting of early deaths may still be informative in an analysis of nuptiality (the marriage rate) or net reproductive success. Thus, it can be desirable to rate the reliability of reproductive histories with regard to intended uses.

Perhaps the most detailed effort to address the biases that can arise in reproductive data obtained from interviews is that by Hill and Hurtado in their book Ache Life History: The Ecology and Demography of a Foraging People (Hill and Hurtado 1996). They identify potential biases and then explain how they assessed the likely importance of those problems in the study population and the methods they used to minimize those problems. They describe in detail how they cross-checked the Ache reproductive histories they collected and how they evaluated their reliability. This included categorizing reproductive histories according to the genealogical relationship between the informant and the subject (“ego”, the person whose reproductive history is being investigated). Through comparison of cases with multiple sources of information gathered from different relatives, they determined which categories were less reliable and in what ways they were so. For example, ego’s nieces and nephews generally provided reliable information about adult deaths but not about childhood deaths; the latter were more reliably obtained from parents, siblings, or ego herself.. Categorizing the reproductive histories allowed them to choose which reproductive histories to include in the samples used in different demographic analyses, enhancing sample sizes without compromising accuracy. Such categorizing also flags people for whom better data are most needed, which may make follow-up fieldwork more efficient.

Hill and Hurtado also discuss means of detecting sample bias – e.g., reflected in skewed sex ratios. They consider how biases are likely to affect the estimation of different fertility and mortality measures and specify why they used certain sources of data for specific analyses. For example their Table 3.4, *Sources of Data for Analyses Reported in Chapters 5–9*, lists the type and source of data for various analyses. Importantly, in discussing reliability of interview data they not only describe their means of assuring reliability (re-interviewing; cross-checking results from different sources – other informants or genealogies), but also present data on the number and type of cross-checks they made and on the degree of agreement among these checks. Such transparency and detail are rare and welcome.

In establishing criteria for evaluating reliability care must be taken not to use criteria that will result in selectivity bias. For instance, consider a reproductive history with no infant deaths. Deaths may have been forgotten or information about them may have been deliberately withheld, but perhaps none occurred. A criterion of ‘exclude the reproductive history if no infant deaths were reported’ might artificially inflate infant mortality rates. Criteria are better based on the source of data (ego or others; multiple sources), but even then care is needed. For instance, if having multiple sources is deemed to improve the quality of reproductive histories, and if second sources are often co-wives, the sample of higher quality reproductive histories may be biased in favor of polygynous households. Women in such households may well have different fertility patterns than those without co-wives (Leslie et al. 1999b).

Researchers in Australia, several European countries, and the US report good repeatability for numerous measures of reproductive history including age at menarche and last menstrual cycle, number of pregnancies, and age at first livebirth. They measured the proportion of women who remained in the same category from the original measurement to remeasurement and the agreement of two measures of continuous variables. Agreement was good, for example, 92% for the number of pregnancies, 100% for age at first livebirth in one

study (Slanger et al. 2007) and similarly high for other measures (Bosetti et al. 2001; Joffe et al. 1995; Lin et al. 2002; Rohan et al. 1988; Tomeo et al. 1999). Older or less-educated women provided less reliable information. Those samples had generally low numbers of pregnancies with many women reporting two or fewer. In a higher fertility setting of rural Senegal, Garenne compared cross sectional survey results with a longitudinal database of maternity history and found little underreporting across the age span of women 15 – 89 years. Underreporting was estimated at 2% of births and 4% of deaths (Garenne 1994; Garenne and Van Ginneken 1994). Reports of such reliability from diverse populations are encouraging, but it would be risky to assume similar reliability in all cases.

CORRECTIONS AND GAP-FILLING

Sound research design coupled with careful data-collection methods go far toward ensuring reliable reproductive histories, but gaps and deficiencies in reproductive history data are seemingly inevitable. There are a number of techniques for handling these deficiencies.

Dating vital events

In the absence of written records (civil or church birth, marriage and death registries; clinic cards, etc.) information about past vital events resides primarily in people's memories. Some means of handling the problem of dating vital events, such as the use of event calendars, are fairly well known and widely applicable. Other means are much more *ad hoc* and situation-specific. The precision of these methods, their attendant sources of error, and therefore their utility and appropriateness, vary markedly with the study population, the specific problem at hand, and the time and resources available for fieldwork. Estimating ages and the timing of developmental or life history events can be an important part of studies with foci other than reproductive histories (e.g., child growth and development; household demography, productivity, and nutritional requirements). The following discussion is thus more broadly relevant.

Discussions with researchers and the occasional published rigorous analysis identify several ways of addressing the problem of dating vital events (Ewbank 1981). One important task is to ascertain if more than one calendrical system is in use. It is important to establish this possibility early in data collection and to record the calendar used by each person providing an interview. To illustrate, consider two people from Tibetan society who report an age of 24 years. The researcher needs to learn the animal year of birth to determine if they are really the same age and to know that the animal year system adds the year of birth to the age. That is, an age of 24 indicates being in the 24th year of life and corresponds to a Western age of 23 years, which indicates the number of full years lived. One of those two people may report 24 years of age in the Tibetan calendar and the other in the Western calendar, in which case one is 23 and the other 24. Calendar converters (e.g. between the Western, Chinese, Persian or various local calendars) can be located or created.

Age ranking, a technique mentioned by Hill and Hurtado (1996) and Howell (1979), calls for asking whether the person of interest is younger or older than persons x and y and so on in the community. The resulting age ranking may be anchored with a historical event or record or a few individuals with relatively certain ages. Howell devotes considerable

attention to the problem of determining the timing of events and especially to age estimation, and provides a detailed description of her age estimation procedure, based on relative aging (rank ordering people by age, then fitting to a model age distribution).

People can often stipulate the timing of vital events in terms of past occurrences that may be political/economic (e.g., their country's independence), cultural (initiation of an age set) or institutional (the construction of a hospital). In some cases the past occurrence may pertain to the physical or biotic environment (e.g., a flood or epidemic). Event calendars constructed from a sequence of memorable occurrences yield absolute (though sometimes imprecise) dates.

Ideally, an event calendar is "fixed" – a single name refers unambiguously to a given year and everyone is knowledgeable about those names, at least for the years relevant to the reproductive histories in question. The Herero of Botswana have a system of year names and, for cultural reasons, most Herero know the name of the year of their birth and those of relatives and friends. This, along with concordances with the Western calendar years as far back as the 1830s, provided Pennington and Harpending (1993) with close to that ideal situation. There were a few problems, such as reuse of a year name, but Pennington and Harpending were able to work out ambiguities, especially for the generations with living members at the time of their study, resulting in enviable accuracy in age estimation.

More common are practices such as found in Turkana, Kenya. There, seasons are referred to in terms of memorable events, but the naming is less formalized and more local: a year or season may have more than one name, even in a single locality, and some names used vary among communities or lineages/clans within the region. It thus took several years to develop and refine the Turkana event calendar, which entailed working out both the sequence of the named seasons and the concordance with Western years (Leslie and Dyson-Hudson 1999; Leslie et al. 1999b). This proved to be extremely useful and well worth the effort but had some limitations – e.g., women were much less adept at associating events in their own reproductive histories with the named seasons in the event calendar than were men. On the other hand, because the names refer to seasons, and there are both wet and dry seasons in most years, events could often be dated more precisely than one year. In cases where people do not habitually name years, it may still be possible to construct an event calendar by first probing for events that are widely remembered, preferably events that are datable in the western calendar (major political events, eclipses, etc.).

Event calendars can be affected by their own patterns of misreporting, such as especially well-known events acting as "attractors" (Hill 1985) – that is, certain years that are very well known (e.g., the year of independence from colonial rule) may be identified as marking events that actually occurred in some earlier or later year. Such misattribution is especially likely when a respondent has a poor memory of the event or knowledge of the event calendar. This is akin to "age heaping" or "digit preference" noted in census data, where ages ending in zero (and to a lesser extent, 5) attract more responses than they should (Barclay 1958). Such heaping generally worsens the farther back in time one goes.

An event calendar can be a great boon to constructing reproductive histories. Even if it is not feasible to develop a year by year event calendar, or if there are a few gaps in it, for some purposes 5-year (or even 10-year) age or time intervals are all that are needed. For instance, ages at death of some children may be established within a narrow range that is sufficient to denote that the death was during infancy or childhood or whether the child survived to reproductive age. If the event calendar is incomplete, or if knowledge of it is spotty, recalling an earlier event and estimating the elapsed time since that event may be sufficient. For example, a woman may recall the year of her marriage and the time to pregnancy, from which one can estimate her age at first birth.

Widely recognized developmental markers or stages afford another means of estimating ages of children and adolescents, and the timing of events such as childhood deaths and birth intervals. These include stages of motor development (crawling, toddling, walking), markers of reproductive maturation (menarche, adrenarche) and markers that are more biobehavioral and less discrete such as weaning or self-feeding. Andrea Wiley and Ivy Pike (Wiley and Pike 1999; Wiley and Pike 1998) present the logic for the utility of developmental stages. They note that it can be fruitful to analyze survival to given developmental stages rather than to a set number of years (e.g., one year of age, for infant mortality). Tooth eruption pattern can be informative (Townsend and Hammel 1990) but is likely to require that the child be present for examination. Cross-cultural studies demonstrate that the timing of such markers varies among populations as well as by environmental factors including nutritional status, maternal education and marital status (Fernald et al. 2012; Prado et al. 2013; Van de Vijver and Tanzer 1997). It is also possible that some markers will not be useful in all cases. For example, Tracer et al. (Tracer et al. 2000) found that not all Au children in Papua New Guinea go through a clear crawling stage. Considering that the accuracy of chronological age varies, too, in some cases the markers may be just as informative.

Thus, in constructing reproductive histories, the interviewer might ask about whether a given child has yet reached these developmental stages, or whether it had reached them before death. If developmental markers are the main means of obtaining ages for young children or infant mortality, and if those are important to the research, then it may be worth considering formal study of the topic in the study population.

Michael Gurven and colleagues (Gurven et al. 2007) determined years of birth and death of Tsimane, lowland forager-horticulturalists in Bolivia, by combining several of the above techniques. For example, they “assigned” (their term) birth years based on ages known from interviews with individuals and relatives, lists of relative age ranks, missionaries’ written records, dated historical events, photo comparison of people with known ages, and developmental stages. They reasoned that the various sources provided somewhat independent estimates of age. As a consequence, they accepted the average of the separate estimates when all clustered within a 3-year range unless there was a reason to believe one or two estimates were better than the others.

Simulating data

When reliable and complete data are difficult to obtain by observation or from community members, simulation may provide estimated values that allow pursuit of many research

questions. One of the best known examples of using simulation to augment empirical data in order to render reproductive histories more useful is Howell's (1979) estimation of !Kung ages. The !Kung in her study did not know their absolute ages but, as is often the case in societies where age helps define status, rights and obligations *vis á vis* others, they knew their relative ages well. Howell was able to rank order members of the population by age and then fit that ordered list to an age distribution chosen from Coale and Demeny's set of models (Coale and Demeny 1966). That is, she converted the rank order into absolute ages so that the number of individuals of each age matched the number expected in a population characterized by a mortality pattern appropriate for the !Kung population. Her analysis of the procedure indicates a 2–3 year average error for the estimates of absolute ages, sufficiently accurate for much subsequent analysis.

Correcting for missing data

It is sometimes desirable and possible to compensate for missing data when estimating population level demographic parameters. For example, correction for the common problem of underreporting of child deaths might be done by estimating rates of early mortality using only reproductive histories of women who are apparently willing to report child deaths (because they did report one or more deaths). That rate could then be used, with a parity-specific binomial expansion, to estimate the proportion of women expected to experience no early child deaths. While better than no correction at all, this approach entails assuming that all women of a given parity experience the same underlying probability of death for each of their births, a questionable assumption. A somewhat better procedure is to estimate the degree of under-reporting by comparing reproductive histories that are based on multiple sources (the woman herself, her husband, co-wives or others) and thus likely to be more reliable, with what the reproductive histories from the woman alone would have indicated, or with single-source reproductive histories in general (Leslie et al. 1999a).

There are other means of correcting for under-reporting of deaths, such as the reverse survival technique (Shryock and Seigel 1976). That technique estimates the number of births x years ago based on the existing age-sex distribution and the age-specific mortality rate that could have produced it. These techniques can markedly improve the picture of what's going on at the aggregate, population level but do not correct individual reproductive histories, our primary concern in this report.

Imputing data.

Imputation methods - replacing missing data with suitable estimates - are used to compensate for missing values. In the context of reproductive histories, dates are frequently missing, and imputing dates can prove quite useful. Commonly, some but not all births in a reproductive history are known. The dates can then be imputed as follows: Spread the n births over a given span (e.g., from a known marriage or first birth date to the present or to a better-established date of another birth). For example, if the dates of one or more births are unknown but fall between two known birth dates in the woman's reproductive history, the missing dates can be estimated as falling evenly spaced between the known "bookends". This allows the inclusion of such reproductive histories in calculations of mean birth intervals, as imputing in this manner will not bias the mean, and it is not likely to seriously

affect estimation of a number of other measures, such as age-specific fertility rates. However, it will artificially reduce the variance of intervals, and thus may affect statistical comparisons of populations or subpopulations. Early and Peters (1990) used this approach along with fertility histories to estimate missing birth dates for Mucajai Yanomama, for whom some birth dates were known from missionaries' records. The notable regularities in Yanomama women's fertility histories (reflected in small standard errors of mean age at first birth, age specific fertility, and birth spacing determined from a sample where dates were known) led to greater confidence in the imputed ages.

Imputation can be viewed as a simulation technique (using a mathematical model of reality), and it can be combined with other simulation techniques. In some cases, it may be preferable to impute dates by a Monte Carlo procedure -- sampling randomly from an appropriate distribution rather than using the mean duration between events (e.g., between two births or between marriage and first birth). This can maintain a reasonable picture of variation at the population level (especially if the procedure is re-run multiple times), but as with straightforward imputing, care should be taken not to reify the simulated dates for use in analyses that focus on relationships between events in individual reproductive histories -- e.g., the relationship between birth interval and survival of the previously born child.

Alternative models for population dynamics and parameter estimation

A number of mathematical techniques can be used to improve estimates of demographic parameters derived from reproductive histories. For example, Gage (Gage 2000; Gage 2001) applied a method of assessing population dynamics (growth and change in age/sex composition) based on birth intervals and parity progression ratios (the proportion of women with a certain number of children who go on to have another) developed by Feeney (Feeney and Feng 1993) using Costa Rican fertility data with poor age estimation. Brass developed a number of techniques for estimating fertility and mortality rates in the face of limited or defective data, and some of these can be used for assessing and correcting underreporting of births and deaths (Brass 1953; Brass 1954; Brass 1975; Brass and Macrae 1984). However, the validity of these procedures depends to varying degrees on population size, whether overall fertility is sufficiently high, and whether or not fertility has changed over time (Leslie and Gage 1989). Caswell developed a technique for calculating variability in lifetime reproductive output using data collected at one time rather than throughout the reproductive career (Caswell 2011). Some of these techniques can provide estimates of population-level rates or characteristics (such as birth intervals or nuptiality rates) and can do so even when ages are not known. For some purposes they can supplement and compensate for inadequate reproductive histories, but they do not provide means for improving the reproductive histories themselves, which is our primary concern here. Thus, we simply point out their potential utility.

Methodological opportunism

Finally, we note that familiarity with a culture or community may reveal idiosyncratic features of languages, kinship systems, or cultural practices that can help solve problems such as age estimation. Here are a few examples.

- Sibling birth order can be determined by the kinship terms used by Yanomama siblings to address one another (Early and Peters 2000).
- Kipsigis women can specify when they went through the clitoridectomy ceremony, which occurs within a restricted age range (Borgerhoff Mulder 1989).
- Turkana men remember the year in which they experienced adrenarche (Campbell et al. 2005).
- The Maasai age grade system not only identifies all men as belonging to a named age set, the members of which are of a known broad age range (roughly 14 years, though subdivisions of the age sets can narrow that range) but the ceremonies that mark the initiation of those age sets into the warrior age grade and later “promotion” to elder status are well known by all. The dates of these ceremonies can serve as markers for events or conditions relevant to reproductive histories. Thus, questions such as “Was your first child born before the Korianga age set’s initiation?” elicit reliable answers.
- Indirect evidence can provide important clues. For example, billboards advertising family planning leads to the expectation of reports of the use of birth control. If respondents do not report any use then perhaps they are withholding information or the question is misunderstood.

It is worth watching for such peculiarities. Their precision, their attendant sources of error, and therefore their utility and appropriateness, vary markedly with the study population, the specific problem at hand, and the time and resources available for fieldwork. They may be useful only in restricted circumstances, but awareness of the possibility of such “opportunistic” approaches may lead to means of improving reproductive histories.

SUMMARY AND CONCLUSIONS

Reproductive histories are important to a variety of topics and problems of interest to biological anthropologists. The causes of incompleteness or bias in collected histories vary with the context in which they were collected. Their consequences vary with their intended use. Some means of handling the problems encountered in documenting reproductive histories, such as the use of event calendars and corroborating informants, are fairly well known and widely applicable. Others, such as the use of developmental markers, deserve wider recognition because they are potentially widely applicable. Still others are *ad hoc* and situation-specific. Using multiple sources of information to arrive at internal consistency and plausibility is a useful guideline, but some women have unusual lives and there will be outliers.

A paper such as this one cannot cover all the possibilities, which apply to varying extents in diverse settings. Many deal with emic issues (understanding from the individual’s point of view rather than the researcher’s) and are culture specific. Those discussed here illustrate the larger issues of good data collection and can lead to implementing solutions where problems are found. We hope that this overview causes readers to consider carefully the nuances of data collection in their particular studies. The simple checklist in Table 1 may prove useful

as a ‘to do’ and as a ‘confirm done’ to prompt planning and reporting when tailored to particular research questions and settings.

All techniques benefit from excellent cultural knowledge, knowledge of local language, interviewing skills, and the time and patience to check and recheck data. Describing the data collection process, and efforts to increase reliability and accuracy, are essential to continued improvement in technique and scientific communication which will reduce measurement error and raise confidence in the reproductive history data collected in communities.

ACKNOWLEDGEMENTS

We thank Sarah Miller-Fellows for assistance in locating references and Mel Goldstein (both of Case Western Reserve University) for reviewing manuscript drafts. CB thanks Geoff Childs (Washington University) and Sienna Craig (Dartmouth College) for productive collaborations to collect reliable reproductive histories.

LITERATURE CITED

- Barclay GW. 1958 *Techniques of Population Analysis*. New York: John Wiley and Sons, Inc.
- Basu AM, and Aaby P, editors. 1998 *The Methods and Uses of Anthropological Demography*. Oxford: Clarendon Press.
- Bland JM, and Altman DG. 1986 Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet* 1(2 8):307–310. [PubMed: 2868172]
- Bollig M 2006 Risk Management in a Hazardous Environment A Comparative Study of Two Pastoral Societies. New York: Springer.
- Bongaarts J 1978 A framework for analyzing the proximate determinants of fertility. *Population and Development Review* 4:105–132.
- Bongaarts J 1987 The proximate determinants of fertility. *Technology in Society* 9(3–4):243–260.
- Bongaarts J, and Potter RG. 1983 *Fertility, biology, and behavior: an analysis of the proximate determinants*. New York: Academic Press.
- Borgerhoff Mulder M 1989 Menarche, menopause and reproduction in the Kipsigis of Kenya. *Journal of biosocial science* 21(2):179–192. [PubMed: 2722914]
- Borgerhoff Mulder M 2001 Optimizing offspring: the quantity-quality tradeoff in agropastoral Kipsigis. *Evolution and human behavior : official journal of the Human Behavior and Evolution Society* 21:391–410.
- Bosetti C, Tavani A, Negri E, Trichopoulos D, and La Vecchia C. 2001 Reliability of data on medical conditions, menstrual and reproductive history provided by hospital controls. *Journal of Clinical Epidemiology* 54(9):902–906. [PubMed: 11520649]
- Brass W 1953 The derivation of fertility and reproduction rates from restricted data on reproductive histories. *Population Studies* 7:137–166.
- Brass W 1954 The estimation of fertility rates from ratios of total to first births. *Population Studies* 7:137–166.
- Brass W 1975 *Methods of Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill, NC: Laboratory for Population Statistics, Carolina Population Center.
- Brass W, and Macrae S. 1984 Childhood mortality estimated from reports on previous births given by mothers at the time of a maternity: I. Preceding-births technique. *Asian and Pacific Forum* 11:5–8.
- Campbell B, Leslie P, Little M, and Campbell K. 2005 Pubertal timing, hormones, and body composition among adolescent Turkana males. *Am J Phys Anth* 128:896–905.
- Campbell K, and Wood J. 1988 Fertility in traditional societies In: P D MP, and S T, editors. *Natural Human Fertility: Social and Biological Determinants*. London: Macmillan p 39–69.
- Caswell H 2011 Beyond R0: demographic models for variability of lifetime reproductive output. *PLoS one* 6(6):e20809.

- Childs G, Goldstein MC, Jiao B, and Beall CM. 2005 Tibetan Fertility Transitions in China and South Asia. *Population And Development Review* 31(2):337–349.
- Cleland J 1996 Demographic data collection in less developed countries 1946–1996. *Population Studies* 50(3):433–450. [PubMed: 11618376]
- Coale AJ, and Demeny P. 1966 Regional model life table and stable populations. Princeton, N. J: Princeton University Press.
- Coast E 2000 Maasai Demography. Doctoral Dissertation in Anthropology. London: University College London.
- Corsi DJ, Neuman M, Finlay JE, and Subramanian SV. 2012 Demographic and health surveys: a profile. *International Journal of Epidemiology* 41(6):1602–1613. [PubMed: 23148108]
- Davis K, and Blake J. 1956 Social Structure and Fertility: An Analytic Framework. *Economic Development and Cultural Change* 4(3):211–235.
- Early JD, and Peters JF. 2000 The Xilixana Yanomami of the Amazon: History, Social Structure, and Population Dynamics. Gainesville: University Press of Florida.
- Ewbank DC. 1981 Age Misreporting and Age-Selective Underenumeration In: Demography CoPa, editor. Washington, D.C.: NAS/NRC.
- Feeney G, and Feng W. 1993 Parity progression and birth intervals in China: The influence of policy in hastening fertility decline. *Population and Development Review* 19(1):61–101.
- Fernald LC, Kariger P, Hidrobo M, and Gertler PJ. 2012 Socioeconomic gradients in child development in very young children: evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences* 109 Suppl 2:17273–17280.
- Filippi V, Ronsmans C, Gandaho T, Graham W, Alihonou E, and Santos P. 2000 Women's reports of severe (near-miss) obstetric complications in Benin. *Studies in family planning* 31(4):309–324. [PubMed: 11198068]
- Firschein IL. 1961 Population Dynamics of the Sickle-Cell Trait in the Black Caribs of British Honduras, Central America. *American journal of human genetics* 13:233–254. [PubMed: 13699768]
- Gage T 2000 Alternative theories of population dynamics: An example using birth intervals and parity progression. *American Journal of Human Biology* 12:274.
- Gage TB. 2001 The age-specific fecundity of mammalian populations: A test of three mathematical models. *Zoo Biology* 20:487–499.
- Garenne M 1994 Do women forget their births? A study of maternity histories in a rural area of Senegal (Niakhar). *Population bulletin of the United Nations*(36):43–54. [PubMed: 12345758]
- Garenne M, and Van Ginneken J. 1994 Comparison of retrospective surveys with a longitudinal follow-up in Senegal: SFS, DHS and Niakhar. *European journal of population = Revue europeenne de demographie* 10(3):203–221. [PubMed: 12288310]
- Goldstein MC, Ben Jiao, Beall CM, and Phuntsog Tsering. 2002 Fertility and Family Planning in Rural Tibet. *The China Journal* 47:19–39.
- Gurven M, Kaplan H, and Supa AZ. 2007 Mortality experience of Tsimane Amerindians of Bolivia: regional variation and temporal trends. *American journal of human biology : the official journal of the Human Biology Council* 19(3):376–398. [PubMed: 17421012]
- Hawkes K 2003 Grandmothers and the evolution of human longevity. *American journal of human biology : the official journal of the Human Biology Council* 15:380–400. [PubMed: 12704714]
- Hill KR, and Hurtado AM. 1996 Aché Life History: The Ecology and Demography of a Foraging People. New York: Aldine de Gruyter 596 p.
- Howell N 2000 Demography of the Dobe !Kung. New York: Aldine de Gruyter.
- Howell N 1979 Demography of the Dobe !Kung New York: Academic Press.
- ICF International. 2012 Measure DHS: Topics: Fertility and Fertility Preferences. Calverton, MD: ICF International.
- Joffe M, Villard L, Li Z, Plowman R, and Vessey M. 1995 A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England. *Journal of Epidemiology & Community Health* 49(3):314–319. [PubMed: 7629471]

- Johnson PL. 1987 The Effects of “Cultural Filters” in collecting Demographic Data In: McDevitt T, editor. *The Survey Under Difficult Conditions Vol 2* New Haven: Human Relations Area File Press p 146–161.
- Julian CG, Wilson MJ, and Moore LG. 2009 Evolutionary adaptation to high altitude: a view from in utero. *American journal of human biology : the official journal of the Human Biology Council* 21(5):614–622. [PubMed: 19367578]
- Kjetland EF, Kurewa EN, Mduluzi T, Midzi N, Gomo E, Friis H, Gundersen SG, and Ndhlovu PD. 2010 The first community-based report on the effect of genital *Schistosoma haematobium* infection on female fertility. *Fertil Steril* 94:1551–1553. [PubMed: 20149365]
- Koblinsky M, Chowdhury ME, Moran A, and Ronsmans C. 2012 Maternal morbidity and disability and their consequences: neglected agenda in maternal health. “*Journal of Health, Population, & Nutrition*” 30(2):124–130.
- Konner M, and Worthman C. 1980 Nursing frequency, gonadal function, and birth spacing among ! Kung hunter-gatherers. *American Association for the Advancement of Science* 207(4432): 788–791.
- Kosova G, Abney M, and Ober C. 2010 Colloquium papers: Heritability of reproductive fitness traits in a human population. *Proceedings of the National Academy of Sciences* 107 Suppl 1:1772–1778.
- Leslie P, and Dyson-Hudson R. 1999 People and Herds In: Little MA, and Leslie PW, editors. *Turkana Herders of the Dry Savanna Ecology and Biobehavioral Response of Nomads to an Uncertain Environment*. Oxford: Oxford University Press p 233–247.
- Leslie P, Dyson-Hudson R, and Fry P. 1999a Population Replacement and Persistence In: Little MA, and PW L, editors. *Turkana Herders of the Dry Savanna Ecology and Biobehavioral Response of Nomads to an Uncertain Environment* Oxford: Oxford University Press p 281–301.
- Leslie P, Dyson-Hudson R, Lowoto E, and Munyesi J. 1999b Ngisonyoka Turkana Event Calendar In: Little MA, and PW L, editors. *Turkana Herders of the Dry Savanna Ecology and Biobehavioral Response of Nomads to an Uncertain Environment*, . Oxford: Oxford University Press p375–378.
- Leslie P, and Gage T. 1989 Demography and human population biology: problems and progress In: Little M, and Haas J, editors. *Human Population Biology: A Transdisciplinary Science*. Oxford: Oxford University Press p 15–44.
- Leslie PW, Fry P, Galvin K, and McCabe J. 1988 Biological, behavioral, and ecological influences on fertility in Turkana pastoralists In: Whitehead E, Hutchinson C, Timmermann B, and Varady R, editors. *Arid Lands: Today and Tomorrow*. Boulder: Westview Press p 705–711.
- Levine RA, and Scrimshaw SC. 1983 Effects of culture on fertility: anthropological contributions In: Bulatao R, editor. *Determinants of Fertility in Developing Countries Vol 2 Fertility Regulation and Institutional Influences*. New York: Academic Press, p 666–695.
- Lin SS, Glaser SL, and Stewart SL. 2002 Reliability of self-reported reproductive factors and childhood social class indicators in a case-control study in women. *Annals of Epidemiology* 12(4): 242–247. [PubMed: 11988412]
- Martin MC, Block JE, Sanchez SD, Arnaud CD, and Beyene Y. 1993 Menopause without symptoms: the endocrinology of menopause among rural Mayan Indians. *American Journal of Obstetrics & Gynecology* 168(6 Pt 1):1839–1835. [PubMed: 8317530]
- Meekers D 1992 The Process of Marriage in African Societies: A Multiple Indicator Approach. *Population and Development Review* 18(1):61–78.
- Merdad L, Hill K, and Graham W. 2013 Improving the Measurement of Maternal Mortality: The Sisterhood Method Revisited. *PloS one* 8(4):e59834.
- Mesoudi A, and Laland KN. 2007 Culturally transmitted paternity beliefs and the evolution of human mating behaviour. *Proceedings of the Royal Society B: Biological Sciences* 274(1615):1273–1278.
- Moreau C, Bherer C, Vezina H, Jomphe M, Labuda D, and Excoffier L. 2011 Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science (New York, NY)* 334:1148–1150.
- O’Connor KA, Holman DJ, and Wood JW. 1998 Declining fecundity and ovarian ageing in natural fertility populations. *Maturitas* 30(2):127–136. [PubMed: 9871907]

- Ober C, Hyslop T, and Hauck WW. 1999 Inbreeding effects on fertility in humans: evidence for reproductive compensation. *American journal of human genetics* 64(1):225–231. [PubMed: 9915962]
- Ober CL, Martin AO, Simpson JL, Hauck WW, Amos DB, Kostyu DD, Fotino M, and Allen FH Jr., 1983 Shared HLA antigens and reproductive performance among Hutterites. *American journal of human genetics* 35(5):994–1004. [PubMed: 6577788]
- Pennington R, and Harpending H. 1993 *The Structure of an African Pastoralist Community: Demography, History and Ecology of the Ngamiland Herero*. Oxford, UK: Oxford University Press.
- Pervin J, Moran A, Rahman M, Razzaque A, Sibley L, Streatfield PK, Reichenbach LJ, Koblinsky M, Hruschka D, and Rahman A. 2012 Association of antenatal care with facility delivery and perinatal survival - a population-based study in Bangladesh. *BMC Pregnancy Childbirth* 12:111. [PubMed: 23066832]
- Prado E, Abubaka A, Abbeddou S, Jimenez E, Some J, Ouédraogo J, and Members of the i-zinc study team. 2013 Extending the Developmental Milestones Checklist for use in a different context in Sub-Saharan Africa. *Extending the Developmental Milestones Checklist for use in a different context in Sub-Saharan Africa*. *Acta Paediatrica*:n/a.
- Rohan TE, Record SJ, and Cook MG. 1988 Repeatability of interview-derived socio-demographic and medical information. *Journal of Clinical Epidemiology* 41(8):763–770. [PubMed: 3418365]
- Shenk MK, Towner MC, Kress HC, and Alam N. 2013 A model comparison approach shows stronger support for economic models of fertility decline. *Proc Natl Acad Sci U S A* 110(20):8045–8050. [PubMed: 23630293]
- Shryock HS, and Seigel JS. 1976 *Methods and Materials of Demography*. New York: Academic Press.
- Siegel JS, and Swanson DA, editors. 2004 *The Methods and Materials of Demography*. 2nd edition 2 ed. San Diego, CA: Elsevier Academic Press.
- Slanger T, Mutschelknauss E, Kropp S, Braendle W, Flesch-Janys D, and Chang-Claude J. 2007 Test-retest reliability of self-reported reproductive and lifestyle data in the context of a German case-control study on breast cancer and postmenopausal hormone therapy. *Annals of Epidemiology* 17(12):993–998. [PubMed: 17855123]
- Strassmann BI, and Garrard WM. 2011 Alternatives to the grandmother hypothesis: a meta-analysis of the association between grandparental and grandchild survival in patrilineal populations. *Human nature (Hawthorne, NY)* 22(1–2):201–222.
- Strassmann BI, and Gillespie B. 2002 Life-history theory, fertility and reproductive success in humans. *Proceedings Biological sciences / The Royal Society* 269(1491):553–562.
- Strassmann BI, and Gillespie B. 2003 How to measure reproductive success? *American Journal of Human Biology* 15(3):361–369. [PubMed: 12704712]
- Strassmann BI, Kurapati NT, Hug BF, Burke EE, Gillespie BW, Karafet TM, and Hammer MF. 2012 Religion as a means to assure paternity. *Proceedings of the National Academy of Sciences* 109(25):9781–9785.
- Tomeo CA, Rich-Edwards JW, Michels KB, Berkey CS, Hunter DJ, Frazier AL, Willett WC, and Buka SL. 1999 Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology (Cambridge, Mass)* 10(6):774–777.
- Tonelli LA, Markowicz KR, Anderson MB, Green DJ, Herrin GL, Cotton RW, Dykes DD, and Garner DD. 1990 Use of deoxyribonucleic acid (DNA) fingerprints for identity determination: comparison with traditional paternity testing methods--Part I. *Journal of Forensic Sciences* 35(6):1265–1269. [PubMed: 2262765]
- Townsend N, and Hammel EA. 1990 Age estimation from the number of teeth erupted in young children: an aid to demographic surveys. *Demography* 27(1):165–174. [PubMed: 2303137]
- Tracer D, Wyckoff S, Wimmer M, and Gardner S. 2000 Prone to crawl: Cultural contingency and early life locomotor development. *Amer J Human Biology* 12(2):278.
- Turke PW. 1988 *Helpers at the nest: childcare networks on Ifaluk In*: Betzig L, Borgerhoff Mulder M, and Burke P, editors. *Human Reproductive Behavior: A Darwinian Perspective*. Cambridge: Cambridge University Press.

- Van de Vijver F, and Tanzer NK. 1997 Bias and equivalence in cross-cultural assessment. *European Review of Applied Psychology* 47:263–279.
- Vitzthum VJ, Spielvogel H, Thornburg J, and West B. 2006 A prospective study of early pregnancy loss in humans. *Fertility and Sterility* 86(2):373–379. [PubMed: 16806213]
- Vitzthum VJ, and Wiley AS. 2003 The proximate determinants of fertility in populations exposed to chronic hypoxia. *High Alt Med Biol* 4(2):125–139. [PubMed: 12855047]
- Wang X, Byars SG, and Stearns SC. 2014 Genetic links between post-reproductive lifespan and family size in Framingham. *Evolution, medicine, and public health* 2013:241–253.
- WHO Department of Reproductive Health and Research. 1997 The sisterhood method for estimating maternal mortality. Geneva and New York: WHO.
- Wiley A, and Pike I. 1999 A test of a model using developmental stage instead of chronological age for collecting and analyzing data on infant and child mortality. *Amer J Human Biology* 11(2):133.
- Wiley AS. 2004 *An Ecology of High-Altitude Infancy A Biocultural Perspective*. Cambridge, UK: Cambridge University Press.
- Wiley AS, and Pike IL. 1998 An alternative method for assessing early mortality in contemporary populations. *American Journal of Physical Anthropology* 107(3):315–330. [PubMed: 9821496]
- Winking J, Stieglitz J, Kurten J, Kaplan H, and Gurven M. 2013 Polygyny among the Tsimane of Bolivia: an improved method for testing the polygyny-fertility hypothesis. *Proceedings of the Royal Society B: Biological Sciences* 280(1756):20123078.
- Wood JW. 1994 *Dynamics of human reproduction : biology, biometry, demography*. New York: Aldine de Gruyter.

Table 1.

Issues to be aware of in designing and carrying out collection of reproductive histories. Solutions and best practices are discussed in the text. Ideally, the means of engaging many of these considerations should be reported in publications.

Data needed

- ☐ Specification of data needed (for immediate research questions and in future)

Sample selection

- ☐ Ascertainment
- ☐ Timing
- ☐ Temporal heterogeneity
- ☐ Censoring and other bias

Enhancing reliability of reproductive histories within sample: preliminary considerations

- ☐ Whom to interview
- ☐ Appropriateness of questions/topics (context and individual-specific)
- ☐ Quantity/quality tradeoffs in data sought
- ☐ Choice of interview setting
- ☐ Language choice
- ☐ Calendars in use
- ☐ Dating vital events
- ☐ Research assistant selection and training
- ☐ Question construction and testing
- ☐ Relationship with study community
- ☐ Use of unexpected or idiosyncratic opportunities

Enhancing reliability of reproductive histories: mid-stream and post-survey

- ☐ Validating data
 - ☐ Cross-checking and follow-up
 - ☐ Means of correcting or compensating for missing data
-