DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 40, ARTICLE 23, PAGES 627–656
PUBLISHED 19 MARCH 2019
https://www.demographic-research.org/Volumes/Vol40/23/
DOI: 10.4054/DemRes.2019.40.23

Research Article

Blood is thicker than bloodshed: A genealogical approach to reconstruct populations after armed conflicts

Diego Alburez-Gutierrez

Contents

Demographic Research: Volume 40, Article 23
Research Article

Blood is thicker than bloodshed: A genealogical approach to reconstruct populations after armed conflicts

Diego Alburez-Gutierrez[1]

Abstract

BACKGROUND
Conducting demographic research on armed conflicts is challenging because conflicts can affect the quality of primary data collection or halt the production of evidence altogether. Most of the existing methods for collecting time-variant demographic data cannot be used with war-affected populations.

OBJECTIVE
This paper introduces the Extended Genealogy Method (EGM), a toolkit for collecting high-quality data for demographic analysis using cross-checked extended genealogies. The paper assesses the quality of the data produced by the EGM by focusing on an empirical application of the method.

METHODS
The EGM uses chain-referral sampling to create extended kinship networks that include all members of a local population. The multiple reporting that results from the sampling strategy is used to reduce the error associated with retrospective reporting.

RESULTS
Data on 3,566 unique individuals and 1,986 marriages were collected from 100 EGM interviews. The paper shows how the EGM-generated data was used to reconstruct the excess mortality from the 1982 Río Negro Massacres in Guatemala, a wartime event that produced very high excess mortality in the population.

CONCLUSIONS
The EGM produced reliable and complete demographic data. The sampling and data processing strategies addressed retrospective and selection bias and helped evaluate data completion. The EGM can be used to reconstruct the demographic history of other local populations.

[1] London School of Economics and Political Science, London, UK and Max Planck Institute for Demographic Research, Rostock, Germany. Email: alburezgutierrez@demogr.mpg.de.

CONTRIBUTION

The method can be applied to reconstruct demographic dynamics in contexts of data scarcity, such as during and after armed conflicts. The EGM produces time-variant social network data useful to study, for example, intergenerational support or the transmission of demographic traits and behaviours over time.

1. Introduction

Our understanding of the demographic consequences of conflict has been constrained by the available data. It is usually not a priority to collect demographic data in the context of armed conflicts using traditional methods such as household surveys (Hill 2004). As a result, data collected before or after the crisis is often used for studying the demographic changes brought about by war. Household survey data (Agadjanian and Prata 2002) and census data (Neupert and Prum 2005) have both been used to study the effects of conflict. However, conflict-affected areas tend to be underrepresented in nationally representative surveys. Mali's northern regions, for example, were not sampled in the 2012–2013 Demographic and Health Survey (DHS) due to access and security concerns (CPS/SSDSPF 2014). This problem was faced by surveys in Nigeria, Sri Lanka, Egypt, and Colombia (Brück et al. 2013). Researchers analysing survey data from conflict-affected regions do not always report these sampling limitations nor consider how they affect their analyses.

In this paper, I propose a data collection method that helps fill this gap and improve our understanding of the effects of mortality crises on population and kinship dynamics. The paper has two objectives. The first is to present an original approach for collecting demographic data in settings of armed conflicts using primary genealogical data. This includes showing how the proposed methodology differs from the existing approaches in anthropology and genetics. I discuss how this method integrates qualitative tools to enhance the collection of quantitative data. Second, I show that the proposed data collection method can be used to produce historical data for the demographic reconstruction of local populations in post-conflict settings. The paper describes how the method was applied to reconstruct excess mortality after a series of mass killings affecting an indigenous population in Guatemala. I evaluate the quality of the evidence produced by this method and its potential for conducting demographic research on small populations for which no other data is available. This is of importance given that time-variant data on conflict-affected population is not commonly available.

1.1 Data sources in the demography of conflict

Most quantitative analyses of armed conflict rely on national registration data, censuses, or surveys. Each of these has drawbacks. Armed conflicts cause statistical offices to discontinue data collection or to produce unreliable evidence (Hill 2004). Censuses are carried out with long intervals, the length of which can be extended by the conflict. Even when census data or national registration data is available, they can be inaccurate – sometimes deliberately so for political goals (Morland 2014). The lack of updated demographic data also affects the accuracy of model-based studies, since population modelling uses demographic parameters as inputs (Alburez-Gutierrez and Segura 2018).

Demographic studies usually analyse household surveys that were conducted before the conflict or at some point after its termination. Some use nationally representative samples (Agadjanian, Dommaraju, and Glick 2008; de Walque 2005; de Walque and Verwimp 2010), but others do not (Heuveline and Poch 2007; Randall 2005; Verwimp 2003). Low-quality baseline data affects the sample design of these surveys. Choosing primary sampling units with probability proportional to size, for example, requires basic population data (Himelein et al. 2016). In addition to this, surveys face security concerns, administrative complexity, and high costs.

Data on conflict-affected populations can also be obtained from qualitative research, which often performs in-depth studies of smaller samples. This approach can be useful to suggest mechanisms through which conflict affects demographic behaviour, but few studies in the demography of conflict have a strong qualitative component (Lubkemann 2002, 2005; Randall 2005). Qualitative data is usually collected to complement quantitative studies, to explore the participants' own perspectives, or to clarify how concepts are understood locally. Heuveline and Poch (2007: 411), for example, accompanied their quantitative analysis of marriage and birth histories with "focus group discussions with different cohorts of women on marriage formation and fertility preferences." Singh et al. (2005) used qualitative research to clarify what refugees from the Sudanese civil war understood to be 'home' when answering a survey.

The lack of longitudinal or time-variant data on war-affected populations is a serious limitation. It is difficult to follow up individuals as they experience violent events or forced displacement. Very few studies have done this. The Chitwan Valley Family Study collected monthly panel data on demographic events in the eponymous Nepalese valley during the 1996–2006 civil war (Axinn, Ghimire, and Williams 2012). The Study of the Tsunami Aftermath and Recovery conducted a longitudinal survey that followed up 30,000 individuals in 10,000 households before and after the 2004 Indian Ocean earthquake and tsunami. These two projects have produced a wealth of information on demographic change (Axinn 2015; Frankenberg et al. 2011),

displacement and migration (Gray et al. 2014; Williams 2015), and the role of community and social organisations in dealing with disaster (Nobles, Frankenberg, and Thomas 2015; Williams 2013).

1.2 Genealogical data for demographic research

There is currently no systematic approach for collecting genealogical data for demographic analysis. Anthropologists have a long tradition of collecting genealogies, but whereas the initial studies of kinship in social anthropology were highly technical (Lévi-Strauss 1969), this approach has been increasingly replaced by a critical stance that seeks to challenge traditional assumptions on gender and family rather than focus on the accuracy of the genealogical data (Kuper 2003; McKinnon 2000). For this reason, anthropological kinship data rarely meets the high-quality standards required for accurate demographic analysis. Unmarried women, marriages without children, and early deaths tend to be underreported in anthropological accounts and historical data (Castilla and Adams 1996; Gamella and Martín Carrasco-Muñoz 2017).

The paper argues that researcher-generated genealogies can be used to reconstruct the demographic consequences of contemporary armed conflict on local populations. No such tool currently exists in the field of demography, where primary genealogical data usually comes from household rosters in surveys. Surveys tend to produce egocentric genealogies that cannot be merged into a unified dataset that records all kinship relationships in the population (Randall and Coast 2014). Existing approaches such as the Kinship Network Questionnaire (Heady and Kohli 2010) or the Kinship Support Tree (Madhavan et al. 2017) are appropriate for collecting egocentric kinship data, but not for generating extended genealogies.[2] Genealogical data can also be reconstructed from secondary sources. Historical population registers exist in Nordic countries (Holden and Boudko 2018; Kolk 2014), North America (Gauvin et al. 2014; Roy-Gagnon et al. 2011), and East Asia (Dong et al. 2015; Zhao 2001), but historical demographic data is usually not available for the low-income countries most affected by armed conflict after World War II (Kaplanis et al. 2018).

Outside the social sciences, population genetics has produced the most systematic and up-to-date protocols for collecting primary genealogies (Poletta, Orioli, and Castilla 2014; Williams-Blangero and Blangero 2006). However, these guidelines were designed for studying genetic inheritance, not demographic processes. Genetic genealogies ('pedigrees') usually do not record childless partnerships and tend to underreport child mortality, making them unsuitable for the study of the social

[2] Egocentric networks are centred on an individual. Extended genealogies (also called 'complete' or 'socio-centric' genealogies) contain data on all families in a population.

processes that underlie demographic change. This is a particular concern for armed conflicts, which tend to increase child mortality rates (Grein et al. 2003).

Extended genealogies have many analytical advantages. They encode the vital events of all members of a population over time, including information on members who were no longer alive when the data was collected. They also record the kinship relations between them, making it possible to derive a wide array of demographic measures and rates. Birth histories, for example, can be easily reconstructed for women and men. It is also possible to observe intergenerational changes in demographic behaviour (e.g., by comparing the fertility of grandparents to that of their grandchildren). Since genealogies are in themselves social networks (more precisely, they are kinship networks) they can be analysed using the principles of social network analysis (Borgatti et al. 2009).

2. The Extended Genealogy Method (EGM): A new approach

This section presents the basic principles of the Extended Genealogy Method (EGM). The method, which uses theory and methods from anthropology, historical genealogy, and population genetics, is an efficient approach for collecting extended genealogies that contain all related individuals that have lived and died in a given population during a specific period.

2.1 Participant selection and reliability

An extended genealogy is generated by combining multiple egocentric genealogies. Studies in population genetics often rely on a census-like approach for this. Genealogical data is collected from all households in a population and post-hoc efforts are made to merge the data and remove duplicates (Choh et al. 2001; Pinkerton et al. 2011). This procedure generally produces one big cluster of related individuals and several minor isolated genealogies, which are usually ignored for the analysis. In a classic study of the genetic epidemiology of the Chagas disease, Williams-Blangero et al. (1997) interviewed 1,758 individuals and found that 954 belonged to a single genealogy. The remaining 804 individuals were distributed amongst 254 unrelated clusters. The EGM uses a chain-referral sampling methodology for collecting and managing genealogical data in local populations.

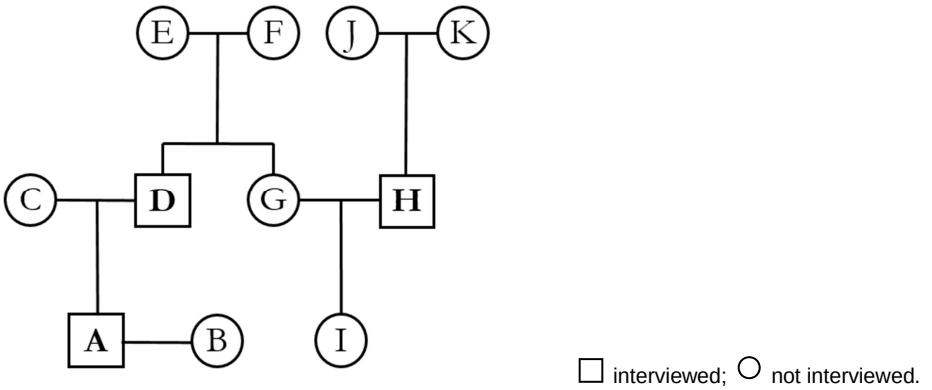
Participant selection is a central component of primary data collection. Chain-referral sampling methods have been used to collect samples from hard-to-reach populations. Respondent-driven Sampling (RDS), for example, applies network

sampling to recruit respondents from the social networks of previous participants in the study (Platt, Luthra, and Frere-Smith 2015). Generalising from these non-representative samples is challenging. Network Scale-up Methods (NSM) have been developed to estimate population-level characteristics, such as total population size (Feehan and Salganik 2016) or composition (Salganik and Heckathorn 2004) using these samples. Scale-up methods are not well suited for generalising dynamic (i.e., time-variant) processes using retrospective cross-sectional data, such as genealogical reporting (Feehan and Salganik 2016). Applying NSM in the context of armed conflict is even more challenging since the models would need to account for the largely unpredictable changes in demographic dynamics caused by armed conflicts.

The EGM interviews follow a standard structure to reduce interview length and minimise respondent fatigue. Respondents are initially asked to produce information on their parents and siblings. The partnerships and offspring of these individuals are then recorded in the interview questionnaire, including nephews and nieces. Figure 1 represents a hypothetical population reconstructed in this way from three genealogical interviews. In this example, a seed interview with individual *A* produces data on individuals *A* to *D*. A second interview with individual *D* produces data on individuals *A* to *I*. The last interview, with individual *H*, yields information on her parents, spouse, and child (*G* to *K*). The respondent for the next interview is always drawn from the pool of individuals known at the time, following a chain-referral logic.

Seven out of the eleven members of this population were reported in at least two independent interviews (individuals *A*, *B*, *C*, *D*, *G*, *H*, and *I*). 'Multiple reporting' in genealogical networks (i.e., where individuals are reported by independent sources) can be used to mitigate the effects of recall bias. Overlapping reports on the same individuals help evaluate data consistency and input missing values. They can also be used to estimate the degree of underreporting in an EGM-reconstructed population, as shown later in the paper.

Figure 1: Hypothetical genealogy demonstrating the logic of the EGM chain-referral sampling



Genealogical data is collected until all members of the population have been recorded in at least one EGM interview. I introduce the concept of 'genealogical saturation' as a criterion for terminating the data collection. The concept is based on the notion of theoretical saturation in qualitative research (Glaser and Strauss 1967), described as "a process in which the researcher continues to sample relevant cases until no new theoretical insights are being gleaned from the data" (Baker and Edwards 2012: 18). Drawing on this notion, I defined genealogical saturation as the point at which all members of the population and the kinship relationships between them have been accurately recorded using EGM interviews. Genealogical saturation is achieved when collecting new genealogical data no longer increases the number of individuals or relationships (of descent or affinity) in a population.

2.2 Getting to know the population of interest

Knowledge of local kinship dynamics and principles of social organisation is essential for making the participant selection process more efficient and for collecting data in a consistent manner. If this information is not available in the published literature, it can be generated from exploratory qualitative research. Researchers must have a clear understanding of marriage dynamics in multi-ethnic societies, particularly if local norms discourage inter-ethnic marriages. Conducting seed interviews with members of different ethnic groups can reduce biases in the ethnic composition of the EGM-generated population. The ethnic distribution of the genealogical population can later be

compared to that of the regional- or country-level population, whenever this information is available.

Background information on the population can later be used to screen for errors and inconsistencies in the data. Typographical and transcription errors, for example, can produce impossible circular relations (i.e., individuals married to themselves). Biological constraints must also be considered. It is unlikely for individuals under 13 years of age and for women over 50 to bear children. Birth intervals shorter than 9 months for the same woman are also improbable. Automated checks can be implemented to identify potential errors in the data.

A second step in the process of collecting genealogies is to determine who counts as a member of the population. This is known as the "boundary specification problem" (Marsden 2011: 371). The boundaries of the population must be defined both vertically (i.e., in terms of generational depth) and horizontally (e.g., should the relatives of an in-marrying partner be recorded?). Researchers must determine the type of relationships that the EGM will record. Families can be defined in strictly biological terms, but they can include adopted or fostered individuals, polygamous unions, or other forms of locally defined kinship relations. Clear guidelines must also be developed regarding data on migrants. Should they be followed up? It is important to keep in mind that the quality of the EGM-generated data will decrease with physical distance and time of separation, meaning that more interviews will be required to ensure that data on migrants is accurately recorded. Ultimately, all criteria must be defined considering the research question, the size of the population, and the resources available to the researchers.

2.3 Data collection tools

This section outlines the structure of the questionnaires that the EGM uses for registering genealogical data. The basic questionnaire consists of three sequential modules for recording household members, individuals, and marriages, respectively. A household was defined as the collection of individuals who regularly slept in one or more dwellings with a shared kitchen.[3] This paper refers to the three modules collectively as the 'EGM questionnaire.'[4] The EGM differs from published guidelines in genealogical research in two important ways. First, it places emphasis on recording all the children ever born and not only those who survived to a certain age or formed a union (this is standard practice in birth histories). Second, the design includes an

[3] The definition was culturally appropriate, as expressed in interviews with local experts (KI-4; KI-9).
[4] The complete EGM questionnaires used in Río Negro are included in the supplementary materials (Appendix B).

intentional degree of redundancy intended to reduce reporting error at the expense of making the interviews longer – examples are given below. Both features were introduced to improve the quality of the data.

The first component of the EGM questionnaire is the 'Household Module.' This module is included to keep a record of current household characteristics and members. It is similar to household survey rosters. It can include an 'Individual ID' field to link the members of the household to the 'Individuals Module' introduced below. The EGM questionnaire used in Río Negro only recorded the dates of birth and civil state of household members. Additional fields can be included to collect information on schooling, employment status, etc., depending on the study objectives.

The second component of the questionnaires is the 'Individuals Module.' In it, every row represents a unique individual. This module records information on all the members of the kinship network, including their unique id, name, sex, date of birth, current age, date of death, age at death, survival status, cause of death, place of birth, parity, total number of marriages, and current location. Additional information can be collected to address specific research questions.

The third component of the questionnaire is a 'Marriages Module,' in which every line represents a separate marriage between two partners. Two fields ('Partner 1' and 'Partner 2') reference the row numbers of the marriage members in the Individuals Module. As an extra precaution, interviewers can be asked to write the partner's forename next to their respective ids. All current and previous marriages must be recorded in this module. The module also collects data on the characteristics of the marriage, including type of union (e.g., marriage, cohabitation), start date and end date, reason for termination, etc. As with the previous component of the questionnaires, new fields can be added to record other information.

The EGM questionnaires can be used together with pre-existing data. Mobile phones or tablets are useful for displaying historical images, such as portraits of relatives. Record linkage techniques can be used to suggest potential matches from the existing genealogical dataset in real time. The EGM can also be integrated with qualitative data collection tools. Interesting questions or themes emerging from the genealogical interviews can be explored in subsequent in-depth interviews or focus group discussions carried out with respondents identified during the genealogical data collection. The rest of the paper exemplifies these points by showing how the EGM was used to reconstruct the demographic history of a war-affected population in Guatemala.

3. Applying the EGM after a mortality crisis

3.1 The context: Río Negro

Río Negro is a relatively isolated rural Maya Achi community in Guatemala that underwent a significant mortality crisis induced by armed conflict in the context of an ongoing fertility transition. In 1979, it was the major village on the basin of the Chixoy River in the central highlands of the country. From 1980 to 1984, in the midst of the country's civil war, the area was caught up in a spiral of violence, especially after construction works started for a state-owned hydroelectric dam that threatened to flood hundreds of acres of arable land (Einbinder 2017).

EGM-generated data showed that more than a third of Río Negro's population died between 1979 and 1983; 94% of these deaths were directly attributable to the conflict. After years of forced displacement, most of the population was resettled in a town built by the INDE Electric Company in the nearby municipal capital of Rabinal. The resettlement area, known as Pacux, was under strict military surveillance until the end of the civil war. The killings in Río Negro were identified as genocide in the report of the UN-backed Truth Commission established as a result of the signing of the Peace Accords in 1996 (CEH 1999).

3.2 Research design

This study was interested in reconstructing the demographic dynamics before, during, and after the 1982 Río Negro Massacres in Guatemala. Therefore, the population of interest for the EGM interviews was defined as all the individuals who ever lived in the community before and up to 2015. Individuals who resided in the community but had no kinship ties to the original population were excluded from the study (e.g., a researcher living in the village for a year). Members of other communities married to Río Negro residents were included, but their own families (e.g., their parents and siblings) were not of interest for the study. These principles provided clear criteria for limiting the scope of the data collection, which would otherwise have extended indefinitely. Applying them systematically made the genealogical interviews more efficient by excluding non-relevant records.

A three-stage mixed methods design was implemented to reconstruct the recent demographic history of Río Negro (1960–2015), including its experience before, during, and after the armed conflict. The first stage of the study was qualitative – involving individual and group interviews and archival research. It generated contextual information for the design of the EGM questionnaire and the sampling. This included

constructing an Event History Calendar (Axinn, Pearce, and Ghimire 1999) to help date events in relation to locally relevant landmarks, such as the building of a church or the establishment of the local maternity ward. The calendar included 276 dated events between 1898 and 2016, including war-related events, political developments, and changes in infrastructure. The final calendar only included events that could be cross-checked using independent archival or historical records. Genealogical data was collected in the second stage using the EGM questionnaire. A form of chain-referral sampling was applied to select the respondents. The last stage was qualitative and aimed at producing evidence to interpret and analyse the quantitative data. FGD were conducted after most of the genealogical data had been collected and partially processed. Preliminary quantitative analysis of the Río Negro data helped identify topics of interest for the FGDs and informed the participant selection criteria for the qualitative discussions.

Fieldwork was carried out between November 2015 and November 2016 in the municipality of Rabinal, Guatemala. Most interviews took place in the resettlement area Pacux, but short visits were made to other locations to collect information on hard-to-reach individuals. The initial three months were spent building rapport, getting acquainted with the local kinship organisation and creating the Event History Calendar. The genealogical data collection took place between January and October 2016. All the EGM questionnaires were pilot tested on the field, and two local research assistants helped with the data collection.[5]

3.3 Data collection and management

Respondents for the two initial or seed interviews were selected following a set of criteria to ensure the quality and appropriateness of the data they provided. The initial qualitative component of the study provided the information needed for the seed selection. Following a chain-referral sampling methodology, the next respondents were chosen from the pool of records produced by the previous genealogical interviews until saturation was achieved. This method of selecting seed respondents could have biased the composition of the EGM-generated data since purposefully sampled seed respondents tend to be better connected than the population average (Platt, Luthra, and Frere-Smith 2015). In a genealogical network, this means that seed respondents would tend to have more relatives than the population mean. This is a concern because it might

[5] Ethics clearance for the project was obtained from Ethics Committees at the London School of Economics and the Universidad del Valle de Guatemala. The study also had the approval of traditional indigenous authorities in Río Negro and Pacux.

bias the population towards members of more extensive kinship networks. However, analysis of the genealogical data showed that this was not the case in Río Negro.[6]

All in all, 112 genealogical interviews were scheduled in Río Negro. Of these, eight were refused, and four could not be finished given the advanced age of the respondents. The 100 completed interviews yielded information on 5,803 individuals and 1,604 marriages. Many of these were duplicates – 62% of all individuals were reported by at least two independent respondents. On average, every individual was recorded in 2.5 separate interviews (SD = 1.26). After removing the duplicate records, 3,566 unique individuals and 1,018 unique marriages remained. The genealogical data was processed in the field as they were collected and transcribed. I developed a suite of graphical user interfaces in the R language to manage the qualitative and quantitative data. R Shiny web applications (Chang et al. 2016) and record linkage techniques (Borg and Sariyar 2016) were used to input, merge, and analyse the data.

On average, each interview recorded data on 58 individuals (SD = 35.07) and 16 marriages (SD = 10.2). The median interview length was one and a half hour (range 0.5–4 hours). 59% of the interviews were answered by a single respondent, and 68% were completed in a single visit to the household. Interviews with women and couples were preferred, since they are known to provide more accurate genealogical accounts than interviews conducted exclusively with men (Poletta, Orioli, and Castilla 2014). The interviews carried out in Río Negro included women in 50% of the cases.

3.4 Data completeness and genealogical saturation

This section discusses how genealogical saturation was evaluated to determine when the data collection should be terminated. Two tests were initially conducted to assess the completeness of the EGM-generated data by comparing it to independent population registers. The EGM data was first matched by name to records in a publicly available dataset of victims of the Río Negro Massacres.[7] For the second test, the genealogical records were matched by name and date of birth to the two local censuses available for Río Negro, one conducted in 1978 and another one in 2008. Matching the EGM data with the census records provided the opportunity to assess genealogical saturation at two fixed points in time – before and after the killings.

The 1978 census was conducted by the INDE Electric Company to serve as a baseline for future compensation programmes. The data, collected between 1978 and 1981, is not entirely reliable, because political violence had already begun when the

[6] Appendix A (supplementary materials) provides a technical description of the seed selection process.
[7] A list of the identified remains exhumed by the Guatemalan Forensic Anthropology Foundation (FAFG): https://www.fafg.org/bd/b_victima.php, accessed April 25, 2018.

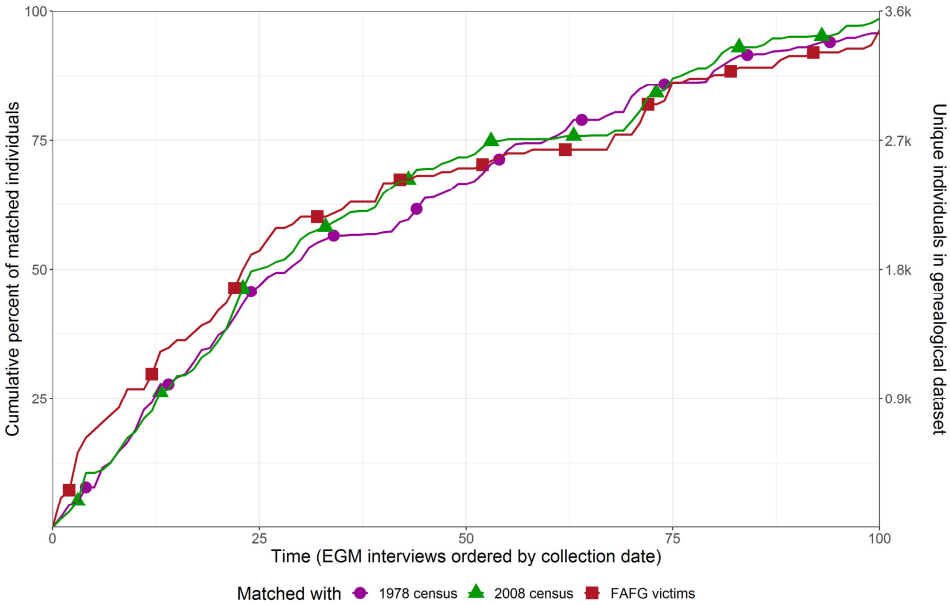
census data was collected. An analysis of the reported ages shows that they are largely rounded (Whipple's Index = 241). The 2008 census was carried out to identify the beneficiaries of the monetary war reparations provided by the Guatemalan government. Being a *de jure* census, it included non-residents as well. Dates are more reliable for this census, since individuals were required to present their national identity card upon registration. The census underreported displaced individuals and children born of rape, according to local experts interviewed for this study.[8]

These matching exercises made it possible to evaluate the progression towards genealogical saturation**.** The vertical axis in Figure 2 represents the completeness rate of the EGM-generated data matched against three independent sources. The graph shows that completeness rates approached 100% after conducting the hundredth genealogical interview. The percentage of omission (2%) is too small to have affected the final population composition significantly. The individuals who were missing from the genealogical dataset were identified as members of three households whose members were all killed during the war and had no living descendants.

The first interviews produced data on the most visible members of the population, as the sampling criteria prioritised accessible respondents. Later interviews recorded information on harder-to-reach individuals, such as internally displaced persons or community members with few living relatives. Initial interviews were more likely to report 'new' individuals who had not been reported before (50% of the census records were matched in the first 25 interviews). As the genealogical network became increasingly dense, the interviews produced more redundant records.

[8] KI-5; field notes, 27 May 2016.

Figure 2: Genealogical saturation estimated by matching EGM-generated records with list of victims and with census records

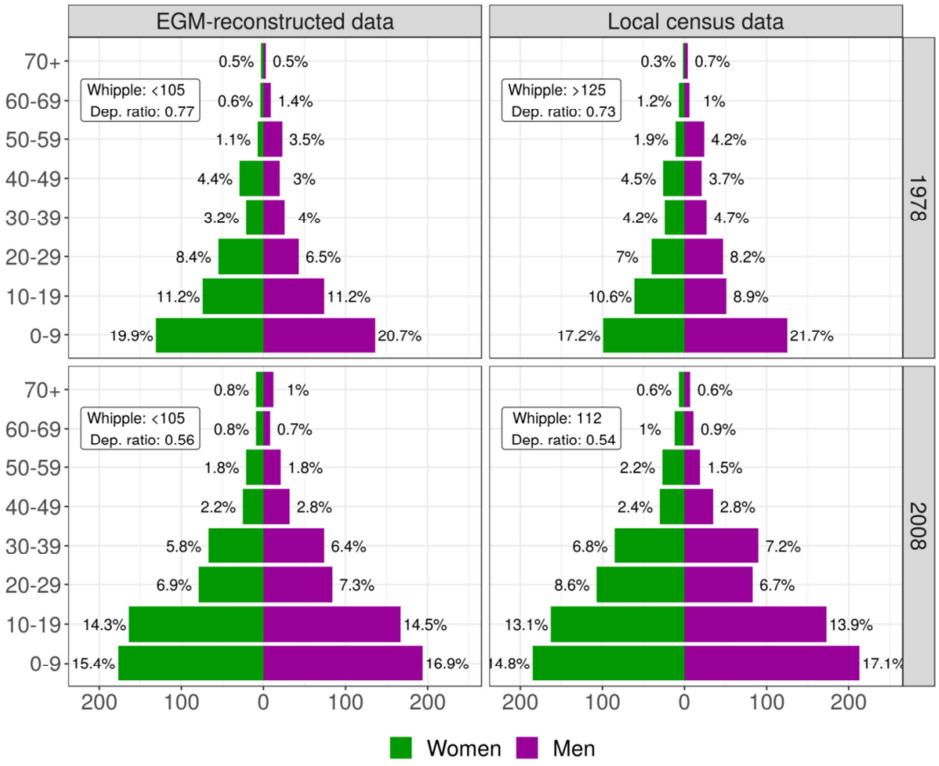


A third test compared the age and sex composition of the EGM-generated population to that registered in the two local censuses. Figure 3 suggests that the population structures were remarkably similar around the time of the killings (1978) and in recent years (2008). Statistical tests confirmed that the age distribution of the EGM-generated data was not significantly different from that of the census data, neither in 1978, χ^2 (2, N = 8) = 9.04; p = 0.25, nor in 2008, χ^2 (2, N = 8) = 10.99; p = 0.14. A comparison of age dependency ratios also confirmed the similarity of the genealogical and census data. The difference between dependency ratios was less than 3% in 1978 and 2008. [9] These results confirmed that the EGM-generated data in Río Negro was similar to that recorded in independent population registers. Furthermore, neither children nor women were systematically underreported, as commonly happens in historical genealogies.[10]

[9] The ratio represents the rate of individuals under 15 or over 64 to individuals aged between 15 and 64. In 1978, it was 0.76 in the EGM data and 0.73 in the census data. The corresponding values for 2008 were 0.56 and 0.54.

[10] Except for births resulting from rape, which could have been underreported in the EGM interviews.

Figure 3: Age-sex structure of the Río Negro population – Comparing genealogical data with census records



Source: Data for the left panels comes from this study. Details on the census data are given in the text.

3.5 Missing data

Multiple reporting helped minimise the share of missing data resulting from non-response. Independent respondents produced duplicated records on the same individuals, and these were later merged to assess data quality and reduce missing values. The share of missing data of eight key demographic variables (date of birth, date of death, place of birth, current location, times married, marital status, children ever born, and cause of death) was reduced from 25% before merging the genealogical datasets to only 11% after their consolidation. Completeness rates were high across all

individual variables, except for the date of non-violent deaths (rates of missing values = 44%). It is possible that interviewers and respondents made more effort to accurately report conflict-related deaths than other types of deaths.

I now examine four scenarios which would have produced significant bias in the EGM data: (1) data on more distant relatives was less accurate; (2) data quality was poorer for individuals who were not alive at the time of the data collection; (3) data on migrants was underreported; and (4) interview length affected reporting quality. Table 1 explores this by considering three data quality indicators: age heaping (the rounding of ages to multiples of five), the share of individuals with no reported parents, and the share of individuals with unknown dates of birth and death. This is important because most individuals whose data was included in the genealogical data were not interviewed themselves – either because they did not survive to 2015, or if surviving, they were not part of the sample of the 100 people interviewed.

Only one of the identified factors appears to have affected data quality considerably – information was less reliable for individuals who were dead at the time of the data collection. This is evidenced in the differential degree of age heaping between dead and living individuals. Social distance does not appear to have impacted data quality. Data on individuals more remotely connected to the respondent (defined using the shortest path to the respondent in a kinship network) appears to be less reliable, but the difference is not pronounced. The age of migrants was more likely to be rounded to the nearest 5 or 10 digits only for those aged 18–47.[11] Migrants were, on average, less connected to the general genealogical network, as evidenced by the fact that their parents were more likely to be unknown. Finally, there was no evidence that respondent fatigue led to poorer reporting. As a matter of fact, more lengthy interviews (i.e., reporting more individuals) were of a better quality. Older and better-informed respondents produced longer interviews with more reliable data. Interviews with couples were the most reliable; those with only male respondents tended to be less complete and accurate.[12]

[11] The results for under 18 (not shown in the table) were similar, with the highest degree of age heaping amongst children who were already dead at the time of the killings.

[12] Measured in terms of (1) share of missing data, (2) accuracy of date reporting, and (3) underreporting of known individuals.

Table 1: Data quality and completeness of EGM-generated data: An evaluation of four potential sources of bias

		Whipple index		Unknown values %		
		Ages 18–47	Ages 23–62	Parents id	Birth date	Death date
1	**Distance to respondent**					
	Closest relative – Q1	87.3	93.6	8.3	2.7	2.5
	Most distant relative – Q4	107.8	96.4	21.6	3.6	1.7
2	**Status in 2015**					
	Alive	100.2	97.8	15.1	3.8	NA
	Dead	120.2	114.1	14.1	4.7	NA
3	**Location in 2015**					
	Local	99.5	100.7	9.1	2.6	1.7
	Migrant	105.7	100.0	26.5	5.8	1.8
4	**Interview length**					
	Shortest – Q1	99.5	122.4	13.9	4.7	1.4
	Longest – Q4	102.5	88.9	15.2	4.9	1.9

3.6 Reconstructing conflict mortality using EGM-generated data

In this section, I show an example of how data from the EGM was used to reconstruct the excess mortality caused by the Río Negro Massacres. According to the genealogical data, 38% of the pre-conflict population was killed during 1982 (366 of the 970 original inhabitants of the village). Respondents in the FGDs reported that physical abuse and rape were common during these events, but these types of violence were not captured by the genealogical interviews.

Mortality was evenly distributed by gender (179 women and 187 men were killed) but not by birth cohort, as Figure 4 shows. More children under 10 were killed in absolute terms, but older males were the most affected after adjusting for population structure. The high incidence of child excess deaths suggests that mortality amongst the youngest was not underreported in the genealogical data. This is important because child excess mortality tends to be underreported in retrospective surveys. The leftmost panel confirms that women and men were killed at similar rates, even if more women in reproductive age were killed in total. This suggests that violence was directed against young women, although further research is needed to corroborate this. The denominators for the mortality rates in the second panel were obtained by conducting a 'pseudo-census' of the genealogical data to keep only the individuals alive at the start of 1982, immediately before the beginning of the mass violence.[13]

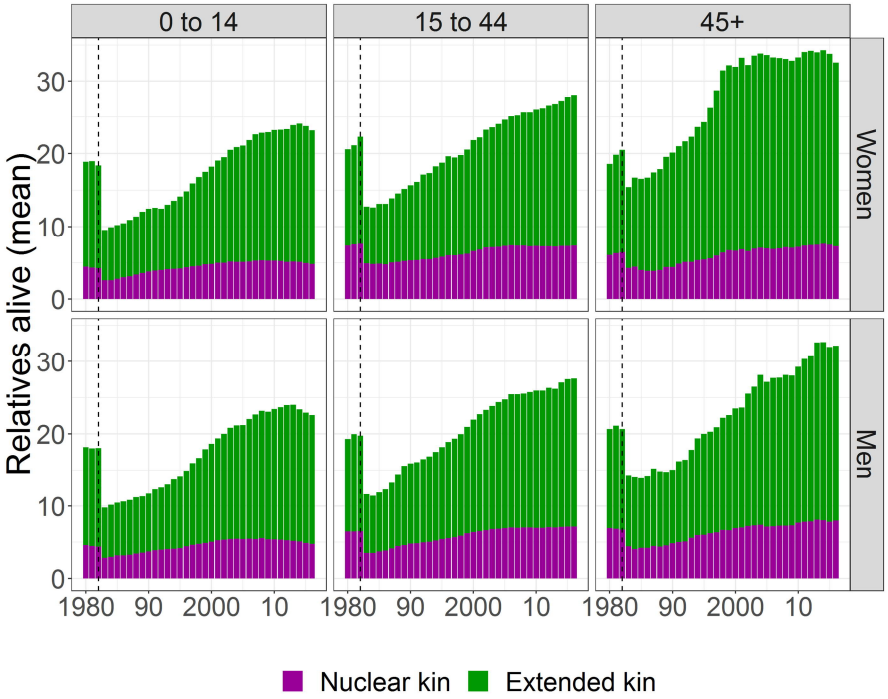
[13] Appendix A in the supplementary materials shows how these pseudo-censuses were conducted.

Figure 4: Age-sex distribution of massacre-derived mortality in Río Negro reconstructed from EGM data

Conflict-derived mortality also affected the survivors of the massacres. Data from this study showed that all the inhabitants of Río Negro lost at least one close relative during the mass killings. Figure 5 shows that women and men in all age groups saw a dramatic reduction in the size of their kinship networks because of the massacres. The mean size of the extended kin networks plummeted by 44% between 1981 and 1983 (from 19.3 to 10.8).[14] Younger individuals were more affected. Children under 15 saw the size of their kinship networks reduced by more than half in the same period (from 14.1 to 6), but adults over 45 experienced only a 25% decrease in the size of their kinship networks (from 13.8 to 10.4). These figures would be very difficult to obtain from traditional household surveys that only consider relatives living in the household (Madhavan et al. 2017).

[14] In the figure, the nuclear kin includes parents, children, siblings, and spouses; the extended kin also includes cousins, aunts, uncles, nephews, grandparents, and grandchildren.

Figure 5: Size of egocentric kinship network in Río Negro over time by sex and age group (1980–2015)



4. Discussion

4.1 A genealogical approach for reconstructing population data

The paper showed that the EGM can be used to produce quality demographic data on populations affected by armed conflicts by exploiting the network structure of kinship relations. The approach builds on previous methodologies for collecting and analysing genealogical data. It also emphasised its uniqueness: the EGM is the first approach for producing high-quality data for demographic analysis from genealogical interviews. The method was designed to address the underreporting and data quality issues affecting anthropological genealogical data (e.g., by reducing the underreporting of infant mortality using multiple reporting techniques). The EGM can be used to produce

the accurate and unbiased reporting required for demographic analysis. Strategies for improving accuracy include controlled redundancy in the questionnaire design and multiple reporting.

In the study of Río Negro, multiple reporting reduced non-response rate by 14% across all variables. The improvement was considerably higher in date of birth reporting, where the number of missing values was halved after integrating data from independent sources. Secondary data sources on the population can be used to assess the accuracy and completeness of the EGM data. Lists of victims and census data were used for the case of Río Negro. Similar data could be obtained from other sources – registries in refugee camps or records kept by local organisations, for example. Even fragmentary data can be used to compare the respondents' accounts with independent records.

The EGM is characterised by its sampling efficiency, which takes advantage of the network properties of kinship systems. The genealogical chain-referral sampling method substantially reduced the amount of data needed to reconstruct an extended genealogy, as compared to examples from the population genetics literature. A population of 3,566 was reconstructed with considerable accuracy from only 100 genealogical interviews. The main individual-level demographic characteristics of the population can be calculated from this data. A flexible research design and dynamic data management tools were essential for this, since constant feedback allowed the selection of the most appropriate respondents at a given point in time during the data collection.

A third strength of the EGM is the successful integration of qualitative tools into its design. In Río Negro, initial qualitative work on local kinship dynamics was essential for establishing an appropriate research design. Qualitative data provided information on vernacular kinship terminology, definitions of marriage, and child naming and rearing practices necessary for interpreting the genealogical data in a relevant conceptual framework. The extended genealogy only recorded biological relations of descent and of marriage or cohabitation, relying on the local definitions and kinship terminology. Lastly, the iterative nature of the research design meant that qualitative could be constantly used to improve the demographic data collection procedures.

4.2 Limitations of the EGM method

Retrospective questions are prone to recall bias. This is a limitation that affects all retrospective data collection methods. The multiple reporting built into the EGM sampling strategy helped address this source of error, but it did not eliminate it

completely. Recalling wartime violent events can be distressing, and proper training should be provided to the interviewers; psychosocial support services should be identified in advance whenever possible.

The EGM can be used to reconstruct historical demographic data. This study successfully reconstructed the last 60 years of Río Negro's demographic history. However, genealogical interviews are unlikely to produce reliable data on the distant past (e.g., demographic events that took place over a century ago). The boundaries of the historical periods that can be reconstructed are limited by the accuracy of the respondents' recollections. Furthermore, the EGM was applied in a small, isolated, and rural community. It might be less successful in urban communities or in localities with high migration (Williams-Blangero and Blangero 2006). Seed respondents must be selected carefully in multi-ethnic societies to avoid skewing the ethnic composition of the genealogical population.

Focusing on a single population sacrifices external validity for greater depth. In this sense, there is a point to be made for analytical generalisations as opposed to empirical ones (Yin 2009). The processes observed in Río Negro, for example, cannot be readily generalised to other populations. But they can help construct analytical frameworks that explain demographic behaviours in the context of other mortality crises.

The method outlined in this paper can be used to collect demographic data, including information on fertility, mortality, and marriage formation. It is not appropriate for collecting other types of information, such as longitudinal data on socioeconomic characteristics or health data. This limits the range of possible analyses. However, it is possible to link the genealogical data to other available data sources to obtain this information. The Río Negro genealogical data was linked to local censuses using record linkage techniques, which made it possible to obtain household-level socioeconomic data. Similar data sources can be available in other contexts.

A final note on ethics. 'Secondary participants' are third parties about whom information is collected, but whose consent is not sought (Marsden 2011: 384). This was the case for the eight individuals who refused to be interviewed in Río Negro. Given the network-like structure of genealogies, their information could be retrieved from a sibling or a cousin. Data was collected in this way under the assumption that kinship information was part of a shared communal knowledge.

4.3 Analysing EGM-generated social network data

Reliance on survey-generated data has encouraged individual-level analysis in demography; it has also limited the possible research questions and widened the gap

between demographic research and social theory (Bachrach 2014; Kirk 1996). Although proposals have recently been made to address this issue (Madhavan et al. 2017), alternative methods are scarce. Social network analysis (Knoke and Yang 2011) is a promising field of study that has only been partially explored in the demography of conflict (Randall 2005), even though armed conflicts are usually assumed to have negative effects on the 'social fabric' (Justino 2011: 13) or the 'social bonds' of a community (Esparza 2005: 387).

A network perspective can help incorporate notions of social norms and institutions as exogenous factors that influence demographic behaviour. This approach provides a straightforward way of quantifying support, transfer of information and many other social processes. Genealogical data is limited in that they do not consider networks of friendship or other forms of association. However, the principles of the EGM can be expanded to record them too. Data on kinship networks can be used to analyse social structure, a concept that has generally been ignored in the demographic literature. The structure of a network around an individual can be taken as an explanatory factor for that individual's behaviour. It is not enough to focus on the availability of networks of support, since the structure of the network does not reveal anything about its function (i.e., whether individuals receive support from their social networks). How effectively these networks are used is important (Cohen and Wills 1985; Rafnsson, Shankar, and Steptoe 2015). Integrating qualitative and quantitative data is key for achieving a more comprehensive understanding of these relationships.

5. Concluding remarks

In this paper, I outlined the principles of the Extended Genealogy Method (EGM), a systematic approach for collecting retrospective demographic data using extended genealogies. Anthropologists, genealogists, and geneticists have collected family histories in the past, but no approach currently exists to ensure that the data meets the high-quality standards of demographic analysis. The EGM provides a unified framework for sampling, managing, and merging overlapping genealogies. It also provides a set of criteria to evaluate data quality and completeness, which no other methodology currently does. The EGM approach relies on a dynamic and efficient sampling strategy that reduces the required sample size for reconstructing a population while maximising data accuracy. The generated data can be used to analyse demographic change at a community and an individual level using traditional demographic analysis or more novel network analysis tools.

Future studies can address different research questions by modifying the EGM questionnaires as necessary. Child fostering and adoptions can be easily studied by

adding a field to record these relationships. Kinship networks can be taken as proxies for networks of support. The EGM can be adapted to study networks of friends or other types of relations. The principles of the EGM can also be applied for reconstructing multiple, unconnected, extended genealogies. This could be useful for studying, for example, kinship networks or networks of support in displacement settings.

This paper focused on the demography of armed conflict by showing how the EGM was used to reconstruct the population of Río Negro in Guatemala after an episode of mass violence thirty years in the past. I have shown that the EGM approach can be used to answer a wide range of research questions, particularly in the context of data scarcity. The UN Refugee Agency has recently become interested in applying this approach to understand networks of support in refugee camps.[15] There is a growing interest in applying social network analysis to understand the transmission of fertility in low- and high-income countries (Bernardi and Klärner 2014; Kohler, Billari, and Ortega 2002). A major limitation of the current studies is the lack of time-variant data on conflict-affected local populations. The EGM has much to contribute to this field of inquiry.[16]

6. Acknowledgments

This research was funded by a PhD Studentship (2014-2018) of the London School of Economics and Political Science. Ernestina Coast, Arjan Gjonça, and Tiziana Leone provided helpful comments on earlier drafts. They are not responsible for any of its content. I am indebted to the people of ADIVIMA in Rabinal for the support during the data collection and to the people of Rio Negro and Pacux who took part in the study.

[15] The author has been involved in designing research proposals using this approach for the UN High Commissioner for Refugees (UNHCR).

[16] Researchers willing to apply the EGM in other settings are encouraged to consult with the author after reviewing the online supplementary materials.

References

Agadjanian, V. and Prata, N. (2002). War, peace, and fertility in Angola. *Demography* 39(2): 215–231. doi:10.1353/dem.2002.0013.

Agadjanian, V., Dommaraju, P., and Glick, J.E. (2008). Reproduction in upheaval: Ethnic-specific fertility responses to societal turbulence in Kazakhstan. *Population Studies* 62(2): 211–233. doi:10.1080/02615470802045433.

Alburez-Gutierrez, D. and Segura, C. (2018). The UNHCR Demographic Projection Tool: Estimating the future size and composition of forcibly displaced populations. Copenhagen: UNHCR (Statistics Technical Series; 2).

Axinn, W.G. (2015). Demographic change: The case of Chitwan Valley in Nepal. *International Journal of Sociology* 45(1): 1–3. doi:10.1080/00207659.2015.1004973.

Axinn, W.G., Ghimire, D., and Williams, N.E. (2012). Collecting survey data during armed conflict. *Journal of Official Statistics* 28(2): 153–171. doi:10.1006/ssre.1998.0641.

Axinn, W.G., Pearce, L.D., and Ghimire, D. (1999). Innovations in life history calendar applications. *Social Science Research* 28(3): 243–264. doi:10.1006/ssre.1998.0641.

Bachrach, C.A. (2014). Culture and demography: From reluctant bedfellows to committed partners. *Demography* 51(1): 3–25. doi:10.1007/s13524-013-0257-6.

Baker, S.E. and Edwards, R. (2012). How many qualitative interviews is enough? Southampton: National Centre for Research Methods (NCRM Review Paper).

Bernardi, L. and Klärner, A. (2014). Social networks and fertility. *Demographic Research* 30(22): 641–670. doi:10.4054/DemRes.2014.30.22.

Borg, A. and Sariyar, M. (2016). RecordLinkage: Record linkage in R: R package version 0.4–10 [electronic resouce]. https://cran.r-project.org/package=RecordLinkage.

Borgatti, S.P., Mehra, A., Brass, D.J., and Labianca, G. (2009). Network analysis in the social sciences. *Science* 323(5916): 892–896. doi:10.1126/science.1165821.

Brück, T., Justino, P., Verwimp, P., and Tedesco, A. (2013). *Measuring conflict exposure in micro-level surveys*. Stockholm: SIPRI.

Castilla, E.E. and Adams, J. (1996). Genealogical information and the structure of rural Latin-American populations: Reality and fantasy. *Human Heredity* 46(5): 241–255. doi:10.1159/000154361.

CEH (1999). Capítulo segundo: Las violaciones de los derechos humanos y los hechos de violencia. In: *Guatemala: Memoria del silencio*. Guatemala City: UNOPS – Comisión para el Esclarecimiento Histórico.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2016). shiny: Web application framework for R: R package version 0.14.2 [electronic resource]. https://cran.r-project.org/package=shiny.

Choh, A.C., Gage, T.B., McGarvey, S.T., and Comuzzie, A.G. (2001). Genetic and environmental correlations between various anthropometric and blood pressure traits among adult Samoans. *American Journal of Physical Anthropology* 115(4): 304–311. doi:10.1002/ajpa.1086.

Cohen, S. and Wills, T.A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin* 98(2): 310–357. doi:10.1037/0033-2909.98.2.310.

CPS/SSDSPF (2014). *Enquête démographique et de santé au Mali 2012–2013*. Maryland: CPS/SSDSPF.

de Walque, D. (2005). Selective mortality during the Khmer Rouge period in Cambodia. *Population and Development Review* 31(2): 351–368. doi:10.1111/j.1728-4457.2005.00069.x.

de Walque, D. and Verwimp, P. (2010). The demographic and socio-economic distribution of excess mortality during the 1994 genocide in Rwanda. *Journal of African Economies* 19(2): 141–162. doi:10.1093/jae/ejp029.

Dong, H., Campbell, C., Kurosu, S., Yang, W., and Lee, J.Z. (2015). New sources for comparative social science: Historical population panel data from East Asia. *Demography* 52(3): 1061–1088. doi:10.1007/s13524-015-0397-y.

Einbinder, N. (2017). *Dams, displacement and development: Perspectives from Rio Negro, Guatemala*. Cham: Springer.

Esparza, M. (2005). Post-war Guatemala: Long-term effects of psychological and ideological militarization of the K'iche Mayans. *Journal of Genocide Research* 7(3): 377–391. doi:10.1080/14623520500190330.

Feehan, D.M. and Salganik, M.J. (2016). Generalizing the network scale-up method: A new estimator for the size of hidden populations. *Sociological Methodology* 46(1): 153–189. doi:10.1177/0081175016665425.

Frankenberg, E., Gillespie, T., Preston, S., Sikoki, B., and Thomas, D. (2011). Mortality, the family and the Indian Ocean Tsunami. *The Economic Journal* 121(554): 162–182. doi:10.1111/j.1468-0297.2011.02446.x.

Gamella, J.F. and Martín Carrasco-Muñoz, E. (2017). The decline of infant and child mortality among Spanish Gitanos or Calé (1871–2005): A microdemographic study in Andalusia. *Demographic Research* 36(33): 945–988. doi:10.4054/DemRes.2017.36.33.

Gauvin, H., Moreau, C., Lefebvre, J.F., Laprise, C., Vézina, H., Labuda, D., and Roy-Gagnon, M.H. (2014). Genome-wide patterns of identity-by-descent sharing in the French Canadian founder population. *European Journal of Human Genetics* 22(6): 814–821. doi:10.1038/ejhg.2013.227.

Glaser, B. and Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gray, C.L., Frankenberg, E., Gillespie, T., Sumantri, C., and Thomas, D. (2014). Studying displacement after a disaster using large scale survey methods: Sumatra after the 2004 Tsunami. *Annals of the Association of American Geographers* 104(3): 594–612. doi:10.1080/00045608.2014.892351.

Grein, T., Checchi, F., Escribè, J.M., Tamrat, A., Karunakara, U., Stokes, C., Brown, V., and Legros, D. (2003). Mortality among displaced former UNITA members and their families in Angola: A retrospective cluster survey. *British Medical Journal* 327(650): 1–5. doi:10.1136/bmj.327.7416.650.

Heady, P. and Kohli, M. (2010). *Family, kinship and state in contemporary Europe: Perspectives on theory and policy*. Frankfurt am Main: Campus.

Heuveline, P. and Poch, B. (2007). The phoenix population: Demographic crisis and rebound in Cambodia. *Demography* 44(2): 405–426. doi:10.1353/dem.2007.0012.

Hill, K. (2004). *War, humanitarian crises, population displacement, and fertility: A review of evidence*. Washington, D.C.: National Academies Press.

Himelein, K., Eckman, S., Murray, S., and Bauer, J. (2016). Second-stage sampling for conflict areas: Methods and implications. Washington, D.C.: World Bank Group (Policy Research Working Paper 7617).

Holden, L. and Boudko, S. (2018). The Norwegian historic population register and migration. *Journal of Migration History* 4(2): 249–263.

Justino, P. (2011). The impact of armed civil conflict on household welfare and policy. *IDS Working Papers* 2011(384): 1–38. doi:10.1111/j.2040-0209.2011.00384_2.x.

Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovits, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D.G., Price, A.L., and Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science* 360(6385): 171–175. doi:10.1126/science.aam9309.

Kirk, D. (1996). Demographic transition theory. *Population Studies* 50(3): 361–387. doi:10.1080/0032472031000149536.

Knoke, D. and Yang, S. (2011). *Social network analysis*. London: Sage.

Kohler, H.-P., Billari, F.C., and Ortega, J.A. (2002). The emergence of lowest-low fertility in Europe during the 1990s. *Population and Development Review* 28(4): 641–680. doi:10.1111/j.1728-4457.2002.00641.x.

Kolk, M. (2014). Multigenerational transmission of family size in contemporary Sweden. *Population Studies* 68(1): 111–129. doi:10.1080/00324728.2013.819112.

Kuper, A. (2003). What really happened to kinship and kinship studies. *Journal of Cognition and Culture* 3(4): 329–335. doi:10.1163/156853703771818073.

Lévi-Strauss, C. (1969). *The elementary structures of kinship*. Boston: Beacon Press.

Lubkemann, S. (2002). Where to be an ancestor? Reconstituting socio-spiritual worlds among displaced Mozambicans. *Journal of Refugee Studies* 15(2): 189–212. doi:10.1093/jrs/15.2.189.

Lubkemann, S. (2005). Migratory coping in wartime Mozambique: An anthropology of violence and displacement in 'fragmented wars'. *Journal of Peace Research* 42(4): 493–508. doi:10.1177/0022343305054093.

Madhavan, S., Clark, S., Beguy, D., Kabiru, C.W., and Gross, M. (2017). Moving beyond the household: Innovations in data collection on kinship. *Population Studies* 71(1): 117–132. doi:10.1080/00324728.2016.1262965.

Marsden, P. (2011). Survey methods for network data. In: Carrington, P.J. and Scott, J. (eds.). *The SAGE handbook of social network analysis*. London: Sage: 370–388.

McKinnon, S. (2000). Domestic exceptions: Evans–Pritchard and the creation of nuer patrilineality and equality. *Cultural Anthropology* 15(1): 35–83. doi:10.1525/can.2000.15.1.35.

Morland, P. (2014). *Demographic engineering: Population strategies in ethnic conflict*. London: Ashgate.

Neupert, R.F. and Prum, V. (2005). Cambodia: Reconstructing the demographic stab of the past and forecasting the demographic scar of the future. *European Journal of Population* 21(2–3): 217–246. doi:10.1007/s10680-005-6854-3.

Nobles, J., Frankenberg, E., and Thomas, D. (2015). The effects of mortality on fertility: Population dynamics after a natural disaster. *Demography* 52(1): 15–38. doi:10.1007/s13524-014-0362-1.

Pinkerton, R.C., Oriá, R.B., Kent, J.W., Kohli, A., Abreu, C., Bushen, O., Lima, A.A.M., Blangero, J., Williams-Blangero, S., and Guerrant, R.L. (2011). Short report: Evidence for genetic susceptibility to developing early childhood diarrhea among shantytown children living in northeastern Brazil. *American Journal of Tropical Medicine and Hygiene* 85(5): 893–896. doi:10.4269/ajtmh.2011.11-0159.

Platt, L., Luthra, R., and Frere-Smith, T. (2015). Adapting chain referral methods to sample new migrants: Possibilities and limitations. *Demographic Research* 33(24): 665–700. doi:10.4054/DemRes.2015.33.24.

Poletta, F.A., Orioli, I.M., and Castilla, E.E. (2014). Genealogical data in population medical genetics: Field guidelines. *Genetics and Molecular Biology* 37(1): 171–185. doi:10.1590/S1415-47572014000200004.

Rafnsson, S.B., Shankar, A., and Steptoe, A. (2015). Longitudinal influences of social network characteristics on subjective well-being of older adults: Findings from the ELSA study. *Journal of Aging and Health* 27(5): 919–934. doi:10.1177/0898264315572111.

Randall, S. (2005). The demographic consequences of conflict, exile and repatriation: A case study of Malian Tuareg. *European Journal of Population* 21(2–3): 291–320. doi:10.1007/s10680-005-6857-0.

Randall, S. and Coast, E. (2014). Poverty in African households: The limits of survey and census representations. *The Journal of Development Studies* 51(2): 162–177. doi:10.1080/00220388.2014.968135.

Roy-Gagnon, M.H., Moreau, C., Bherer, C., St-Onge, P., Sinnett, D., Laprise, C., Vézina, H., and Labuda, D. (2011). Genomic and genealogical investigation of the French Canadian founder population structure. *Human Genetics* 129(5): 521–531. doi:10.1007/s00439-010-0945-x.

Salganik, M.J. and Heckathorn, D.D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology* 34(1): 193–240. doi:10.1111/j.0081-1750.2004.00152.x.

Singh, K., Karunakara, U., Burnham, G., and Hill, K. (2005). Using indirect methods to understand the impact of forced migration on long-term under-five mortality. *Journal of Biosocial Science* 37(6): 741–760. doi:10.1017/S002193200400700X.

Verwimp, P. (2003). Testing the double-genocide thesis for central and southern Rwanda. *Journal of Conflict Resolution* 47(4): 423–442. doi:10.1177/0022002703254478.

Williams, N.E. (2013). How community organizations moderate the effect of armed conflict on migration in Nepal. *Population Studies* 67(3): 353–369. doi:10.1080/00324728.2012.754927.

Williams, N.E. (2015). Mixed and complex mixed migration during armed conflict: Multidimensional empirical evidence from Nepal. *International Journal of Sociology* 45(1): 44–63. doi:10.1080/00207659.2015.1005434.

Williams-Blangero, S. and Blangero, J. (2006). Collection of pedigree data for genetic analysis in isolate populations. *Human Biology* 78(1): 89–101. doi:10.1353/hub.2006.0023.

Williams-Blangero, S., Vandeberg, J.L., Blangero, J., and Teixeira, A.R.L. (1997). Genetic epidemiology of seropositivity for Trypanosoma Cruzi infection in Rural Goias, Brazil. *American Journal of Tropical Medicine and Hygiene* 57(5): 538–543. doi:10.4269/ajtmh.1997.57.538.

Yin, R.K. (2009). *Case study research: Design and methods*. Thousand Oaks: Sage.

Zhao, Z. (2001). Chinese genealogies as a source for demographic research: A further assessment of their reliability and biases. *Population Studies* 55(2): 181–193. doi:10.1080/00324720127690.