Blood is Thicker than Bloodshed: A Genealogical Approach to Reconstruct Populations after Armed Conflicts

Diego Alburez-Gutierrez

Max Planck Institute for Demographic Research

gatemonte@gmail.com

Abstract

This paper introduces the Extended Genealogy Method (EGM), a new approach for collecting high-quality data for demographic analysis using extended genealogies. The EGM uses chain-referral sampling to create socio-centric family networks. The sampling and data processing strategies address retrospective and selection bias and ensure data completeness. To evaluate the quality of the method, the paper shows how the EGM was used to reconstruct the demographic history of a population in Guatemala affected by mass violence in 1982. Data on 3,566 unique individuals and 1,986 partnerships were collected from 100 EGM interviews. The paper describes how the data were collected, processed, and checked for systematic sources of error. It also shows how they were used to reconstruct the demographic dynamics in the population for the 1955-2015 period. The EGM can be applied reconstruct the demographic dynamics of local populations in contexts of data scarcity, such as during and after armed conflicts.

Introduction

Our understanding of the demographic consequences of conflict has been constrained by the scarcity of data. It is usually not a priority to collect demographic data using traditional methods such as household surveys in the context of mortality crises (Hill, 2004). As a result, data collected before or after the crisis are 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, to name only a few studies. 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 is a common problem – surveys carried out in Nigeria, Sri Lanka, Egypt, and Colombia have also suffered from this limitations (Brück et al., 2013). Researchers analysing survey data form conflict-affected regions do not always report these sampling limitations nor consider how they affect their analyses.

This paper proposes a data collection method that contributes to fill this gap and to improve our understanding of the effects of mortality crises on population and family dynamics. This article has two objectives. The first is to present an original approach for collecting demographic data in settings of armed conflicts using genealogical data. This includes showing how the proposed methodology differs from the existing approaches in anthropology and genetics. It discusses how this method integrates qualitative tools to enhance the collection of quantitative data.

Second, it shows that the proposed data collection method can be used to produce retrospective data for conducting demographic reconstructions of local populations in post-conflict settings. For this, it describes how the method was applied to reconstruct excess mortality after a series of mass killings affecting an indigenous population in Guatemala. The paper evaluates the quality of the evidence produced by the new data collection method and its potential for conducting demographic research on other populations for which no other data exist. This is of particular importance given that time-variant data on conflict-affected population are not widely available.

Data sources in the demography of conflict

Most quantitative analyses of armed conflict have relied on national registration data, censuses, or household surveys. Each of these data sources has drawbacks. Armed conflicts cause statistical offices to discontinue data collection or to produce unreliable data (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 are available, they can be inaccurate – sometimes deliberately so for political goals (Morland, 2014). The lack of accurate data also affects model-based approaches such as population projections, because they use 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 (de Walque, 2005; e.g. Agadjanian et al., 2008; de Walque and Verwimp, 2010), but others do not (Verwimp, 2003; Randall, 2005; Heuveline and Poch, 2007). Low quality baseline data affect 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). On top of 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 (non-representative) samples. This approach can be useful to suggest mechanisms through which conflict affects demographic behaviour. Few studies in the demography of conflict have a strong qualitative component (Lubkemann, 2002, 2005; Randall, 2005). Qualitative data are usually collected to complement quantitative studies, to explore the participants' own perspectives, or to clarify local definitions. Heuveline & Poch (2007, p. 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-varying data on war-affected populations limits our understanding of how conflict violence affects local populations. It is difficult to follow up individuals as they experience violent events or forced displacement and very few studies have done so. The Chitwan Valley Family Study collected monthly panel data on demographic events in the eponymous Nepalese valley during the 1996-2006 civil war (Axinn et al., 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 (Frankenberg et al., 2011; Axinn, 2015); displacement and migration (Gray et al., 2014; Williams, 2015); and the role of community and social organisations in dealing with disaster (Williams, 2013; Nobles et al., 2015).

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 (see 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 (Mckinnon, 2000; Kuper, 2003). For this reason, anthropological kinship data rarely meet the high data quality standards required for demographic analysis. Unmarried women, marriages without children, and early deaths tend to be underreported in historical and anthropological accounts (Castilla and Adams, 1996; Gamella and Carrasco-Muñoz, 2017). Demographic studies that have reconstructed and analysed genealogies from archival sources, such as parish records (Hammel and Gullickson, 2004) and historical family trees (Zhao, 2001), warn that genealogies recorded in administrative sources suffer from similar limitations.

Outside the social sciences, medical genetics has produced the most systematic and up-to-date protocols for collecting genealogies (Williams-Blangero and Blangero, 2006; Poletta et al., 2014). However, these guidelines were designed for studying genetic inheritance, not demographic processes. 'Pedigrees' (genetic genealogies) usually do not record childless partnerships and tend to underreport child mortality, making them unsuitable for the study of the social processes that underlie demographic change. This is a particular concern for this study since armed conflicts increase child mortality rates.

Genealogies have many analytical advantages. On the one hand, 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 were collected). They record the relationship between members of a population as well. This means that it is possible to derive a wide array of demographic measures and rates from them. Birth histories, for example, can be easily reconstructed for women and men. It is also possible to observe changes in demographic behaviour over generations in the same family (e.g. compare the fertility of parents to that of their children). Since genealogies are in themselves social networks (more precisely, they are family networks) they can be analysed using the principles of social network analysis (Borgatti et al., 2009). This paper uses this approach to evaluate the influence of social support and social pressure on demographic behaviour, to give only a few examples.

The Extended Genealogy Method (EGM)

This section presents the basic principles of the data collection method known as the Extended Genealogy Method (EGM). This method uses theory and methods from anthropology, historical research, and human genetics. The method is an efficient approach for collecting genealogies that contain all related individuals that have lived and died in a given population during a specific period ('extended genealogies'). This section introduces the design of the EGM, emphasising respondent selection and data collection tools.

No such tool currently exists in the field of demography. Most family network data in demographic studies come from survey household rosters. These usually assume that families are limited to the household. Therefore, they tend to produce a series of unrelated family trees, one for each household interviewed (Randall and Coast, 2014). The resulting genealogies are ego-centric, meaning that they are defined exclusively in relation to the head of house. These cannot be used to produce extended genealogies since there is no way of merging individual ego-centric networks to create a unified data base of all family relations in a population (Milligan, 2010; Madhavan et al., 2017).

Participant selection and reliability

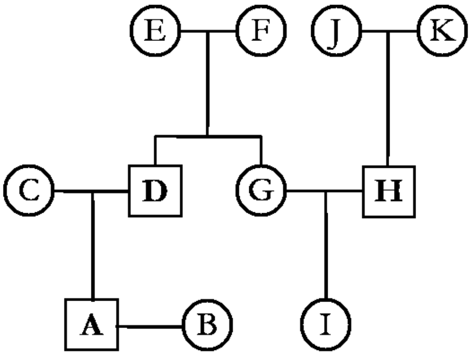
Constructing an extended genealogy requires merging multiple family histories into a unique genealogical dataset. Genetic studies often rely on a census-like approach where genealogical data are collected from all households in a given population. Post-hoc efforts are made to merge the data and remove duplicates (e.g. Choh et al., 2001; Pinkerton et al., 2011). This procedure generally produces one large cluster of related individuals and several minor isolated genealogies. The latter are usually ignored for the analysis. In a classical 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 distributed amongst 254 unrelated family clusters, were subsequently dropped from the analysis. The rest of the section outlines a more efficient approach for collecting and managing genealogical data in local populations.

Participant selection is a central component of any data collection exercise. The EGM relies on a form of researcher-led chain-referral sampling (Platt et al., 2015) to improve data collection efficiency by removing the need to interview every household in a population. The sampling strategy starts with a given seed respondent – which is purposefully chosen to meet criteria of accessibility, marital status, and age – and selects the next case from the list of individuals generated in the first interview. Multiple initial cases, or seeds, can be chosen and interviews conducted in parallel. More details on seed selection are given in the next section.

Interviews should 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 EGM questionnaire (including nephews and nieces). Figure 1 represents a population reconstructed in this way from three hypothetical genealogical interviews. In this example, a seed interview with individual *A* produced data on individuals *A* through to *D*. A second interview with individual *D* produced data on individuals *A* to *I*. The last interview, with individual *H*, yielded information on her parents, spouse, and child (*G* to *K*). Note that the respondents for the next interviews were always drawn from the pool of individuals known at the time.

Seven out of the eleven members of this hypothetical population were reported in at least two independent interviews (individuals *A, B, C, D, G, H,* and *I*). I call this 'multiple reporting'. Genealogical networks with a higher degree of multiple reporting (i.e. where individuals are reported by more independent sources) provide additional opportunities for cross-checking and inputting missing data. They can also be used to estimating the degree of underreporting (e.g. child mortality or childless marriages), as is explained below.

Figure 1. Diagram exemplifying the genealogical chain-referral sampling (hypothetical family group)



☐ interviewed; ○ not interviewed.

This process should be repeated until all members of the population have been enumerated. I introduce the concept of 'genealogical saturation' as a criterion for evaluating to degree to which all members of a population have been registered in the genealogical interviews. The criterion is useful for determining whether more interviews need to be carried out. The concept is based on the notion of theoretical saturation, which in qualitative research is understood 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, p. 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 given population. The genealogical chain-referral sampling results in a considerable overlap in reporting – someone might be reported as a mother in one interview and as a sister in another. 'Multiple reporting' is a feature of this sampling method that helps mitigate the effects of recall bias, since multiple observations from independent sources exist for the genealogical records. It also addresses the known limitation that reporting reliability decreases for more distant relatives (Heady, 2010).

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 – e.g. is there a unique local definition of a partnership? If this information is not available in the published literature, it can be generated from exploratory qualitative research. The same data can later be used to interpret the results of the quantitative analysis. Understanding local social norms can help screen for errors and inconsistencies in the data. A high incidence of inbreeding, homosexual unions, or concurrent partnerships might reflect flaws in the data if these practices are not accepted locally. Typographical and transcription errors can produce impossible circular relations such as an individual married to themselves. Finally, it is useful to consider biological constraints. 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 easily be implemented to identify these cases.

A first 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, p. 371). The boundaries of the population have to be defined both 'vertically' (i.e. in terms of generational depth) and 'horizontally' (e.g. should the family of an in-marrying partner be recorded?). Clear criteria must be developed regarding 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 may be required to ensure that the data on migrants are accurately recorded. Researchers must also define the type of relationships that the EGM will record. Families can be defined in strictly biological terms, but they can include adopted individuals, or other forms of locally defined kinship relations. All of these criteria must be defined considering the research question, the size of the population, and the available resources.

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

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 partnerships, respectively. This paper refers to the three modules collectively as the 'EGM questionnaire'. 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 that survived to a certain age or formed a union (this is standard practice in birth histories). Second, the design includes an 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, as is discussed later.

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 but can include an 'Individual ID' field to link the members of the household to the 'Individuals Module' (see below). The EGM questionnaire used in Río Negro only collected data on date of birth and civil state of household members. Additional fields can be included to collect information on schooling, employment status, etc., depending on the objectives of the study.

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

The third component of the questionnaire is a 'Unions Module', in which every line represents a separate union between two partners. Two fields ('Partner 1' and 'Partner 2') reference the row numbers of the union 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 unions must be recorded in this module. The module also records data on the characteristics of the union, such as type of union (e.g. marriage, cohabitation), start date and end dates of the union, reason for termination, etc. As with the previous component of the questionnaires, new fields can be added to record other information as required by the research design.

The EGM uses the principles of relational database design to record kinship relations. The Individuals Module and the Unions Module are tables linked by ID fields that uniquely identify rows in each table. Table 1 shows how the hypothetical genealogy shown in Figure 1 would be recorded in this tabular format (including only individuals *A-G* for simplicity's sake). The 'Individuals Module table' includes two columns that register the unions associated with an individual. The ‘Parent union ID’ field refers to the ID of the union formed by the parents of the individual. The field ‘Couple union ID’ refers to the ID of the union(s) between the individual and their partner(s). A similar relational field is included in the ‘Unions Module table’ (the colour of the columns indicate how the fields are linked across the two tables). The resulting tables are redundant by design: the repetition is intended to reduce typographical and transcription errors.

The EGM layout can be modified according to the research aims and the data available. Existing information such as archival records and registry data can be incorporated into the data collection. Digital data collection tools can be used to display additional information in real-time as new data are entered in mobile phones or tablets. This can include showing contextual data and portrays, or suggesting dates and potential matches from the existing genealogical records (more on this below). The EGM can be easily 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 this paper exemplifies these points by showing how the EGM was applied to reconstruct the demographic history of a war-affected population in Guatemala.

Table 1. Relational database design in EGM questionnaires: Individuals and Unions Module tables (recording a hypothetical genealogy)

Individuals Module table

Ind.ID	Ind. name	Parent union ID	Couple union ID	…
1	A	2	3	…
2	B	NA	3	…
3	C	NA	2	…
4	D	1	2	…
5	E	NA	1	…
6	F	NA	1	…
7	G	1	NA	…

Unions Module table

Union ID	Ind. ID		Ind. name	…
1	Partner 1	5	E	…
	Partner 2	6	F	…
2	Partner 1	3	C	…
	Partner 2	4	D	…
3	Partner 1	1	A	…
	Partner 2	2	B	…

Applying the EGM after a mortality crisis

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 (see Johnston, 2005 for a detailed history).

EGM-generated data showed that more than a third of Río Negro's population died between 1979 and 1983; 94% of these deaths are 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 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).

Research design

This section gives details of the implementation of the EGM. A mixed methods design was implemented to reconstruct the demographic history of Río Negro, including its experience throughout the armed conflict. An initial qualitative stage of the study generated contextual information for the design of the EGM questionnaire and the sampling.

Fieldwork was carried out between November 2015 and November 2016 in the municipality of Rabinal, Guatemala. Most interviews took place in the resettlement area, although short visits were made to other locations to collect information on hard-to-reach family groups. The initial three months were spent building rapport, getting acquainted with the local kinship organisation and creating an Event History Calendar (Axinn et al., 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 final calendar included 276 dated events between 1898 and 2016, including war-related events, political developments, and infrastructure projects. The calendar only included events that could be cross-checked using independent archival or historical records.

Data collection and management

Genealogical data were collected between January and October 2016. Respondents for the two initial or seed interviews were selected according to three criteria: (a) they possessed relevant genealogical knowledge, (b) they had been born before the 1982 massacres, and (c) they resided in Pacux. The initial qualitative component of the study provided the information needed to identify potential respondents for the seed interviews. Following a chain-referral sampling methodology, the next respondents were chosen from the pool of records produced by the previous interviews until genealogical saturation was achieved (more on this below).

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 et al., 2015). In a genealogical network, this means that seed respondents would tend to have more relatives than the population average. This is a concern because it might bias the population towards members of more extensive family networks. However, analysis of the genealogical data showed that his was not the case in Río Negro, where seed respondents actually had a lower-than-average family size.

We conducted 112 genealogical interviews. 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 unions. Many of these were duplicates – 62% of all individuals were reported by at least two independent respondents. On average, each individual was recorded in 2.5 separate interviews (sd = 1.26). After removing the duplicate records, 3,566 unique individuals and 1,018 unique unions remained. Given the iterative nature of the research design, the genealogical data had to be 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 unions (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 et al., 2014). The interviews carried out in Río Negro included women in 50% of the cases.

Genealogical saturation and data completeness

This section discusses how genealogical saturation was evaluated to determine when the data collection should be terminated. Three tests were conducted to assess the completeness of the genealogical data produced by the EGM in Río Negro (i.e. to determine whether genealogical saturation has been achieved). All of them required comparing the EGM-generated data to independent population registers.

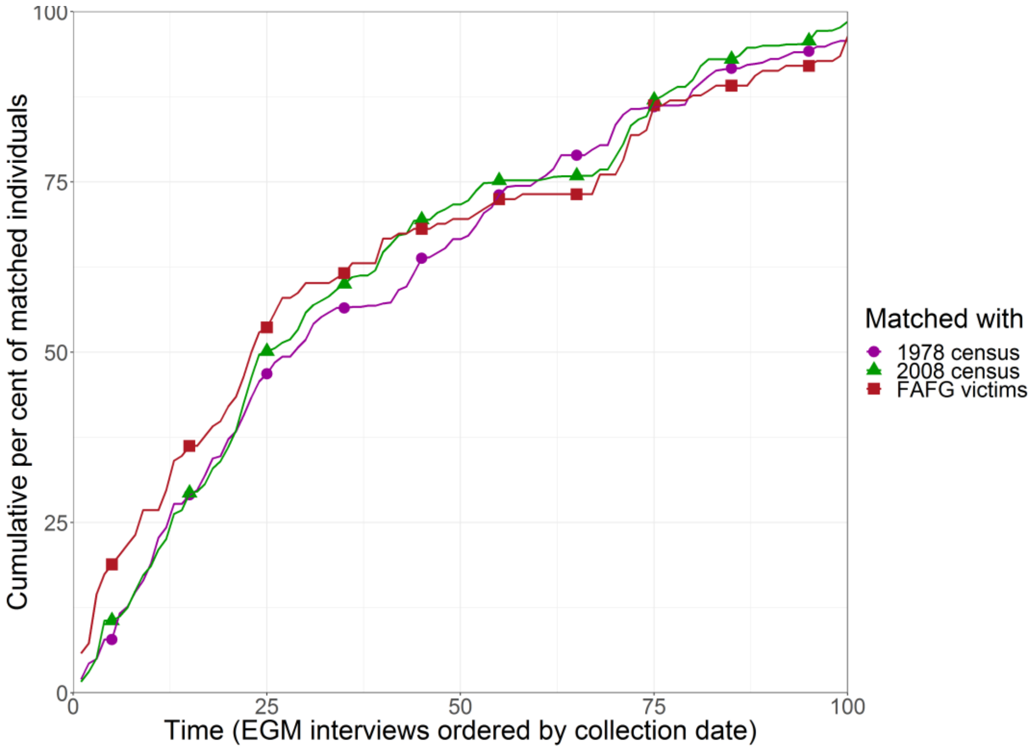
Initially, the EGM data was matched by name to a list of victims of the Río Negro Massacres exhumed by the Guatemalan Forensic Anthropology Foundation (FAFG).[1] This list is incomplete since not all victims have been exhumed yet. Nevertheless, all the listed victims should have been reported during the EGM interviews. For the second and third tests, the genealogical records were matched by name and date of birth with the two local censuses available for Río Negro, one conducted in 1978 and another one in 2008.[2] 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.

Figure 2. Genealogical saturation: completeness rates estimated by matching genealogical records with census data and independent list of victims

---

[1] https://www.fafg.org/bd/b_victima.php, accessed April 25 2018.

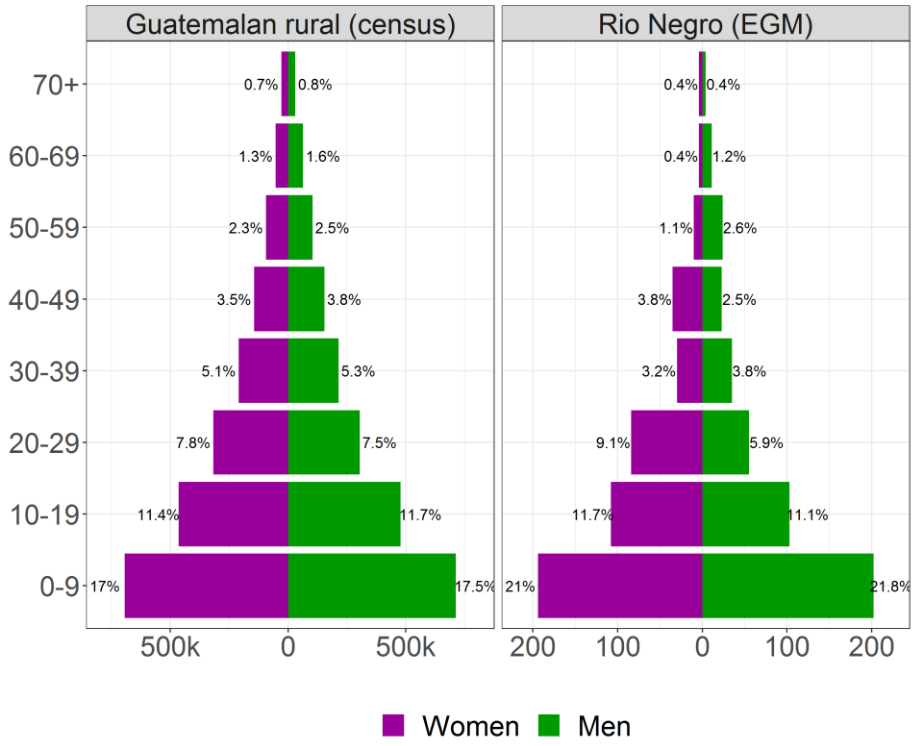
[2] The 1978 census was conducted by the INDE Electric Company to serve as a baseline for future compensation programmes. The 2008 census was carried out to identify the beneficiaries of the monetary war reparations provided by the Guatemalan government.



These matching exercises made it possible to evaluate the progression towards genealogical saturation**.** The vertical axis in Figure 2 represents the share of records from the independent registers matched to the EGM data over time. The vertical axis 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 100 genealogical interviews. The individuals who were missing altogether from the genealogical dataset were identified as members of three households that perished during the war and had no living descendants. The percentage of omission is too small to have affected the final population composition significantly.

The image also shows an important characteristic of the sampling strategy chosen for collecting genealogical data. The sampling criteria prioritised accessible respondents. This means that the first interviews produced data on the more visible members of the population. Later interviews recorded information on harder-to-reach individuals. This included internally displaced members of the population or individuals 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.

Figure 3. A comparison of the age-sex structure of the Guatemalan national population and the village of Río Negro in 1981



Source: Data from the 1981 Guatemala census (left) and from this study (right).

Anthropological and historical genealogies tend to underreport women and children. To evaluate whether this was the case in Río Negro, it is useful to compare the population structure of the village, reconstructed from the genealogical data, with that of the Guatemalan rural population as a whole. A visual examination of Figure 3 shows that the age and sex distribution was remarkably similar for both populations in 1981, one year before the mass killings. This provides additional confirmation that the EGM produced reliable demographic data since Río Negro was a typical indigenous community before the mass killings, according to the 1978 INDE census report. This, together with the preceding analysis of genealogical saturation, supports the conclusion that no particular demographic group was underreported in the genealogical data. The next section evaluates the quality of the data even further to determine whether the EGM performs better than the existing methods for collecting genealogies.

Data quality

The system of multiple reporting built into the sampling of the EGM 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.[3] The share of missing data of eight key demographic variables (date of birth, date of death, place of birth, current location, number of 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, with the exception of the date of non-violent deaths (rates of missing values = 44%). The events of 1982 overshadowed other, less dramatic deaths. It is possible that interviewers and respondents made more effort to accurately report conflict casualties than non-conflict deaths.

The paper now considers four potential sources of bias in the genealogical data. This was essential because most of the EGM-generated records were 'secondary participants' (i.e. relatives of the respondent). The sources of error were evaluated by considering whether (1) data on more distant relatives were less accurate; (2) data quality was poorer for individuals who were not alive at the time of the data collection; (3) data on migrants were underreported; and (4) interview length affected reporting quality. Three indicators of data quality were considered to determine the presence of these sources of bias: age heaping, the share of individuals with no reported parents, and the share of individuals for whom the dates of birth and death were unknown (Table 2).

Table 2. Data quality and completeness of EGM-generated data: an evaluation of four potential sources of bias

	Whipple index for ages		Unknown values %		
	18-47	23-62	Parents id	Birth date	Death date
1 Distance to respondent					

---

[3] Multiple systems were implemented to identify and merge duplicate records to avoid double counting. Records were matched by name, date of birth, date of death, and a range of other variables including mother's name, mother's date of birth, etc. This minimised the number of duplicate records, although inevitably some remained in the final database.

	Closest relative - Q1	95.2	96.2	9.0	3.1	2.5
	More distant relative - Q4	103.9	94.5	19.1	4.0	1.6
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

Whipple Index key: < 105 very accurate; 105-110 relatively accurate; 110-125 acceptable; 125-175 bad.

Overall, these sources of bias did not affect data quality systematically, with one exception. Information was less reliable for individuals who were already 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 family 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 digit only for those aged 18-47. 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. those reporting more individuals) were of a better quality. Older and better-informed respondents produced longer interviews with more reliable data. As expected, interviews with couples were the most reliable; those with only male respondents tended to be less complete and accurate.[4]

This section showed that the quality of the EGM-generated data was good for Río Negro. The next section shows a number of examples of how data from the EGM were used to reconstruct the historical population of the village.

Reconstructing the historical population of Río Negro using EGM-generated data

The size and composition of a population are elemental demographic measures. Table 3 shows this breakdown for Río Negro in five selected years (1981, 1983, 1993, 2003, and 2013). The table gives the exact size of each demographic group and makes it possible to compare the distribution of the population over time. Each of these yearly populations were reconstructed by conducting 'pseudo-censuses' on the genealogical data (i.e. by filtering only those individuals who were alive in any given year). The necessary variables for this filtering (date of birth and date of death) were obtained from the genealogical data. Pseudo-censuses could only be carried out after fully deduplicating the EGM-generated records to avoid artificially inflating population size or over-representing any particular age group.

---

[4] Measured in terms of share of missing data, accuracy of date reporting, and underreporting of known individuals.

Table 3. Age and sex distribution of population in Río Negro created conducting pseudo-censuses of the genealogical data (selected years)

Year	1981 (pre-killings)		1983 (post-killings)		1993		2003		2013	
Age group	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
0 to 4	106	103	61	65	121	145	175	169	139	147
5 to 9	88	99	75	80	98	106	185	164	143	145
10 to 14	74	56	56	51	57	62	120	142	172	165
15 to 19	34	47	27	36	72	78	97	106	184	162
20 to 24	48	26	20	21	54	50	57	62	120	139
25 to 29	36	29	25	26	27	35	70	78	96	98
30 to 34	18	16	12	9	20	21	53	50	57	59
35 to 39	12	19	10	10	25	22	27	34	68	76
40 to 44	17	12	6	7	12	7	20	20	53	48
45 to 49	18	11	9	1	9	9	25	22	23	32
50 to 54	4	13	6	5	6	6	11	7	19	20
55 to 59	6	11	3	6	7	1	9	9	25	20
60 to 64	2	9	1	2	5	4	6	6	10	6
65 to 69	2	2	1	3	3	6	7	1	9	9
70 to 74	3	1	0	1	0	2	4	4	5	4
75 to 79	0	0	1	0	1	2	2	6	7	0
80+	1	3	0	0	0	0	1	3	5	9
Total	469	457	313	323	517	556	869	883	1135	1139

Rio Negro has been a young population over time. The share of adults over 45 years of age has been consistently small, whilst children under 15 constituted a clear majority before the year 2003. There were signs of population ageing after this year, with the population under 15 constituting a smaller share of the total population in 2013 Nevertheless, the population remained relatively young at this point.

There was a clear 'dip' in total population size in 1982, resulting from the Río Negro Massacres. According to the genealogical data, 38% of the pre-conflict population was killed in 1982 (366 of the 970 original inhabitants of the village). A comparison of the 1981 and 1983 populations evidences that the 1982 massacres affected members of all age groups. 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 shows 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.

It is remarkable that, even though the mass killings produced extensive mortality, they did not alter the male to female ratio in any age group considerably. Men and women were similarly affected by the violent events (179 women and 187 men were killed), even if more women in reproductive age were killed in total. 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. This suggests that violence was directed against young women in particular.

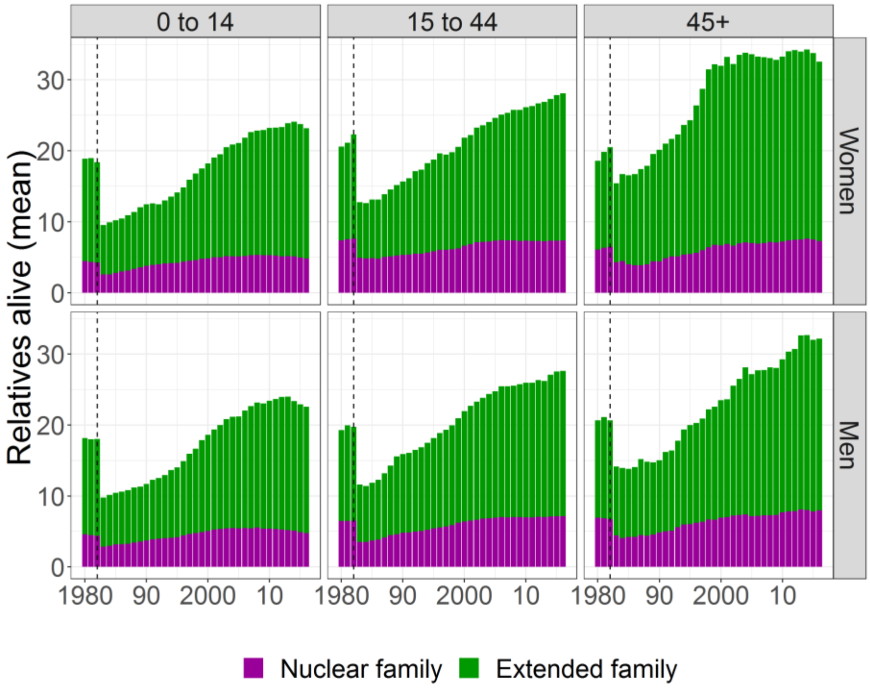
Conflict mortality also affected the survivors of the violence. Data from this study showed that all the inhabitants of Río Negro lost at least one member of the extended family during the mass killings. The genealogical dataset produced by the EGM can be used to quantify the effects of the mass violence on family structure. This is easiest done by considering changes in family size for both the nuclear and extended families.[5] **¡Error! La autoreferencia al marcador no es válida.** shows that women and men in all age groups saw a dramatic reduction in the size of their family networks after the massacres. The average extended family size, for example, plummeted by 44% between 1981 and 1983

---

[5] In this study, nuclear family refers to parents, children, siblings, and spouses; the extended family also includes cousins, aunts, uncles, nephews, grandparents, and grandchildren.

(from 19.3 to 10.8). This reduction was more dramatic for younger individuals. Children under 15 saw the average size of their extended families more than halved (from 14.1 to 6) in the same period. Adults over 45, on the other hand, experienced only a 25% reduction in their average family size (from 13.8 to 10.4). These figures would be very difficult to obtain from traditional household surveys, which tend to focus on the family living in the household (Madhavan et al., 2017).

Figure 4. Reconstructed average family size in Río Negro over time by sex and age group (1980-2015)



Discussion

A genealogical approach for reconstructing population data

This 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 family histories. 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. Independent population registers 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 also characterised by its sampling efficiency, which takes advantage of the network structure of genealogies. The genealogical chain-referral sampling method substantially reduced the amount of data needed to reconstruct an extended genealogy, as compared to examples from the medical genetics literature. A population of 3,566 was reconstructed with considerable accuracy from only 100 genealogical interviews. This paper showed how individual-level demographic characteristics of the population can be calculated from these data. A flexible research design and dynamic data management tools 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 family dynamics was essential for establishing an appropriate research design. Qualitative data provided information on vernacular kinship terminology, partnerships definitions, 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.

The extent to which kinship relations can be established between members of a population is determined by the generational depth of the data; mainly, by the number of ancestors known for every induvial. This is known as ancestry depth. An individual with known parents and grandparents has an ancestry depth of two – one step from individual to parents; another step from parents to grandparents. A depth of two, the minimal required to identify cousins and parents' siblings, was is available for 70% of the inhabitants of Río Negro. Cousins, aunts, and uncles could not be identified for one third of the population. Most of these were individuals in older cohorts who did not provide information about their parents.

Limitations of the EGM method

Retrospective questions are prone to recall bias. This limitation also affects other 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 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 paper successfully reconstructed the last 60 years of Río Negro' 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).

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 particular 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 crisis.

The method outlined in this paper can be used to collect demographic data, including information on fertility, mortality, and union 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, for example, was linked to local censuses using record linkage techniques, which made it possible to obtain household-level socioeconomic data. Similar data sources are likely to 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, p. 384). This was the case for the eight individuals who refused to be interviewed in Río Negro. Given the networked structure of genealogies, their information could be retrieved from a sibling or a cousin. Data were collected in this way under the assumption that kinship information is part of a shared communal knowledge.

Analysing the EGM-generated social network data

Reliance on survey-generated data has encouraged individual-level analysis in demography; it has also limited the choice of research questions and widened the gap between demographic research and social theory (Kirk, 1996; Bachrach, 2014). 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, even though armed conflicts are usually assumed to have negative effects on the 'social fabric' (Justino, 2011, p. 13) or the 'social bonds' of a community (Esparza, 2005, p. 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 notions such as social support, transfer of information, and similar processes. Genealogical data are 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. Family network data can be used to analyse social structure, a concept that has generally been ignored in the demographic literature. But discussing community-level dynamics does not preclude discussion of individual agency.

The structure of a network around an individual can be taken as an explanatory factor for that individual's behaviour. It is not sufficient to focus on the availability of networks of support, since, in itself, the structure of the network does not reveal anything about its function (i.e. whether individuals actually do receive support from their social networks).

How effectively these networks are used is important (Cohen and Wills, 1985; Rafnsson et al., 2015). Integrating qualitative and quantitative data is key for achieving a more comprehensive understanding of these relationships.

Genealogical data can easily be transformed to perform l demographic analysis. As mentioned before, this can be achieved by conducting pseudo-censuses on the data. This was exemplified in Figure 3, which only shows only those individuals alive at the start of 1981. Time-variant demographic rates can be reconstructed in this way. It is also possible to follow-up individuals or family groups over time in order to explore longitudinal changes in demographic behaviour. The association between variables of interest can be explored using traditional descriptive, regression, or survival analyses.

Concluding remarks

This paper outlined the principles of the Extended Genealogy Method (EGM), a systematic approach for collecting retrospective demographic data using socio-centric family networks. Anthropologists, genealogists, and geneticists have collected family histories in the past, but no approach currently exists to ensure that the data meet the specific data quality requirements for 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 at 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. Family networks can be taken as proxies for networks of support. The EGM can be adapted to study networks of friendship or other types of relations. The principles of the EGM can also be applied to reconstruct multiple, unconnected, extended genealogies. This could be useful for studying, for example, family networks or networks of support in displacement settings. The United Nations High Commissioner for Refugees (UNHCR) has recently become interested in applying this approach to understand networks of support in refugee camps. There is a growing interest in applying social network analysis to understand the transmission of fertility in low- and high-income countries (Kohler et al., 2002; Bernardi and Klärner, 2014). A major limitation of the current studies is the lack of longitudinal social network data. A genealogical approach has much to contribute to this field of inquiry.

This article 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. It showed that the EGM approach can be used to answer a wide range of research questions, particularly in the context of data scarcity. Researchers in other areas of study will no doubt find it helpful.

References

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

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

Alburez-Gutierrez, D., and C. Segura. 2018. The UNHCR Demographic Projection Tool: estimating the future size and composition of forcibly displaced populations, *UNHCR Statistics Technical Series* (1): 1–22.

Axinn, W. G. 2015. Demographic Change: The Case of Chitwan Valley in Nepal, *International Journal of Sociology* 45(1): 1–3.

Axinn, W. G., D. Ghimire, and N. E. Williams. 2012. Collecting Survey Data during Armed Conflict, *Journal of Official Statistics* 28(2): 153–171.

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

Bachrach, C. A. 2014. Culture and Demography: From Reluctant Bedfellows to Committed Partners, *Demography* 51(1): 3–25.

Baker, S. E., and R. Edwards. 2012. How many qualitative interviews is enough?, *National Centre for Research Methods Review Paper* (Unpublished): 1–42.

Bernardi, L., and A. Klärner. 2014. Social networks and fertility, *Demographic Research* 30(1): 641–670.

Borg, A., and M. Sariyar. 2016. *RecordLinkage: Record Linkage in R. R package version 0.4-10.* . Retrieved from https://cran.r-project.org/package=RecordLinkage

Borgatti, S. P., A. Mehra, D. J. Brass, and G. Labianca. 2009. Network Analysis in the Social Sciences, *Science* 323(April): 892–896.

Brück, T., P. Justino, P. Verwimp, and A. Tedesco. 2013. *Measuring Conflict Exposure in Micro-Level Surveys*. Stockholm: SIPRI.

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

CEH. 1999. Capítulo segundo: Las violaciones de los derechos humanos y los hechos de violencia (Tomo I) [Chapter II: Human Rights violations and the acts of violence (Volume I)], In *Guatemala: Memoria del Silencio [Guatemala: Memory of Silence]*. Guatemala City: UNOPS - Comisión para el Esclarecimiento Histórico.

Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson. 2016. *Shiny: Web Application Framework for R. R package version 0.14.2.* . Retrieved from https://cran.r-project.org/package=shiny

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

Cohen, S., and T. A. Wills. 1985. Stress, Social Support, and the Buffering Hypothesis, *Psychological Bulletin* 98(2): 310–357.

CPS/SSDSPF. 2014. *Enquête Démographique et de Santé au Mali 2012-2013 [Mali DHS 2012-2013]*. Maryland: CPS, INSTAT, INFO-STAT and ICF International.

de Walque, D. 2005. Selective mortality during the Khmer Rouge period in Cambodia, *Population and Development Review* 31(2): 351–368.

de Walque, D., and P. Verwimp. 2010. The Demographic and Socio-economic Distribution of Excess Mortality during the 1994 Genocide in Rwanda, *Journal of*

African Economies 19(2): 141–162.

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.

Frankenberg, E., T. Gillespie, S. Preston, B. Sikoki, and D. Thomas. 2011. Mortality, the Family and the Indian Ocean Tsunami, *The Economic Journal* 121(August): 162–182.

Gamella, J. F., and E. Carrasco-Muñoz. 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.

Gray, C. L., E. Frankenberg, T. Gillespie, C. Sumantri, and D. Thomas. 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.

Hammel, E. a, and A. Gullickson. 2004. Kinship structures and survival: maternal mortality on the Croatian-Bosnian border 1750-1898, *Population Studies* 58(2): 145–59.

Heady, P. 2010. Appendix 3: sampling, fieldwork and data quality, In P. Heady and M. Kohli (Eds.), *Family, Kinship and State in Contemporary Europe, Vol. 3*. Frankfurt am Main: Campus.

Heuveline, P., and B. Poch. 2007. The Phoenix population: demographic crisis and rebound in Cambodia., *Demography* 44(2): 405–26.

Hill, K. 2004. *War, humanitarian crises, population displacement, and fertility: a review of evidence*. Washington, DC: National Academies Press.

Himelein, K., S. Eckman, S. Murray, and J. Bauer. 2016. Second-Stage Sampling for Conflict Areas: Methods and Implications, *World Bank Group Policy Research Working Paper* (7617): 1–31.

Johnston, B. R. 2005. *Chixoy Dam Legacy Issues Study. Volume 1: Executive summary*. Santa Cruz, California: Center for Political Ecology.

Justino, P. 2011. The Impact of Armed Civil Conflict on Household Welfare and Policy, *IDS Working Papers* 2011: 1–38.

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

Knoke, D., and S. Yang. 2011. *Social network analysis. Quantitative Applications in the Social Sciences*. London: Sage Publications.

Kohler, H.-P., F. C. Billari, and J. A. Ortega. 2002. The Emergence of Lowest-Low Fertility in Europe During the 1990s, *Population and Development Review* 28(4): 641–680.

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

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.

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.

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

Marsden, P. 2011. Survey Methods for Network Data, In P. J. Carrington and J. Scott (Eds.), *The SAGE Handbook of Social Network Analysis*. London: Sage Publications.

Mckinnon, S. 2000. Domestic Exceptions: Evans-Pritchard and the Creation of Nuer Patrilineality and Equality, *Cultural Anthropology* 15(1): 35–83.

Milligan, G. 2010. Appendix 1: a brief explanation of the Kinship Network Questionnaire, In P. Heady and M. Kohli (Eds.), *Family, Kinship and State in Contemporary Europe, Vol. 3*. Frankfurt am Main: Campus.

Morland, P. 2014. *Demographic Engineering: Population Strategies in Ethnic Conflict*. London: Ashgate.

Neupert, R. F., and V. Prum. 2005. Cambodia: Reconstructing the Demographic Stab of the Past and Forecasting the Demographic Scar of the Future, *European Journal of Population* 21: 217–246.

Nobles, J., E. Frankenberg, and D. Thomas. 2015. The Effects of Mortality on Fertility: Population Dynamics After a Natural Disaster, *Demography* 52(1): 15–38.

Pinkerton, R. C., R. B. Oriá, J. W. Kent, A. Kohli, et al. 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.

Platt, L., R. Luthra, and T. Frere-Smith. 2015. Adapting chain referral methods to sample new migrants: Possibilities and limitations, *Demographic Research* 33(1): 665–700.

Poletta, F. a., I. M. Orioli, and E. E. Castilla. 2014. Genealogical data in population medical genetics: Field guidelines, *Genetics and Molecular Biology* 37(1): 171–185.

Rafnsson, S. B., A. Shankar, and A. Steptoe. 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.

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.

Randall, S., and E. Coast. 2014. Poverty in African Households: the Limits of Survey and Census Representations, *The Journal of Development Studies* 51(2): 162–177.

Singh, K., U. Karunakara, G. Burnham, and K. Hill. 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.

Verwimp, P. 2003. Testing the Double-Genocide Thesis for Central and Southern Rwanda, *Journal of Conflict Resolution* 47(4): 423–442.

Williams-Blangero, S., and J. Blangero. 2006. Collection of Pedigree Data for Genetic Analysis in Isolate Populations, *Human Biology* 78(1): 89–101.

Williams-Blangero, S., J. L. VandeBerg, J. Blangero, and A. R. L. Teixeira. 1997. Genetic epidemiology of seropositivity for Trypanosoma cruzi infection in rural Goias, Brazil, *American Journal of Tropical Medicine and Hygiene* 57(5): 538–543.

Williams, N. E. 2013. How community organizations moderate the effect of armed conflict on migration in Nepal, *Population Studies* 67(3): 353–69.

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.

Yin, R. K. 2009. *Case Study Research: Design and Methods*. Thousand Oaks, Calif: Sage Publications.

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.