American Journal of Epidemiology **Submitted Manuscript**[1]

Special Collection: N/A

Title: Estimating death rates in complex humanitarian emergencies using the network survival method

Authors: Casey F. Breen, Saeed Rahman, Christina Kay, Joeri Smits, Abraham Azar, Steve Ahuka, Dennis M. Feehan

ORCiD IDs: Casey F. Breen (0000-0003-3702-4746), Dennis M. Feehan (0000-0001-7008-4375)

Correspondence Address: Casey F. Breen, 42-43 Park End Street Oxford, OX1 1NF, UK (casey.breen@demography.ox.ac.uk)

Joint Authorship: N/A

Affiliations:
- Casey F. Breen: Leverhulme Centre for Demographic Science, Department of Sociology, and Nuffield College. University of Oxford. Oxford, United Kingdom.
- Saeed Rahman: IMPACT Initiatives. Geneva, Switzerland.
- Christina Kay: IMPACT Initiatives. Goma, Democratic Republic of the Congo.
- Joeri Smits: IMPACT Initiatives, Geneva, Switzerland; Tufts University. MA, USA
- Abraham Azar: IMPACT Initiatives, Geneva, Switzerland
- Steve Ahuka: School of Public Health, University of Kinshasa, Kinshasa, Democratic Republic of the Congo
- Dennis M. Feehan: Department of Demography, University of California, Berkeley. CA, United States.

Key words: Humanitarian Emergencies, Mortality Estimation, Network Survival Method

Acknowledgments[2]**:** We are grateful to Fiston Muhigirwa Stanislas and Anny Mwamini Ngoie for their excellent supervision of fieldwork. We thank Paul Tshiminyi Munkamba and Tresor Katanga Tubale for assistance with enumerator training. For helpful conversations and comments, we thank Aashish Gupta, Diego Alburez-Gutierrez, Payal Hathi, Ridhi Kashyap, Patricia McManus, Michelle Niemann, Bruno Masquelier, Jonathan Polonsky, and participants of the PAA 2024 session "Environmental and Climate Influences on Morbidity and Mortality," the BSPS 2023 session "Data Science: Modelling Kinship," the 2023 Quetelet Conference, the 2023 Oxford Hard-to-Reach Populations Conference, the Population Studies Group seminar series at LSHTM, the Leverhulme Centre for Demographic Science seminar series at University of Oxford, the Population Cluster Group seminar series at NYU-Abu Dhabi, and the Measuring Adult Mortality Community of Practice Webinar by the United Nations Statistical Commission.

[1] N/A indicates not applicable.
[2] Study investigators, conference presentations, preprint publication information, thanks.

Funding: The authors acknowledge financial support from the Leverhulme Trust (Grant RC-2018-003) for the Leverhulme Centre for Demographic Science, the Bill and Melinda Gates Foundation (INV-045370), the Berkeley Population Center (P2C HD 073964), and the Berkeley Center for the Economics and Demography of Aging (5P30AG012839).

Conflict of Interest: N/A

Disclaimer: N/A

Data Availability Statement: All data and replication code is available from:

Estimating death rates in complex humanitarian emergencies using the network survival method

April 29, 2025

Abstract

Reliable estimates of death rates in complex humanitarian emergencies are critical for assessing the severity of a crisis and for effectively allocating resources. However, in many humanitarian settings, logistical and security concerns make conventional methods for estimating death rates infeasible. We develop and test a new method for estimating death rates in humanitarian emergencies using reports of deaths in survey respondents' social networks. To test our method, we collected original data in Tanganyika Province of the Democratic Republic of the Congo ($N = 5{,}311$), a setting where reliable estimates of crude death rates (CDR) are in high demand. Qualitative fieldwork suggested testing two different types of personal networks as the basis for CDR estimates: deaths among immediate neighbors and deaths among kin. We compare our network-based estimates (0.44 deaths per 10,000 person-days) against a standard retrospective household mortality survey, which estimated a CDR nearly twice as high (0.81 deaths per 10,000 person-days). Given that both methods are equally plausible, our findings underscore the need for further validation and development of both methods.

1 Introduction

Reliable estimates of death rates are essential for addressing complex humanitarian emergencies. These estimates are crucial for crisis assessment, resource allocation, preserving the historical record of tragedies, and supporting advocacy [1–3]. Recent estimates of mortality in humanitarian emergencies have guided effective responses to armed conflicts [4–6], famine [7], and war crimes [8].

The most reliable way to learn about death rates is generally through data from a high-quality civil registration and vital statistics system (CRVS). However, during complex humanitarian emergencies, this is often not feasible. In some settings, high-quality CRVS systems may not exist, while in others, the system may deteriorate over the course of the emergency [9]. For instance, at the time of the 2010 earthquake in Haiti, there was no high-quality CRVS system [10], and even if there had been, the earthquake caused a near-total collapse of civic infrastructure and processes [11]. Alternative methods for estimating death rates are therefore needed. Existing methods fall into three broad classes:

First, retrospective household mortality surveys are a widely-used approach for estimating death rates [12–16]. These surveys typically involve asking a probability sample of households about vital events and household composition during a recall period [17]. Household surveys are time-consuming and costly, and even when well-executed, can be prone to various errors leading to underestimation or overestimation of death rates [18–20]. For example, Jarrett et al. (2020) [21] conducted a careful validation exercise, comparing deaths reported in a surveillance system and a retrospective household survey. They found that over half of the deaths reported in the survey were either outside the recall window, occurred in a different household, or were fabricated. In practical terms, in humanitarian emergencies, obtaining a high-quality probability sample is often challenging or impossible. For instance, data collection was paused for three weeks in response to major security concerns, including the attack and burning of a data collection office in a 2004 mortality household survey in the Democratic Republic of the Congo [22]. Therefore, household mortality surveys are generally not a feasible strategy for estimating real-time mortality in humanitarian emergencies [23].

Second, prospective demographic surveillance systems can be established for monitoring deaths [24]. In a prospective demographic surveillance system, trained enumerators visit homes and administer surveys, collecting data on deaths, births, and migration for prespecified time intervals (weekly, monthly, etc.). Ideally, this approach would provide real-time death rates, but in practice, updates occur only when new deaths are reported, which may happen only a few times per year. Additionally, properly enumerating the population denominator can take several months. Moreover, such surveillance systems are expensive, difficult to maintain, and often deteriorate in complex humanitarian emergencies [2, 25, 26].

Third, key informant reporting involves selecting key informants to report on mortality within a predefined community, such as a village or neighborhood [27]. Using capture-recapture methods, these data can be combined with lists of deaths from other sources to estimate death rates [28, 29]. This approach is more cost-efficient than surveillance systems or retrospective surveys, but a validation study conducted in four separate study sites found this approach undercounts deaths among children under five [27]. In certain types of humanitarian emergencies, selecting appropriate key informants may be challenging, and informants may struggle to accurately report on displaced populations. Future empirical work will be useful in furthering our understanding of the settings in which this method can be successfully applied.

Each of these methods is important, but has limitations that are exacerbated in humanitarian emergencies. There remains an urgent need for specialized methods to estimate timely death rates in humanitarian emergencies [23]. In this study, we adapt a method called network survival to the challenge of estimating death rates during a complex humanitarian emergency in which operational constraints prevent direct access to populations. The network survival method was originally developed to estimate national death rates [30]. Our study builds on this earlier work by introducing several key methodological innovations, including: (i) employing a non-probability sampling approach that allows remote data collection without an on-the-ground presence; (ii) using a short retrospective window to facilitate high-frequency mortality estimates; (iii) incorporating qualitative work to inform the specific choice of networks for reporting; and (iv) refining methods for blending two death rate estimates.

2 Study design and data collection

2.1 Study site

In order to empirically test our new method, we needed a setting that satisfied two criteria: (1) it should have characteristics similar to other places where humanitarian emergencies have emerged in the past; and (2) it should be possible to obtain a probability sample that could produce a set of estimates using a standard retrospective household survey. We chose three health zones in the Tanganyika Province of the Democratic Republic of the Congo: Kalemie, Nyemba, and Nyunzu (Figure 1).

These health zones border one another in the easternmost part of the country, which is characterized by high death rates and historical insecurity problems that have caused humanitarian emergencies to arise in the past [31, 32]. This region is an example of the kind of setting where humanitarian emergencies may emerge and methods for estimating death rates are critically needed. Further, in collaboration with our partner organization, IMPACT Initiatives, we determined that it would be possible to obtain a probability sample of households to produce estimates from a standard probabilistic household survey.

2.2 Design and data collection

Our design called for two separate data collection projects that produced several different estimates of the *crude death rate* (CDR) (Figure 2). The first data collection project used a new approach called network survival to produce CDR estimates from a sample that could realistically be obtained during a humanitarian emergency. In such emergencies, a conventional probability survey would likely be infeasible due to security and logistical challenges. Instead, we collected a non-probability quota sample designed to imitate a setting where operational constraints prevent humanitarian actors from reaching insecure areas, but populations may be moving back and forth to access services and markets or evacuating an insecure area ("quota sample"). The network survival method uses survey questions to collect information about deaths and exposure among respondents' personal networks (e.g. kin, neighbors); thus, it is possible to learn about people and places that cannot directly be

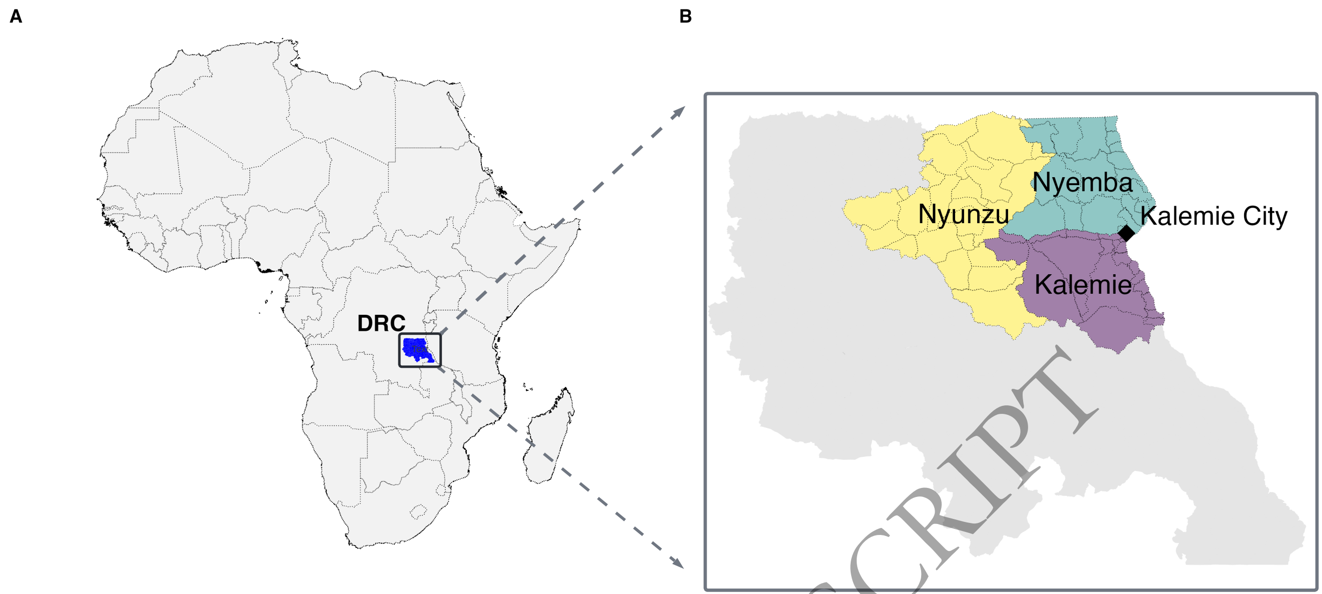

Figure 1: (*A*) Map of Africa with the Democratic Republic of the Congo's Tanganyika Province highlighted in blue. (*B*) Inset of Tanganyika Province, with our three focal health zones highlighted. The quota survey respondents were all sampled in Kalemie City, labeled with a black diamond. The household survey respondents were sampled in their respective health zones.

reached by the study team. As described below, data collected from the quota sample allow us to produce several different CDR estimates based on the network reports. We also asked quota sample respondents retrospective questions about deaths in their households.

The second data collection project aimed to produce a set of CDR estimates using the standard approach: a retrospective probabilistic household survey. We employed a probability sampling design to obtain a sample of households in the study area ("probability sample"). Enumerators visited these households in person and interviewed respondents using a standard survey instrument, which included questions about deaths among household members. Respondents were also asked the same network reporting questions used in the quota sample survey. Retrospective probabilistic household surveys are known to have some flaws [18–20] and they are generally not feasible in a humanitarian emergency. However, there is no perfect way to estimate mortality from a survey, and retrospective household surveys have been widely adopted by governments and non-governmental organizations (NGOs) [33, 34]. In our study, the probability-based household estimates serve two purposes: (1) they allow us to separate out the effect of non-probability sampling and the effect of reporting about network members on the network survival estimates; and (2) they allow us to have a set of estimates produced by a standard (if imperfect) method that can be used to contextualize and compare estimates from the new method.

Our primary comparisons are outlined in Figure 2. From both the quota sample and probability sample, we produce separate estimates of the CDR using the household method and the network survival method. Within each sample, we compare the household and network estimates to understand the difference attributable to methods. Within each method, we compare the probability and quota estimates to understand the difference attributable to sampling strategy.

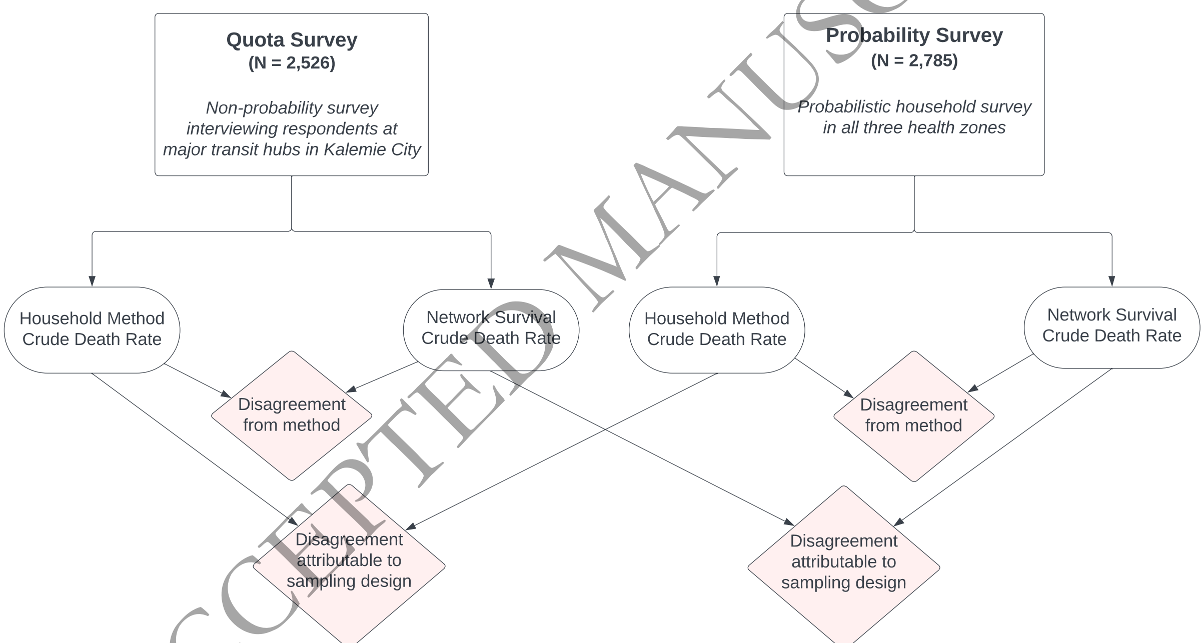

Figure 2: Illustration of study design and comparisons.

The quota survey targeted adults aged 18 and older using a non-probability, quota-based sampling strategy to intercept people coming into Kalemie City from all three health zones. Specifically, trained survey enumerators sampled respondents at transit hubs and service sites including ports, markets, taxi stands, foot paths, and health clinics in Kalemie City. The quotas were established based on gender and health areas, the geographic units below health zones. A total of 2,526 interviews were conducted in Kalemie City between March

1$^{\text{st}}$ and June 29$^{\text{th}}$, 2023. Respondents answered demographic and socioeconomic questions before reporting deaths within their kin and neighbor networks. Respondents were asked to report about deaths that occurred between January 1$^{\text{st}}$, 2023, and their interview date.

An important limitation of our quota sample is that it only accounts for gender and geography, and does not account for other dimensions of selection, such as age or socioeconomic status. In general, there is a tradeoff between representativeness and feasibility when implementing quotas: more complex quotas improve representativeness but can be challenging to implement, while simpler quotas may not be representative of the general population. Despite these limitations, quota surveys have been effective across a range of public health settings where probability sampling was infeasible due to logistical or financial constraints, including studies on conflict-affected populations in South Sudan [35], interpersonal contact during the COVID-19 pandemic [36], and mortality estimation in Ebola-affected regions [37].

The survey used a randomized order for the kin and neighbor modules and broke questions into subcategories to reduce cognitive load. To help reach more remote areas of the Nyunzu Health Zone, we established a secondary sampling site in Nyunzu Town. However, the sole enumerator in Nyunzu Town had limited direct supervision, leading to some potential data quality concerns. As this enumerator only conducted 30% of overall interviews for respondents living in Nyunzu, we dropped these data from our main analysis. A robustness check demonstrated no statistically significant differences in our estimates if these data were or were not included in our analysis (see Section S4 for details).

The probability survey was fielded directly after the quota survey, from July 21$^{\text{st}}$, 2023 to September 1$^{\text{st}}$, 2023. The probability survey was conducted in-person at 2,785 households in randomly sampled clusters across all three health zones. Respondents reported on deaths since January 1$^{\text{st}}$, 2023, the same reference date as the network survey. The probability survey included the complete network survival module, allowing us to produce both a standard household estimate and a network survival estimate of the CDR.

2.3 Quantity of interest

Our primary quantity of interest is the *crude death rate* (CDR), expressed in deaths per 10,000 people per day. These are the units typically used in complex humanitarian settings to estimate CDRs [4, 38, 39]. To convert to the standard units used in demography, deaths per 1,000 people per year, multiply the CDR by 36.5 (for details, see Section S2.3). Mathematically, the CDR M is given by $M = \frac{D}{N} \times 10,000$, where D is the number of deaths that occurred in a given time period, and N is the total *person-days* of exposure. Our primary quantity of interest was the CDR pooled across all three health zones from January 1$^{\text{st}}$, 2023 to June 29$^{\text{th}}$, 2023.

2.4 Formative Fieldwork

To help inform the design of our study, we conducted eight focus groups and 25 open-ended interviews in the study setting. The primary goal of this formative research was to identify the specific personal network(s) for respondents to report on: networks that are large enough for us to learn lots from each interview, but small enough to accurately recall and report death [40]. The formative research also informed other study parameters, including the recall period length, the method for estimating network size, and the selection of transit hubs and service sites (e.g., ports, taxi stands, markets) for sampling respondents. Our qualitative fieldwork suggested using two different types of personal networks as the basis for death rate estimates shown in Table 1: deaths among immediate neighbors and deaths among kin.

One potential issue in mortality estimation studies is recall bias, which occurs when respondents systematically forget or otherwise misreport past events, leading to inaccuracies in reported deaths. In the context of mortality estimation, this can result in underreporting of deaths, particularly when respondents struggle to recall exact dates or fail to report deaths that occurred further in the past. To mitigate recall bias, we selected a significant and locally memorable reference event—New Year's Day, January 1st, 2023—as the starting point for the recall period. Prior research has shown that anchoring recall to well-known events, such as Ramadan or the death of a political leader, improves accuracy in reporting deaths [39]. Our qualitative research revealed that New Year's Day was highly salient in this setting and helped respondents better remember whether a death occurred in the recall period. We also selected a relatively short recall window, which at maximum was eight months long, to further minimize recall bias.

Network Tie	Group
Household	Respondent's Household
Neighbor	1st Closest Neighbor Household
Neighbor	2nd Closest Neighbor Household
Neighbor	3rd Closest Neighbor Household
Neighbor	4th Closest Neighbor Household
Neighbor	5th Closest Neighbor Household
Kin	Respondent's Grandchildren
Kin	Respondent's Children
Kin	Respondent's Siblings
Kin	Respondent's Cousins
Kin	Respondent's Aunts/Uncles
Kin	Respondent's Parents
Kin	Respondent's Grandparents

Table 1: Household, kin and neighbor network subgroups.

3 Estimation

3.1 Network survival method

Building off the broader network reporting literature for studying hard-to-reach populations [41–44], the network survival method can be thought of as a generalization of the sibling method [45–49] and the network scale-up method [41]. The network survival method has generated estimates that closely align with those produced by international health organizations in a similar setting in Rwanda using a probability survey [30]. Further, in Brazil, over 25,000 respondents were probabilistically sampled across 27 different cities [50], and the network method estimates were benchmarked against the gold-standard vital statistics collected by the Brazilian government. The estimates aligned closely with the ground truth estimates from vital statistics, and were 15% more accurate at modest sample sizes (N $\approx$ 1,000) than the standard sibling method.

The core idea of the network survival method is to ask respondents to report about deaths occurring within their personal networks. Specifically, the network method asks a survey respondent to answer a series of questions that can be used to determine (i) how many people are in the respondent's personal network; and (ii) how many people in the respondent's personal networks died in a given time period. These network reports are then combined to estimate a death rate:

$$\widehat{M} = \left(\frac{\widehat{D}}{\widehat{N}}\right) \times 10,000 = \left(\frac{\widehat{y}_{F,D}}{\widehat{y}_{F,N}}\right) \times 10,000, \tag{1}$$

where $\widehat{M}$ is an estimator for the CDR; $\widehat{D}$ is an estimator for the number of deaths in the population; $\widehat{N}$ is an estimator for the amount of exposure; F is the frame population (i.e., the universe of people eligible to respond to the survey); $\widehat{y}_{F,D}$ is an estimate of the total number of reported deaths among personal network members over the reference period; and $\widehat{y}_{F,N}$ is an estimate of the total amount of exposure among personal network members reported over the reference period.

3.2 Network survival estimator

To use this estimator in our study, we must specify estimators for $y_{F,D}$ and $y_{F,N}$. The expression for estimating reported deaths, $y_{F,D}$, can be written as:

$$\widehat{y}_{F,D} = \sum_{i \in s} w_i \; y_{i,D}, \tag{2}$$

where s is the sample; w_i is a weight for respondent $i \in s$; and $y_{i,D}$ is the number of deaths among personal network members reported by respondent i. The expression for estimating reported exposure, $\widehat{y}_{F,N}$, can be written as

$$\widehat{y}_{F,N} = \sum_{i \in s} w_i \; d_i \; E_i, \tag{3}$$

where d_i is the reported number of people in i's personal network and E_i is the number of days of exposure respondent i reported about their personal network. The product of d_i and E_i estimates the total amount of exposure reported by respondent i in person-days.

Putting Equation 2 and Equation 3 together with Equation 1, we have the estimator we use in our study:

$$\widehat{M} = \left(\frac{\sum_{i \in s} w_i \; y_{i,D}}{\sum_{i \in s} w_i \; d_i \; E_i} \right) \times 10,000. \tag{4}$$

Equation 4 is convenient because it expresses the estimator in terms of respondent-specific weights w_i.

3.3 Producing estimates from quota sample

Our design called for quotas by gender and health area, the geographic units below health zones. This allowed us to closely match the overall target population's gender and geographic distribution. However, the quota did not account for selection with respect to socioeconomic status, age, or other characteristics. Quota sample respondents were wealthier and the youngest and oldest age segments were underrepresented compared to the general population (see Figure S1 for more details). To address this, we developed weighting strategies intended to mimic the availability of increasingly rich external data (Figure 3): unweighted estimates relying solely on our quota sample, estimates using WorldPop gridded population data for poststratification weights, and estimates with inverse probability weights (IPW) based on respondents' age, sex, household size, household age composition, and ownership of assets that correlate with household wealth. We construct IPW weights using logistic regression to model inclusion probability in the quota sample based on a pooled quota and probability sample [51].

Inverse probability weighting (IPW) is the preferred weighting approach for this method when sufficient auxiliary data is available, as it can more easily adjust for a broad range of characteristics. However, IPW relies on the availability of high-quality auxiliary data to weight against (e.g., a reliable census or probability survey), and measurement errors in auxiliary data may bias the adjusted estimates. Furthermore, IPW can lead to unstable weights when probabilities of selection are very small, resulting in high variance. Finally, all weighing strategies can only account for the measured dimensions of selection, and cannot address bias from unmeasured factors. For more details on weighting, see Section S2.6.

To assess sampling uncertainty, we constructed 10,000 bootstrap resamples. Each resample was drawn with respect to gender, health zone, and survey month, mirroring our original quota sampling design. From each bootstrap resample, we calculated a point estimate of the CDR. Using the 2.5th and 97.5th percentiles of these CDR estimates, we constructed a 95% uncertainty interval. This approach quantifies the uncertainty in our estimates due to the randomness of the sampling process.

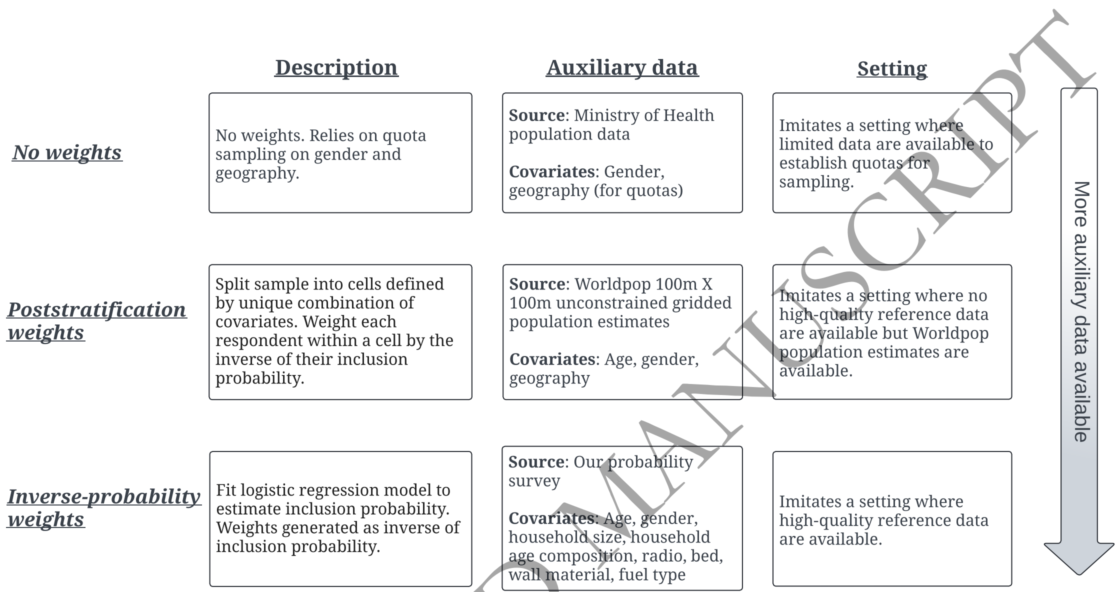

Figure 3: Overview of different weighting strategies. We developed weighting strategies intended to mimic the availability of increasingly rich external data.

We produced separate estimates using reports about neighbor and kin networks. In addition, we use a blended estimator to produce a combined estimate based on both the kin and the neighbor network reports [40]. Specifically, the blended estimate is based on averaging together the estimate from each network in a principled way. Suppose we have two estimators for N, $\widehat{N}^A$ and $\widehat{N}^B$. We define the blended estimate with pooling weight θ as:

$$\underbrace{\widehat{N}}_{\text{Blended Estimator}} = \underbrace{\theta\widehat{N}^A}_{\text{Weighted Estimator A}} + \underbrace{(1-\theta)\widehat{N}^B}_{\text{Weighted Estimator B}} \tag{5}$$

where θ is a weight $\in [0, 1]$. The advantage of this blended approach is that we expect it to produce smaller mean squared error (MSE) than either kin or neighbor estimate alone, because the estimate is based on more information. But this comes at the cost of additional assumptions; see Section S2.7 and Feehan et al. 2016 [40] for a detailed discussion and derivation of the optimal weight.

3.4 Producing estimates from probability sample

We produced CDR estimates from the probability sample using two methods: the standard household method and the network survival method. For the standard household method, we calculated person-time observed for each individual based on relevant dates within the recall period, such as date of birth, death, joining the household, or leaving the household. We then calculated the CDR by dividing the number of deaths by the total person-time observed and re-scaling to express as deaths per 10,000 person days. To generate network survival estimates from our probability sample, we apply the same estimator used for the quota sample. However, we do not use survey weights, as we consider the probability sample to be self-weighting. To make the probability survey estimates directly temporally comparable to the quota sample estimates, we restrict the probability sample to deaths and exposure reported during the same recall period as the quota sample (January 1^{st}, 2023 to June 29^{th}, 2023).

4 Results

First, we analyze our estimates from the quota sample. Figure 4 shows the distribution of household and network sizes. The average household size is 7 people. In comparison, the average kin network size is 26.7 people, and the average neighbor network size is 29.5 people. Correspondingly, respondents report many times more deaths in their neighbor and kin networks than in their own household.

Figure 5 presents three sets of CDR estimates, based on kin reports, neighbor reports, and a blended combination of the two. For each, we calculate three estimates: unweighted, poststratified (adjusted for gender, age, and geography), and inverse probability weighted (IPW, incorporating all available sociodemographic information). This allows us to assess how weighting adjustments impact our estimates.

To incorporate respondent-specific weights, we weight both death reports and exposure contributed by each individual, as described in Equation 4. Poststratification increases estimates slightly, except for household-based estimates, which decline modestly. In contrast, IPW raises estimates by approximately 40% for both kin and neighbor networks. Despite some variation, kin and neighbor estimates remain consistent across weighting strategies. Household estimates are noisier but generally align with network-based estimates.

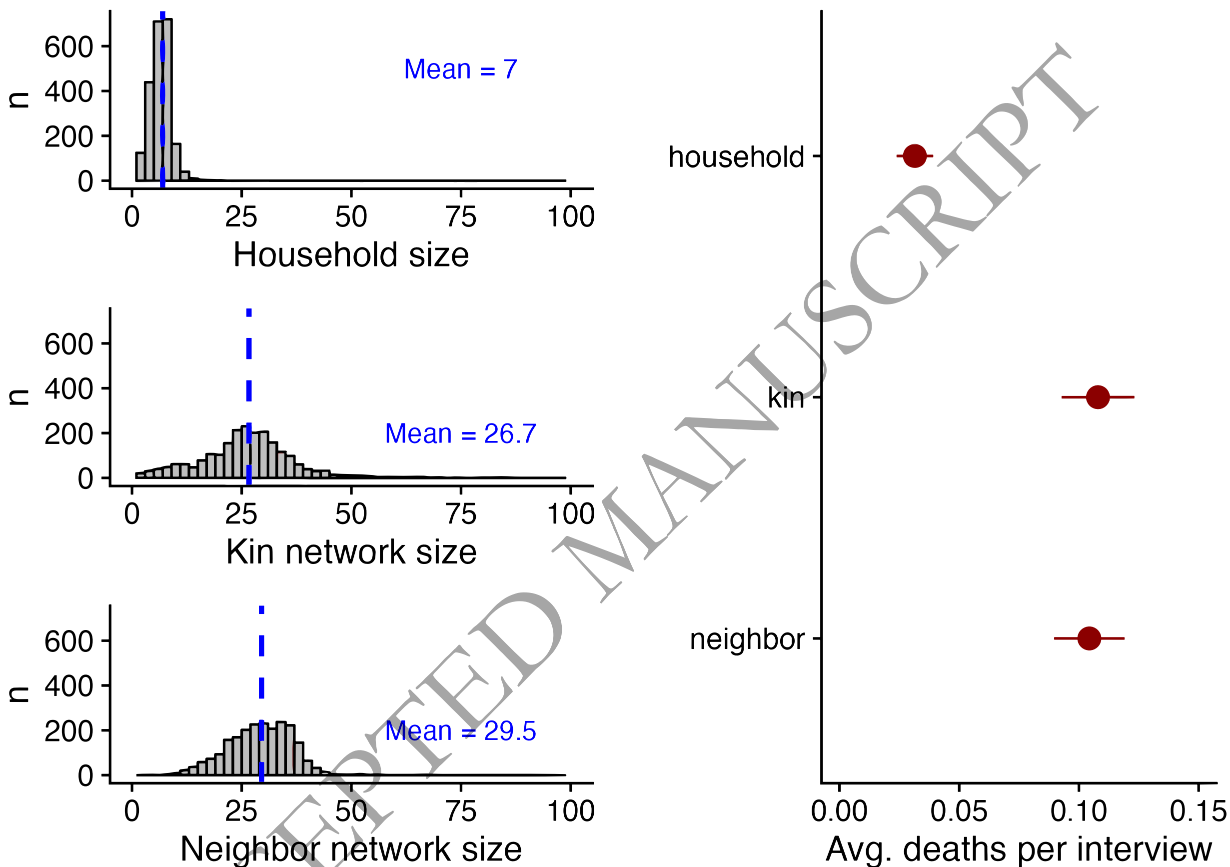

Figure 4: Network size and average deaths per interview from quota sample. Uncertainty bars show 95% confidence intervals.

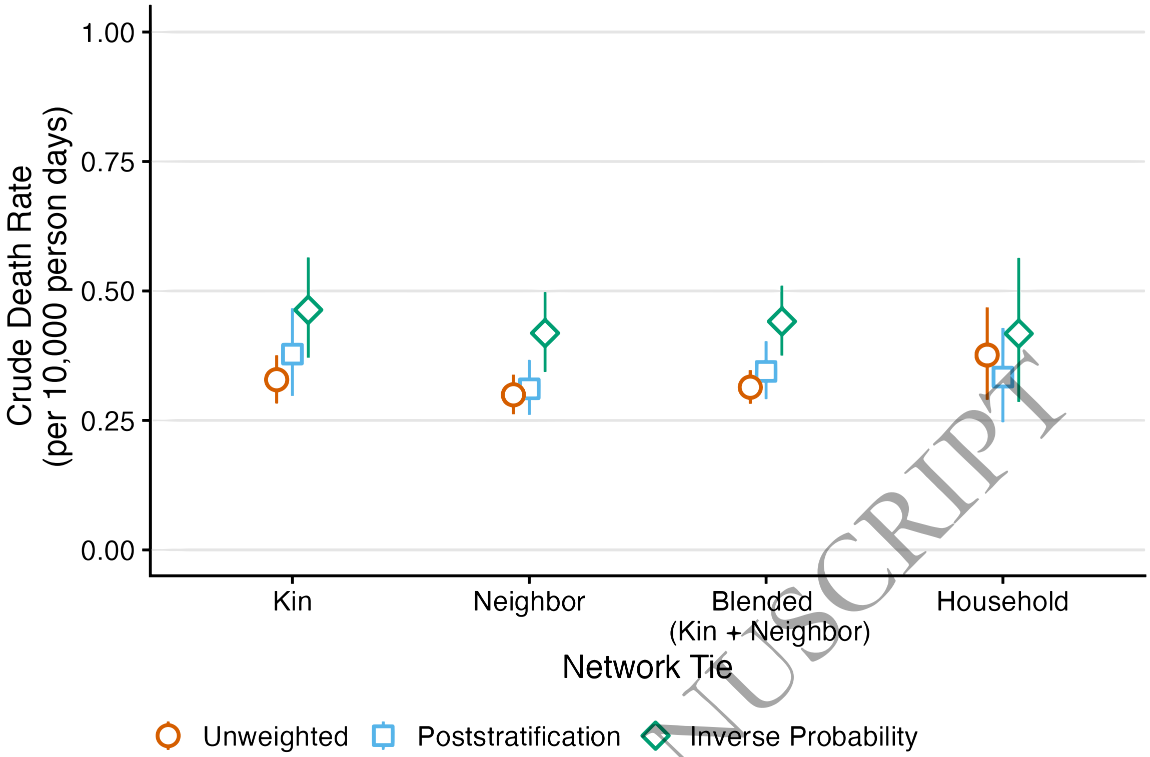

Figure 5: Network survival method estimates of the crude death rate (CDR) from the quota sample under three different weighting schemes. The CDR is expressed in units of deaths per 10,000 person days. Uncertainty bars show 95% confidence intervals.

Next, we compare estimates from our quota sample with estimates from our probability sample. For ease of comparison, we focus on what we would expect to be our best network estimate from our quota sample: our blended estimates with inverse-probability weights. The blended network IPW estimate is 0.44 (95% CI: 0.38–0.51), closely aligning with the kin (0.46, 95% CI: 0.37–0.56) and neighbor (0.42, 95% CI: 0.34–0.50) IPW estimates. This estimate combines information from neighbor and household reports and weights to account for selection into our quota sample.

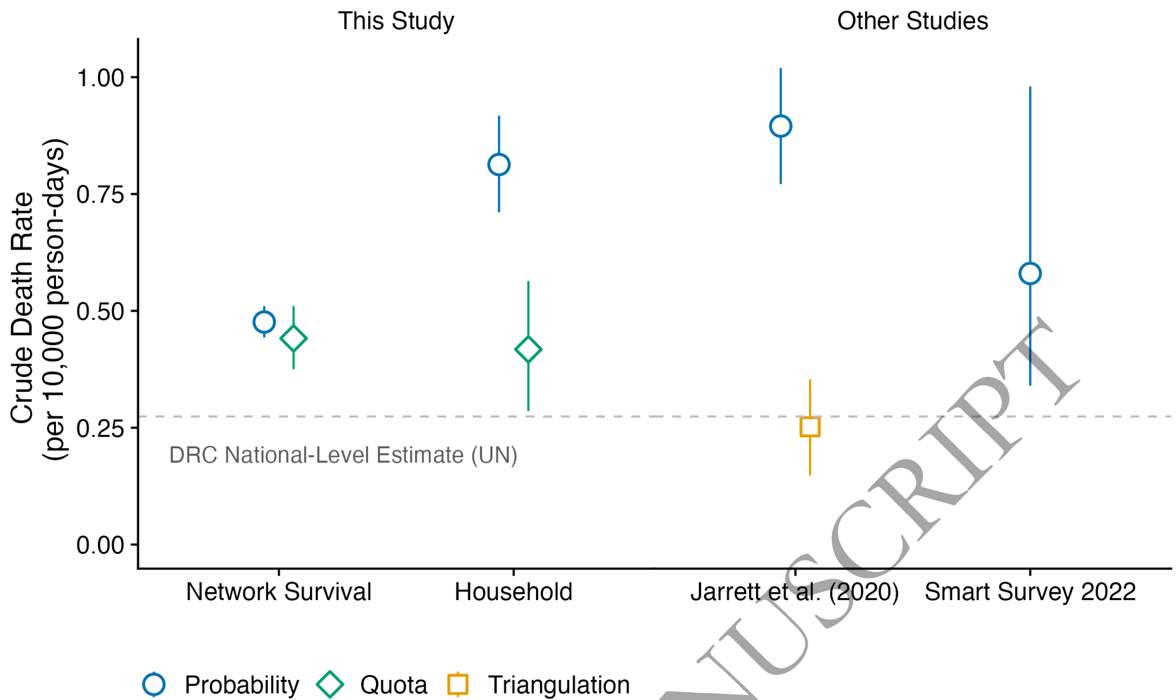

Figure 6: Comparison with CDR estimates from other studies. The quota sample estimates are weighted using inverse probability weights; the network survival estimates are blended estimates from both kin and neighbor networks. The Jarrett et al. [21] study was conducted in the Fizi province in 2011. The 2022 SMART survey was conducted in the Kalemie Health Zone in November 2022. Uncertainty bars show 95% confidence intervals.

Figure 6 presents the full set of comparisons between both arms of our study and external estimates. The blended network CDR estimate from the quota sample of 0.44 (95% CI, 0.38–0.51) aligns closely with the blended network estimate from the probability sample of 0.48 (95% CI, 0.44–0.51). Additionally, within our quota sample, CDR estimates based on household reports are consistent with both network estimates. However, the probability sample household estimate is substantially larger than both our quota sample household estimate and all network estimates.

To help contextualize this disagreement, we make several comparisons with other external studies, noting that these external estimates are neither perfectly temporally nor geographically aligned with ours. Our first comparison is to a Standardized Monitoring and Assessment of Relief and Transitions (SMART) survey conducted in November 2022 in the

Kalemie Health Zone [52]. This study asked respondents to report on deaths after August 1st and before the November interview date, an observation window approximately six months before our observation window. This survey found a CDR of 0.58 (95% CI, 0.34–0.98), slightly higher than the quota blended IPW estimate (0.44, 95% CI: 0.38–0.51).

Next, we compared our estimates with those from Jarrett et al. [21]. This study collected data in the Fizi Health Zone in 2011 and 2012. The Fizi Health Zone borders our study area to the north (see Section S2.8). Despite these estimates being over 10 years old and from a neighboring health zone with potentially differing contexts of conflicts, disease outbreaks, and population dynamics, they still provide valuable insights into the reliability of standard household-based CDR estimates in these settings. A standard, probabilistic household survey found a CDR of approximately 0.9 (95% CI,0.77–1.02). In order to understand how accurate the probabilistic household survey was, the authors conducted a separate surveillance of households and then re-interviewed all household respondents to reconcile any discrepancies between the surveillance and household survey. This careful reconciliation process found only approximately 28% of deaths reported in the household study had legitimately occurred in the study period; of the 72% of erroneous death reports, these "deaths" were either outside the recall period (32%), not within the household (48%), or false reports (20%). After this reconciliation process, the authors estimated a CDR of 0.25 (95% CI, 0.15–0.35), and hypothesized that strategic overreporting was responsible. This study adds to a growing body of literature highlighting that household surveys—even when well-executed and administered—may produce biased estimates of mortality [18–21]. These studies reinforce the need to be cautious when interpreting our own household-based estimates, which may potentially be affected by similar biases, including strategic over reporting.

We performed a series of validation checks to confirm the quality of our survey responses and network reports (see Section S3 for full details). First, to test for respondent fatigue during interviews, we compared responses based on the randomized order of our network modules. Regardless of the survey order, respondents reported nearly identical average household sizes and average numbers of deaths, indicating consistent and reliable responses. Second, we performed internal validity checks for network reports, focusing on relationships we expected to be reciprocal. The results showed no significant deviations from reciprocity, further confirming the reliability of the network reports. Finally, we compare the age composition of the quota sample (household, kin, and neighbor networks) with the age composition of the probability sample, finding high overall agreement.

5 Discussion

In this study, we introduce a new method for estimating death rates by adapting the network survival method to non-probability settings, as demanded by the unique constraints of humanitarian emergencies. We conducted formative fieldwork to help us pick which personal networks to ask respondents to report on, recall period length, and other important design parameters. We assessed the performance of this method in a realistic setting and conducted a probabilistic household survey for comparison. Although the limitations of household-based mortality surveys are well-documented [18–21] they are widely used and so we see them as a useful comparator, if not a gold standard. This comparison helps us better understand the plausibility of both sets of estimates and different potential sources of error relative to the ground truth. Taken together with external estimates, our results highlight a large amount of uncertainty surrounding the true underlying CDR in our focal health zones. Despite the lack of a reliable ground truth to benchmark our CDR estimates against, our study had several key findings.

Our quota sample taken at transit hubs and service sites in Kalemie City was positively selected with respect to socioeconomic status compared to the broader population. This was expected, as our quotas for the sample were only on gender and geographic region (health area), and did not address selection into the sample with respect to age or socioeconomic status. In our quota samples, our weighted estimates were substantially higher than our unweighted estimates. This suggests that adjusting for socioeconomic selection into the non-probability sample is crucial for producing accurate estimates and indicates that, as expected, people with lower socioeconomic status in this setting had neighbors and kin with higher mortality. After reweighting to adjust for selection, our network estimates from the quota sample aligned closely with our network estimates from our probability sample. Despite the major differences in sample design, the reweighted quota and probability samples produced nearly identical network estimates, demonstrating the effectiveness of the reweighting approach in this setting. Further, both network estimates were consistent with our estimated household CDR from the quota sample. However, the CDR estimate from the probability sample household reports was substantially higher than any other estimate.

This lack of agreement is surprising. Although our study cannot speak definitively to this discrepancy, we can speculate on possible explanations. Given the high level of non-government organization (NGO) activity in this area, respondents in the probability sample may have been motivated to answer in a way that would maximize their chances of receiving aid, similar to the 'strategic misreporting' hypothesized by Jarret et al. 2020 [21]. This incentive would be stronger in the probability sample, where enumerators visited respondents' households and could potentially return to deliver aid. In contrast, respondents in the quota sample likely had lower expectations of receiving aid, as they generally lived far away from Kalemie City and did not provide specific addresses or locations for follow-up. Our study included a verbal autopsy for reported household deaths, asking detailed questions about causes, which may have reduced the likelihood of fabricated deaths, but not false reports outside recall window or household. There may also have been a *memorial effect*, where the emotional salience of household members who passed away recently but prior to the observation window may result in over-reporting. This may be stronger in the probability sample where people are interviewed in households, as the environment itself may remind respondents of deceased household members, making the emotional salience stronger.

On the other hand, it is also possible that the network survival method underestimated the true CDR. Respondents may have forgotten about deaths or been unaware of deaths in their extended networks. The quota and probability samples also had slightly different recall periods. Our qualitative research, however, suggested that these factors are unlikely to produce errors big enough to explain the difference between the household and network estimates: respondents to our qualitative study reported that deaths were very salient and perceived themselves to be highly aware of deaths in their kin and neighbor networks. Our validity checks also find no cause for concern about data quality, though we cannot definitively rule out undiagnosed problems with the network reports.

The comparison with external estimates offered additional insights. The most directly comparable study, conducted in the Kalemie Health Zone approximately six months before our study produced estimates that aligned with the network survival estimates [52]. Another study, conducted 12 years earlier in a neighboring health zone [21], used a prospective mortality surveillance system to evaluate the accuracy of deaths reported on a probability-based household mortality survey similar to our household survey. The results revealed significant overreporting of deaths on the household survey. The authors hypothesized that the large presence of local and international NGOs may have led respondents to strategically make false reports about deaths in hopes of receiving aid [21]. Similar overreporting may help explain the discrepant household CDR estimate from the probability sample (Section S2.9).

There are several important next steps for future research, broadly falling into two key areas: additional validation efforts and methodological advancements. In terms of validation efforts, this study motivates more empirical work to validate and assess the performance of both the standard household survey method and the network survival method in conflict settings. An ideal validation study would take place a setting where high-quality, gold-standard mortality estimates can be obtained, such as a demographic surveillance site. This study design would allow for a systematic comparison of conventional household retrospective mortality surveys and the network survival method benchmarked against the surveillance-based estimates. An independent reconciliation of any reported discrepancies could be conducted to investigate inconsistencies, helping to determine the extent of strategic overreporting, missed reporting of true deaths (false negatives), and recall bias. Such a study would provide helpful evaluation of both standard household and network survival approaches.

From a methodological standpoint, future work could consider alternative model-based approaches to adjust for non-probability sampling. We investigated several different weighting strategies, but future work could explicitly model mortality for subgroups and incorporate upweighting, similar to multilevel regression with poststratification [53]. Additionally, while CDRs are a standard metric for measuring mortality in humanitarian emergencies, they depend on the overall age distribution of the population, limiting cross-context comparisons. Network-based methods could be extended to estimate age-specific death rates (requiring collecting more detailed information on the ages of all network members) that explicitly account for differential age structure across network ties. This is particularly important because, although the kin, neighbor, and household networks' age composition in crude age categories largely aligns (Figure S5), there are subtle age composition differences across networks. Finally, in our study, we analyzed all data after data collection was completed. Respondents reported deaths occurring within an average recall period of approximately six months. Future studies could explore the feasibility of shorter recall windows and implement a streamlined pipeline to generate estimates on a more regular basis.

The method introduced in this paper addresses a long-standing call for the development of new tools to estimate mortality in humanitarian emergencies [23]. We combined the network survival method with a quota sampling approach. This design could be deployed remotely in settings where operational constraints prevent humanitarian actors from reaching insecure areas, meaning it could potentially be applied to estimate death rates in a wide range of humanitarian emergencies. For example, a research team could establish a checkpoint outside of an ongoing humanitarian emergency. At this checkpoint, they could collect a quota sample, with quotas established based on gender, geographic region (based on when the emergency started), and other relevant characteristics. The survey instrument would collect information on deaths among immediate neighbors and deaths among kin, or some other network informed by qualitative research. With a sufficiently large sample size, CDR estimates could be generated monthly. The resulting CDR estimates could help track mortality over time, guide aid distribution, and support advocacy efforts for stronger interventions.

References

[1] Peter Heudtlass, Niko Speybroeck, and Debarati Guha-Sapir. Excess mortality in refugees, internally displaced persons and resident populations in complex humanitarian emergencies (1998-2012) - insights from operational data. *Conflict and Health*, 10:15, 2016. ISSN 1752-1505. doi: 10.1186/s13031-016-0082-9.

[2] Francesco Checchi. Estimation of population mortality in crisis-affected populations: Guidance for humanitarian coordination mechanisms. Technical report, 2018. URL https://healthcluster.who.int/publications/m/item/estimation-of-population-mortality-in-crisis-affected-populations.

[3] Francesco Checchi and Les Roberts. Documenting Mortality in Crises: What Keeps Us from Doing Better? *PLoS Medicine*, 5(7):e146, July 2008. ISSN 1549-1676. doi: 10.1371/journal.pmed.0050146. URL https://dx.plos.org/10.1371/journal.pmed.0050146.

[4] Francesco Checchi, Christopher I. Jarvis, Kevin Van Zandvoort, and Abdihamid Warsame. Mortality among populations affected by armed conflict in northeast Nigeria, 2016 to 2019. *Proceedings of the National Academy of Sciences*, 120(30):e2217601120, July 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2217601120. URL https://pnas.org/doi/10.1073/pnas.2217601120.

[5] Benjamin Coghlan, Pascal Ngoy, Flavien Mulumba, Colleen Hardy, Valerie Nkamgang Bemo, Tony Stewart, Jennifer Lewis, and Richard J. Brennan. Update on mortality in the Democratic Republic of Congo: Results from a third nationwide survey. *Disaster Medicine and Public Health Preparedness*, 3(2):88–96, June 2009. ISSN 1938-744X. doi: 10.1097/DMP.0b013e3181a6e952.

[6] Olivier Degomme and Debarati Guha-Sapir. Patterns of mortality rates in Darfur conflict. *Lancet (London, England)*, 375(9711):294–300, January 2010. ISSN 1474-547X. doi: 10.1016/S0140-6736(09)61967-X.

[7] Francesco Checchi and Courtland Robinson. Mortality among populations of southern and central Somalia affected by severe food insecurity and famine during 2010-2012. Technical report, May 2013. URL https://reliefweb.int/report/somalia/mortality-among-populations-southern-and-central-somalia-affected-severe-food.

[8] Patrick Ball, Wendy Betts, Scheuren Fritz, Dudukovich Jana, and Jana Asher. Killings and Refugee Flow in Kosovo March - June 1999. Technical Report A Report to the International Criminal Tribunal for the Former Yugoslavia, 2022. URL https://www.icty.org/x/file/About/OTP/War_Demographics/en/s_milosevic_kosovo_020103.pdf](https://www.icty.org/x/file/About/OTP/War_Demographics/en/s_milosevic_kosovo_020103.pdf).

[9] Olusesan Ayodeji Makinde, Clifford Obby Odimegwu, Mojisola O. Udoh, Sunday A. Adedini, Joshua O. Akinyemi, Akinyemi Atobatele, Opeyemi Fadeyibi, Fatima Abdulaziz Sule, Stella Babalola, and Nosakhare Orobaton. Death registration in Nigeria: A

systematic literature review of its performance and challenges. *Global Health Action*, 13 (1):1811476, December 2020. ISSN 1654-9880. doi: 10.1080/16549716.2020.1811476.

[10] Bénédique Paul, David Jean Simon, Vénunyé Claude Kondo Tokpovi, Mickens Mathieu, and Clavie Paul. Prevalence and factors associated with undocumented children under-five in Haiti. *International Journal for Equity in Health*, 23(1):169, August 2024. ISSN 1475-9276. doi: 10.1186/s12939-024-02255-8. URL https://doi.org/10.1186/s12939-024-02255-8.

[11] David McEntire, Abdul-Akeem Sadiq, and Kailash Gupta. Unidentified Bodies and Mass-Fatality Management in Haiti: A Case Study of the January 2010 Earthquake with a Cross-Cultural Comparison. *International Journal of Mass Emergencies & Disasters*, 30(3):301–327, November 2012. ISSN 0280-7270. doi: 10.1177/028072701203000303. URL https://doi.org/10.1177/028072701203000303.

[12] R. I. Glass, W. Cates, P. Nieburg, C. Davis, R. Russbach, H. Nothdurft, S. Peel, and R. Turnbull. Rapid assessment of health status and preventive-medicine needs of newly arrived Kampuchean refugees, Sa Kaeo, Thailand. *Lancet (London, England)*, 1(8173): 868–872, April 1980. ISSN 0140-6736. doi: 10.1016/s0140-6736(80)91365-3.

[13] Rolando J. Acosta and Rafael A. Irizarry. A Flexible Statistical Framework for Estimating Excess Mortality. *Epidemiology (Cambridge, Mass.)*, 33(3):346–353, May 2022. ISSN 1044-3983. doi: 10.1097/EDE.0000000000001445. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10200579/.

[14] Julie DaVanzo. A Household Survey of Child Mortality Determinants in Malaysia. *Population and Development Review*, 10:307–322, 1984. ISSN 0098-7921. doi: 10.2307/2807966. URL https://www.jstor.org/stable/2807966.

[15] Linda A. Bartlett, Shairose Mawji, Sara Whitehead, Chadd Crouse, Suraya Dalil, Denisa Ionete, Peter Salama, and Afghan Maternal Mortality Study Team. Where giving birth is a forecast of death: Maternal mortality in four districts of Afghanistan, 1999-2002. *Lancet (London, England)*, 365(9462):864–870, 2005. ISSN 1474-547X. doi: 10.1016/S0140-6736(05)71044-8.

[16] P. Salama, F. Assefa, L. Talley, P. Spiegel, A. van Der Veen, and C. A. Gotway. Malnutrition, measles, mortality, and the humanitarian response during a famine in Ehiopia. *JAMA*, 286(5):563–571, August 2001. ISSN 0098-7484. doi: 10.1001/jama.286.5.563.

[17] K. Lisa Cairns, Bradley A. Woodruff, Mark Myatt, Linda Bartlett, Howard Goldberg, and Les Roberts. Cross-sectional survey methods to assess retrospectively mortality in humanitarian emergencies. *Disasters*, 33(4):503–521, October 2009. ISSN 1467-7717. doi: 10.1111/j.1467-7717.2008.01085.x.

[18] Ian M. Timæus. Measurement of Adult Mortality in Less Developed Countries: A Comparative Review. *Population Index*, 57(4):552–568, 1991. ISSN 0032-4701. doi: 10.2307/3644262. URL https://www.jstor.org/stable/3644262.

[19] Bruno Lankoandé, Bruno Masquelier, Pascal Zabre, Hélène Bangré, Géraldine Duthé, Abdramane B. Soura, Gilles Pison, and Sié Ali. Estimating mortality from census data: A record-linkage study of the Nouna Health and Demographic Surveillance System in Burkina Faso. *Demographic Research*, 46:653–680, April 2022. ISSN 1435-9871. doi: 10.4054/DemRes.2022.46.22. URL https://www.demographic-research.org/articles/volume/46/22.

[20] Kenneth Hill, Peter Johnson, Kavita Singh, Anthony Amuzu-Pharin, and Yagya Kharki. Using census data to measure maternal mortality: A review of recent experience. *Demographic research*, 39:337–364, 2018. ISSN 1435-9871. doi: 10.4054/DemRes.2018.39.11. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6903798/.

[21] Prudence Jarrett, Frank J. Zadravecz, Jennifer O'Keefe, Marius Nshombo, Augustin Karume, and Les Roberts. Evaluation of a population mobility, mortality, and birth surveillance system in South Kivu, Democratic Republic of the Congo. *Disasters*, 44 (2):390–407, April 2020. ISSN 1467-7717. doi: 10.1111/disa.12370.

[22] Benjamin Coghlan, Richard J. Brennan, Pascal Ngoy, David Dofara, Brad Otto, Mark Clements, and Tony Stewart. Mortality in the Democratic Republic of Congo: A nationwide survey. *The Lancet*, 367(9504):44–51, January 2006. ISSN 0140-6736, 1474-547X. doi: 10.1016/S0140-6736(06)67923-3. URL https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(06)67923-3/fulltext?pubType=related.

[23] Working Group for Mortality Estimation in Emergencies. Wanted: Studies on mortality estimation methods for humanitarian emergencies, suggestions for future research. *Emerging Themes in Epidemiology*, 4(1):9, December 2007. ISSN 1742-7622. doi: 10.1186/1742-7622-4-9. URL https://ete-online.biomedcentral.com/articles/10.1186/1742-7622-4-9.

[24] Sarah Bowden, Kai Braker, Francesco Checchi, and Sidney Wong. Implementation and utilisation of community-based mortality surveillance: A case study from Chad. *Conflict and Health*, 6(1):11, November 2012. ISSN 1752-1505. doi: 10.1186/1752-1505-6-11.

[25] P. B. Spiegel, M. Sheik, B. A. Woodruff, and G. Burnham. The accuracy of mortality reporting in displaced persons camps during the post-emergency phase. *Disasters*, 25 (2):172–180, June 2001. ISSN 0361-3666. doi: 10.1111/1467-7717.00169.

[26] Chimeremma Nnadi, Andrew Etsano, Belinda Uba, Chima Ohuabunwo, Musa Melton, Gatei wa Nganda, Lisa Esapa, Omotayo Bolu, Frank Mahoney, John Vertefeuille, Eric Wiesen, and Elias Durry. Approaches to Vaccination Among Populations in Areas of Conflict. *The Journal of Infectious Diseases*, 216(suppl_1):S368–S372, July 2017. ISSN 0022-1899. doi: 10.1093/infdis/jix175. URL https://doi.org/10.1093/infdis/jix175.

[27] Francesco Checchi, Bayard Roberts, and Oliver Morgan. A New Method to Estimate Mortality in Crisis-Affected Populations: Validation and Feasibility Study. Version 2. Technical report, 2009. URL https://www.fantaproject.org/sites/default/files/resources/EM_Method_%20March_2009.pdf.

[28] Bayard Roberts, Oliver W. Morgan, Mohammed Ghaus Sultani, Peter Nyasulu, Sunday Rwebangila, Egbert Sondorp, Daniel Chandramohan, and Francesco Checchi. Economic Feasibility of a New Method to Estimate Mortality in Crisis-Affected and Resource-Poor Settings. *PLOS ONE*, 6(9):e25175, September 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0025175. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0025175.

[29] Bayard Roberts, Oliver W. Morgan, Mohammed Ghaus Sultani, Peter Nyasulu, Sunday Rwebangila, Mark Myatt, Egbert Sondorp, Daniel Chandramohan, and Francesco Checchi. A new method to estimate mortality in crisis-affected and resource-poor settings: Validation study. *International Journal of Epidemiology*, 39(6):1584–1596, December 2010. ISSN 1464-3685. doi: 10.1093/ije/dyq188.

[30] Dennis M. Feehan, Mary Mahy, and Matthew J. Salganik. The Network Survival Method for Estimating Adult Mortality: Evidence From a Survey Experiment in Rwanda. *Demography*, 54(4):1503–1528, August 2017. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-017-0594-y. URL https://read.dukeupress.edu/demography/article/54/4/1503/167730/The-Network-Survival-Method-for-Estimating-Adult.

[31] Mattias Schedwin, Aurélie Bisumba Furaha, Richard Kapend, Pierre Akilimali, Espoir Bwenge Malembaka, Helena Hildenwall, Tobias Alfvén, Thorkild Tylleskär, Mala Ali Mapatano, and Carina King. Under-five mortality in the Democratic Republic of the Congo: Secondary analyses of survey and conflict data by province. *Bulletin of the World Health Organization*, 100(7):422–435, July 2022. ISSN 0042-9686. doi: 10.2471/BLT.22.287915. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9243684/.

[32] United Nations. DR Congo: 'Widespread and systematic' violence linked to clashes over gold, August 2021. URL https://news.un.org/en/story/2021/08/1097762.

[33] Claudine Prudhon and Paul B. Spiegel. A review of methodology and analysis of nutrition and mortality surveys conducted in humanitarian emergencies from October 1993 to April 2004. *Emerging Themes in Epidemiology*, 4(1):10, June 2007. ISSN 1742-7622. doi: 10.1186/1742-7622-4-10. URL https://doi.org/10.1186/1742-7622-4-10.

[34] Olivier Degomme and Debarati Guha-Sapir. Mortality and nutrition surveys by Non-Governmental organisations. Perspectives from the CE-DAT database. *Emerging Themes in Epidemiology*, 4(1):11, June 2007. ISSN 1742-7622. doi: 10.1186/1742-7622-4-11. URL https://doi.org/10.1186/1742-7622-4-11.

[35] Jennifer Scott, Sarah Averbach, Anna Merport Modest, Michele Hacker, Sarah Cornish, Danielle Spencer, Maureen Murphy, and Parveen Parmar. An assessment of attitudes toward gender inequitable sexual and reproductive health norms in South Sudan: A community-based participatory research approach. *Conflict and Health*, 7(1):24, November 2013. ISSN 1752-1505. doi: 10.1186/1752-1505-7-24. URL https://doi.org/10.1186/1752-1505-7-24.

[36] Dennis M. Feehan and Ayesha S. Mahmud. Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nature Communications*, 12(1):

893, February 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-20990-2. URL https://www.nature.com/articles/s41467-021-20990-2.

[37] Shuo Feng, Karen A. Grépin, and Rumi Chunara. Tracking health seeking behavior during an Ebola outbreak via mobile phones and SMS. *npj Digital Medicine*, 1(1): 1–8, October 2018. ISSN 2398-6352. doi: 10.1038/s41746-018-0055-z. URL https://www.nature.com/articles/s41746-018-0055-z.

[38] Holly E. Reed and Charles B. Keely. Understanding Mortality Patterns in Complex Humanitarian Emergencies. In *Forced Migration & Mortality*. National Academies Press (US), 2001. URL https://www.ncbi.nlm.nih.gov/books/NBK223340/.

[39] Francesco Checchi and Les Roberts. Interpreting and using mortality data in humanitarian emergencies: A primer for non-epidemiologists. Technical report, 2005. URL https://odihpn.org/publication/interpreting-and-using-mortality-data-in-humanitarian-emergencies/.

[40] Dennis M. Feehan, Aline Umubyeyi, Mary Mahy, Wolfgang Hladik, and Matthew J. Salganik. Quantity Versus Quality: A Survey Experiment to Improve the Network Scale-up Method. *American Journal of Epidemiology*, 183(8):747–757, April 2016. ISSN 0002-9262. doi: 10.1093/aje/kwv287. URL https://doi.org/10.1093/aje/kwv287.

[41] Peter D. Killworth, Eugene C. Johnsen, Christopher McCarty, Gene Ann Shelley, and H.Russell Bernard. A social network approach to estimating seroprevalence in the United States. *Social Networks*, 20(1):23–50, January 1998. ISSN 03788733. doi: 10.1016/S0378-8733(96)00305-X. URL https://linkinghub.elsevier.com/retrieve/pii/S037887339600305X.

[42] Charles Kadushin, Peter D. Killworth, H. Russell Bernard, and Andrew A. Beveridge. Scale-Up Methods as Applied to Estimates of Heroin use. *Journal of Drug Issues*, 36(2): 417–440, April 2006. ISSN 0022-0426, 1945-1369. doi: 10.1177/002204260603600209. URL http://journals.sagepub.com/doi/10.1177/002204260603600209.

[43] Peter D. Killworth, Christopher McCarty, H. Russell Bernard, Gene Ann Shelley, and Eugene C. Johnsen. Estimation of Seroprevalence, Rape, and Homelessness in the United States Using a Social Network Approach. *Evaluation Review*, 22(2):289–308, April 1998. ISSN 0193-841X, 1552-3926. doi: 10.1177/0193841X9802200205. URL http://journals.sagepub.com/doi/10.1177/0193841X9802200205.

[44] Dennis M. Feehan and Matthew J. Salganik. Generalizing the Network Scale-up Method: A New Estimator for the Size of Hidden Populations. *Sociological Methodology*, 46(1): 153–186, August 2016. ISSN 0081-1750, 1467-9531. doi: 10.1177/0081175016665425. URL http://journals.sagepub.com/doi/10.1177/0081175016665425.

[45] Ian M. Timæus. Estimation of mortality from orphanhood in adulthood. *Demography*, 28(2):213–227, May 1991. ISSN 0070-3370.

[46] Bruno Masquelier. Adult Mortality From Sibling Survival Data: A Reappraisal of Selection Biases. *Demography*, 50(1):207–228, February 2013. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-012-0149-1. URL https://read.dukeupress.edu/demography/article/50/1/207/169659/Adult-Mortality-From-Sibling-Survival-Data-A.

[47] Stéphane Helleringer, Gilles Pison, Almamy M. Kanté, Géraldine Duthé, and Armelle Andro. Reporting Errors in Siblings' Survival Histories and Their Impact on Adult Mortality Estimates: Results From a Record Linkage Study in Senegal. *Demography*, 51(2):387–411, April 2014. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-013-0268-3. URL https://read.dukeupress.edu/demography/article/51/2/387/169432/Reporting-Errors-in-Siblings-Survival-Histories.

[48] Ziad Obermeyer, Julie Knoll Rajaratnam, Chang H. Park, Emmanuela Gakidou, Margaret C. Hogan, Alan D. Lopez, and Christopher J. L. Murray. Measuring Adult Mortality Using Sibling Survival: A New Analytical Method and New Results for 44 Countries, 1974–2006. *PLOS Medicine*, 7(4):e1000260, April 2010. ISSN 1549-1676. doi: 10.1371/journal.pmed.1000260. URL https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000260.

[49] Dennis M. Feehan and Gabriel M. Borges. Estimating Adult Death Rates From Sibling Histories: A Network Approach. *Demography*, 58(4):1525–1546, August 2021. ISSN 0070-3370. doi: 10.1215/00703370-9368990. URL https://doi.org/10.1215/00703370-9368990.

[50] Dennis M. Feehan and Matthew Salganik. Validating survey-based estimates of adult mortality with high-quality vital statistics: Evidence from 27 cities, January 2024. URL https://osf.io/x5ywv.

[51] Michael R. Elliott and Richard Valliant. Inference for Nonprobability Samples. *Statistical Science*, 32(2):249–264, 2017. ISSN 0883-4237. URL https://www.jstor.org/stable/26408228.

[52] Smart Survey. Enquete Nutritionnelle Smart Territoire De Kalemie. Technical report, studfee, 2022. URL https://reliefweb.int/report/democratic-republic-congo/enquete-nutritionnelle-smart-territoire-de-kalemie-province-du-tanganyika-periode-d

[53] David K. Park, Andrew Gelman, and Joseph Bafumi. Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. *Political Analysis*, 12(4):375–385, 2004. ISSN 1047-1987, 1476-4989. doi: 10.1093/pan/mph024. URL https://www.cambridge.org/core/product/identifier/S1047198700000905/type/journal_article.

Supplementary Materials

Novel Estimates Reveal Subnational Heterogeneities in Disease-Relevant Contact Patterns in the United States

Casey F. Breen Saeed Rahman Christina Kay Joeri Smits
Abraham Azar Steve Ahuka Dennis M. Feehan

August 23, 2025

Table of Contents

S1 Formative fieldwork

S1.1 Fieldwork Overview

The network survival method is highly flexible in that it can be used to produce estimated CDRs based on deaths reported in many different kinds of personal networks—friends, co-workers, kin, acquaintances, neighbors, etc. [1]. This flexibility can be advantageous because it means that researchers can adapt the method to different settings and study goals. But it also means that care must be taken in choosing which network to use as the basis of mortality estimates.

To help inform the design of our study, we conducted formative research in the study setting. The main goal of the formative research was to help us pick the specific personal network(s) to ask respondents to report on. However, the formative field work also helped inform several other key study design parameters, such as length of the recall period, method for estimating network size, and transit hubs (e.g., ports, taxi stands) for sampling respondents.

Our formative fieldwork was conducted in two stages. In the first stage, we conducted eight two-hour focus groups, each with four to eight participants. Our focus groups were split up by age and gender to maximize participant participation. For instance, if both younger and older men were placed in a group, cultural norms would dictate that only the older men would dominate discussions. We conducted four focus groups in the relatively urban Kalemie City and four focus groups in a rural village of Tabac Congo.

In these focus groups, we asked respondents a series of open-ended questions on how they learned about deaths in their community. Using a translator, we conducted the interviews in either French and Swahili depending on group preference. We used the scripted questions below, probing or asking follow-up questions as necessary:

1. How do people in your community learn about deaths? How do you personally learn about deaths?

2. Generally, how long after a death does it take to learn about it?

3. Do you learn about deaths from in-person conversations? From phone calls? Social media? Text message?

4. How well-informed are people about the details of the death? (Age/sex/cause/etc.)

5. Are deaths stigmatized at all? Is there any reason people would not report deaths?

Next, we asked a series of questions about different candidate personal networks. We tested different social ties, including (i) people you have had a meal with in the past year; (ii) people you talk to once a week; (iii) blood-related kin; (iv) immediate neighbors; and (v) acquaintances you talked to in the last year. For each tie definition, we asked questions on the following topics:

1. Under [tie definition], can you directly count how many people you know? If not, can you guess how many people you know under [tie definition]? How confident do you feel in your answer?

2. Under [tie definition], what kind of people are you connected to? Similar people? Random other people?

3. How much do you know about other people you are connected to through [tie definition]?

4. Would you know if someone in this network passed away in the past one month? Three months? Six months?

5. For people you are connected to by [tie definition], are you more likely to know whether certain groups of people died (men vs. women, young vs. old)?

Respondents were prompted to first directly answer the questions and then to engage in a broader discussion with other respondents. From these discussions, several insights emerged. First, respondents in nearly all focus groups reported learning about deaths predominantly from word of mouth. Social media, especially commemorative posts on WhatsApp status updates, was another common way for respondents to learn about deaths in urban, but not rural, settings. In every focus group, respondents reported a high degree of certainty when reporting on deaths occurring in their extended kin or their immediate neighbors.

Respondents often gave nonsensically large answers when they were asked how many people they had a meal with in the past month or how many people they knew in groups of known size (e.g., how many teachers do you know?). This suggested that respondents were better able to count the number of people they were connected to in stronger social connections, such as blood-related kin.

The formative fieldwork also gave us the opportunity to gain insight into several other key study parameters, including the recall period. While in some humanitarian emergencies, circumstances may dictate the length of the recall window (e.g., the month directly following an earthquake), in more protracted humanitarian emergencies, this is a parameter researchers can vary. The choice of a recall period is important, as a recall period that stretches too far into the past may reduce respondents' ability to accurately recall and report about deaths. On the other hand, asking about too short a recall period may result in not enough information about deaths being collected to accurately estimate death rates.

S1.2 Insights from formative fieldwork

The qualitative data suggested that it would be valuable to use a significant and memorable reference event to start the recall period. Such locally recognizable events help respondents more accurately recall dates of death or approximate periods when deaths occurred; studies have used New Year's Day, Ramadan, and even the death of a prominent political figure [2]. Our qualitative research indicated that New Year's Day was a very salient event in this setting, helping respondents accurately determine whether a death occurred before or after this date. Based on this finding, we selected New Year's Day, January 1st, 2023, as our reference event.

These qualitative data also suggested that respondents were able to accurately report about deaths occurring in their extended kin network and their network of immediate neighbors. We used these insights to draft a preliminary set of survey questions and conducted 18 individual cognitive interviews. In these individual interviews, we asked respondents to talk out loud through answers and explain their rationale for their answers. This led to a series of minor wording changes and clarifications of definitions (e.g., being more explicit in our wording that kin only includes blood relatives). For example, to help respondents count the number of people they were connected to in these networks, we broke down categories into smaller groups (e.g., number of female cousins age 0–4, number of male cousins age 0–4, number of female cousins age 5–18, number of male cousins age 5–18, number of female cousins 18+, number of male cousins 18+).

Respondents reported being able to accurately report on the size of their kin and immediate neighbor networks using this approach; this suggests that their total network size could be estimated using an approach called the *summation method*, which asks respondents to report on the number of people they are connected to in specific discrete categories and then sums those reports up to get an estimate of total personal network size [3, 4]. The advantage of the summation method is that it helps break down a personal network such as "extended kin" into subgroups that are easier to count. In the context of this study, it was particularly helpful as respondents often struggled to count the number of people they were connected to in larger groups.

To summarise, we conducted formative field work in our focal health zones. This formative research revealed that respondents were confident they could accurately report on deaths occurring after New Year's in two of their personal networks: their extended kin network and their immediate neighbor network. Further, respondents reported being able to confidently report on deaths after New Year's and the size of extended kin networks and immediate neighbor networks using the summation method.

S2 Study design

S2.1 Quota survey

The quota survey was designed to test our new network-based approach by asking respondents to report on mortality in their kin network and in their neighbor network. The frame population—the universe of people eligible to respond to the survey—was all adults over age 18 who reported living in one of the three focal health zones: Nyunzu, Nyemba, and

Kalemie. We used a non-probability, quota-based sampling strategy to sample respondents at major transit hubs, such as ports, markets, taxi stands, foot paths, and health clinics in Kalemie City, the capital of Tanganyika Province. We chose this diverse set of transit hubs in hopes of sampling as representative a sample as possible. The number of interviews per site type is shown in Table S1.

Site type	n
Health facility	358
Market	833
Other transport	1136
Port	113
Taxi	211

Table S1: Study Sites

Our quotas specified a target number of respondents in cells defined by gender and by all of the health areas[1] that lie in Nyunzu, Nyemba, and Kalemie. These quotas were established based on available population data from the Ministry of Health using vaccination campaign micro-planning information.

A total of 2,526 interviews were conducted from March 1st, 2023 to June 29th, 2023. We emphasize that we recommend using probability sampling wherever possible; however, in this study our goal was to explicitly test this non-probability quota sampling design, because it is the kind of data collection strategy that would be feasible during a humanitarian crisis.

The quota survey proceeded as follows (see Section S6 for the full survey instrument). After obtaining informed consent, respondents were asked a series of screener questions to determine eligibility for the survey. If respondents were eligible to participate in the survey based on quotas, they first answered a series of questions about their demographic and socioeconomic characteristics. Respondents were asked about age, sex, education level, occupation, and a set of questions to construct a wealth index: owning a bed, owning a radio, material of the exterior walls of their dwelling unit, and primary fuel used for cooking.

Next, respondents were asked to report on deaths in their kin and neighbor networks in

[1] There are currently 26 provinces in DRC. These provinces are subdivided into a total of 519 health zones (also called Zones de Santé), and each health zone is further divided into Health Areas (also called Aires de Santé). See https://data.humdata.org/dataset/drc-health-data for more information.

separate modules. We selected these two networks based on a series of focus groups and cognitive interviews conducted to determine the specific social ties that respondents could accurately report on. The order of the kin and neighbor modules were randomly assigned to allow us to assess potential question ordering effects.

Respondents reported on the number of connections they had in different subgroups (e.g., "How many male cousins do you have under the age of 5?"). We then immediately asked respondents to report on the number of deaths in these groups. In both modules, we broke these questions into finer subcategories to reduce cognitive load on respondents and improve the accuracy of reporting. After completing the network modules, respondents were asked a series of questions about births, migration, measles, and cholera in their personal networks. Finally, if respondents reported a death in their household, they were asked a series of detailed follow-up questions about the timing and an abbreviated WHO verbal autopsy.

S2.2 Recall period

Respondents were asked to report on deaths occurring between the reference date, January 1st, 2023, and the interview date. Since the network survey was in the field for four months, a respondent's recall period varied depending on their interview date. Our estimator compensates for different recall periods among respondents by including a term E_i, representing the total exposure days each respondent reported about their personal networks (Equation 3). We estimate the total amount of exposure reported by a respondent, in person-days, by taking the product of the length in days of the respondent's recall period and their personal network size.

Notably, this rolling recall period resulted in more reports about deaths and exposure at the beginning of the observation window and fewer towards the end. For instance, respondents interviewed in March could not report on deaths in June. We pooled information on all deaths and exposure from January 1st, 2023, to June 29th, 2023, assuming that the CDR remained constant throughout the period. This assumption seems reasonable, because we found negligible changes in estimated death rates over time. However, in different contexts with a stronger time trend in mortality, researchers might need to produce separate estimates for shorter time periods and average them together.

S2.3 Crude death rate units

We chose to express our CDR as deaths per 10,000 people per day. This contrasts with units more commonly used in demography: deaths per 1,000 people per year. To convert the CDRs reported in this paper to units of deaths per 1,000 people per year, simply multiply by 36.5.

1. As there are 365 days in a year, to convert from a daily rate to an annual rate, we multiply the CDR by 365.

2. The humanitarian CDR is expressed per 10,000 people, whereas the demographic CDR is expressed per 1,000 people. To account for this, we divide by 10.

The conversion factor is therefore calculated as:

$$\text{Conversion factor} = \frac{365 \text{ days per year}}{10} = 36.5 \tag{S1}$$

The rationale for expressing the CDR in these units is twofold. First, conditions in humanitarian disasters can fluctuate significantly on a daily basis, and so CDRs are often calculated for a time periods much shorter than a year (in contrast to conventional demographic CDRs, which are typically calculated for a year); estimating mortality over shorter time periods can capture these fluctuations better than an annual measure [5]. Second, this daily CDR is used as the basis for defining humanitarian emergencies; for example, the Center for Disease Control (CDC) defines a humanitarian crisis as more than 1 death per 10,000 persons per day [2].

S2.4 Mortality clustering within social networks

Mortality clustering—the non-random concentration of deaths within specific groups, such as households, social networks, and villages—has potential implications for the network survival method estimates presented in this paper. In certain settings, mortality may exhibit stronger clustering within extended networks than within households. For instance, for certain infectious diseases, deaths may cluster among immediate neighbors who spread the disease to each other. Whether clustering is greater within extended networks or within households will ultimately be context-specific.

This potential clustering should not introduce bias into the point estimates but does lead to greater variability and increased uncertainty. Our bootstrap resampling procedure explicitly captures this uncertainty and accounts for the clustering of mortality within networks. We recommend that all future studies incorporating the network survival method implement a similar bootstrap procedure for uncertainty quantification.

S2.5 Probability survey

In addition to the quota survey, we collected a probabilistic, retrospective household mortality survey (probability survey) administered between July 24$^{\text{th}}$, 2023 and September 2$^{\text{nd}}$, 2023. We sample 2,785 households from our focal health zones of Nyunzu, Nyemba, and Kalemie. The sampling frame was constructed from population data from the Ministry of Health derived from vaccination campaign micro-planning information. Using these population data, we defined primary sampling units, generally at the village level. We randomly sampled 38 primary sampling units in Kalemie,[2] 40 primary sampling units in Nyunzu, and 44 primary sampling units in Nyemba. Within primary sampling units, households were selected using random sampling. Within households, the household head, or in their absence another adult over the age of 18, was surveyed. The household survey asked detailed information about deaths occurring within their household after January 1$^{\text{st}}$, 2023 and the full set of network method questions. If respondents reported a death within their own household, a supervisor then followed up the same day with a verbal autopsy questionnaire to collect detailed information about the cause of the death.

We consider the estimate from the probability survey to be valuable as a comparator, but we note that it is not a gold standard. Like all estimates based on a retrospective household survey, our estimates may be prone to different sources of error, including sampling error, response errors, and frame errors that may affect the accuracy of the household-based estimate [6–9].

[2] Due to insecurity in parts of Kalemie, we were only able to access 38 PSUs in the Kalemie health zone because of security issues in the southern areas at the time of the survey.

S2.6 Weighting strategies

For our quota survey, we used a non-probability quota sample with quotas based on gender and health area. This design led our respondents to match the overall gender and geographic distribution of our target population very closely. However, because our design did not choose respondents probabilistically, there are still ways that selection bias may affect the composition of our survey respondents. Specifically, our sample overrepresents higher SES individuals and middle-aged respondents (Figure S1). Given the observable selection into our sample, we develop a few different weighting strategies to adjust for potential selection into our network survey sample.[3]

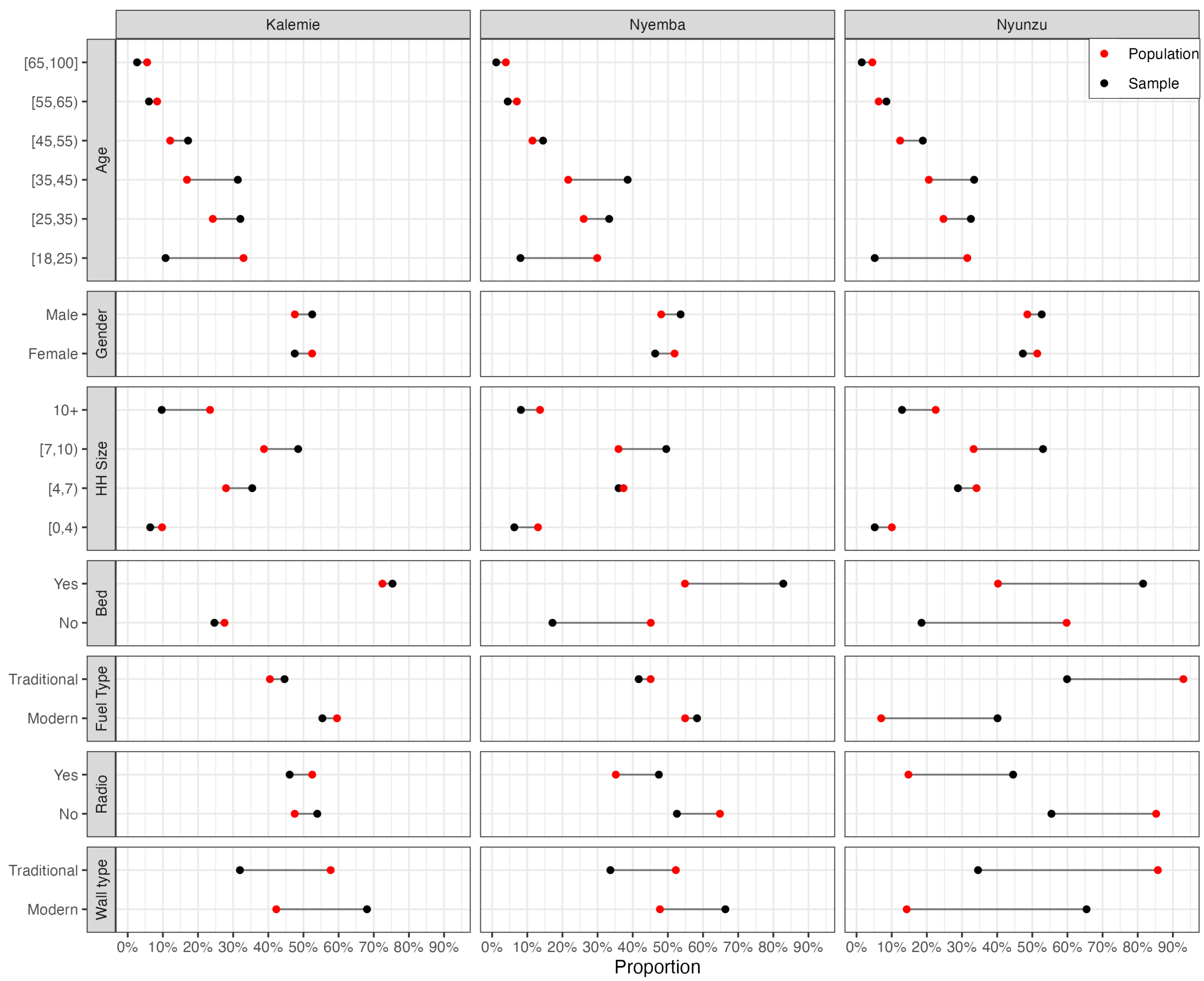

Figure S1: Difference in respondent composition between quota and household samples.

[3]The framework we adopt for inference from a non-probability sample is sometimes called *quasi-randomization* [10].

We construct three different sets of estimates using different weighting strategies imitating different data availability settings. Our first set of estimates are unweighted. This unweighted strategy relies exclusively on our quota sample based on geographic region (health area, the geographic unit beneath health zones) and gender. This gives us a baseline set of estimates not adjusting for any of the selection into the sample.

Our second set of estimates imitates a setting where no auxiliary data specific to our setting is available to help construct survey weights. Instead, we use modeled data from the WorldPop gridded population estimates—which are available all over the world—to construct poststratification weights [11]. We use the 2020 age and gender-structured, gridded cells with a resolution of 100m. We construct weighting targets by taking the intersection of these gridded cells with administrative boundaries for each of the three health zones using administrative boundaries from the GADM project. We then construct post-stratification weights on the following cells: age (18–24, 25–34, 35–44, 45–54, 55–64, 65+), gender (female and male), and health zone (Nyunzu, Nyemba, and Kalemie).

For our final set of estimates, we construct survey weights using logistic regression to model inclusion probability. We use our household survey, which in this setting represents the most accurate set of reference estimates. There was no other representative household survey large enough to serve as a reliable reference survey. In other settings, a recent household survey, such as a Demographic and Health Survey (DHS), might serve as a reliable reference. To take advantage of this, researchers must design their non-probability survey instrument carefully to ensure harmonization with the reference survey. Question wording for sociodemographic and household questions was identical between the quota and probability surveys.

Specifically, we combined together our quota and household surveys, and fit models to estimate inclusion probability:

$$w_i = \frac{1}{\hat{P}(S_i = 1)} \tag{S2}$$

where w_i is a weight defined as the inverse probability of being included in the sample ($S_i = 1$). We estimate three separate regression models, one for each health zone, using the following specification:

$$
\begin{aligned}
\text{logit}(\Pr(\text{inclusion} = 1|\mathbf{X})) = &\beta_0 + \beta_{(\text{gender})} + \beta_{(\text{age class})} + \beta_{(\text{hh size})} \\
&\beta_{(\text{radio})} + \beta_{(\text{bed})} + \beta_{(\text{wall material})} + \beta_{(\text{modern fuel type})} + \\
&\beta_{(\text{hh count age 0–4})} + \beta_{(\text{hh count age 5–17})} + \beta_{(\text{hh count age 18+})}
\end{aligned} \tag{S3}
$$

where inclusion denotes the dependent variable indicating whether an individual is included within a specific zone. Independent variables comprise both continuous and categorical predictors: gender (male, female), age class (18–24, 25–34, 35–44, 45–54, 55–64, 65–100), household size (0-3, 3-6, 7+), household possession of a radio, household possession of a bed, household having a modern constructed wall type, household's primary fuel source being modern, and number of household members under age 5, between age 5 and 18, and over age 18. The regression coefficients and goodness of fit statistics are reported in Table S2.

	Kalemie	Nyemba	Nyunzu
(Intercept)	0.044***	0.016***	0.004***
	(0.012)	(0.005)	(0.002)
Gender (Male)	1.166	1.140	0.894
	(0.108)	(0.106)	(0.116)
Age class [25,35)	3.077***	3.473***	5.341***
	(0.445)	(0.536)	(1.312)
Age Class [35,45)	5.882***	5.009***	7.884***
	(0.889)	(0.791)	(1.980)
Age Class[45,55)	5.112***	4.405***	12.363***
	(0.860)	(0.803)	(3.318)
Age Class[55,65)	2.943***	2.579***	16.544***
	(0.618)	(0.617)	(5.279)
Age Class[65,100]	2.018*	1.715	3.865**
	(0.558)	(0.652)	(1.798)
Household Size Size 4–7	0.879	1.101	0.796
	(0.197)	(0.256)	(0.273)
Household Size 7–10	0.718	1.076	0.810
	(0.181)	(0.283)	(0.309)
Household Size 10+	0.194***	0.444**	0.192***
	(0.056)	(0.135)	(0.080)
Owns radio	0.645***	1.356**	2.411***
	(0.067)	(0.136)	(0.354)
Owns Bed	1.258	3.790***	3.300***
	(0.150)	(0.440)	(0.483)
Modern House Material	4.189***	2.325***	6.918***
	(0.425)	(0.259)	(0.946)
Use Modern Fuel	0.821	0.484***	4.129***
	(0.090)	(0.056)	(0.681)
Under 5 Count (1)	1.674***	2.314***	2.095***
	(0.216)	(0.301)	(0.410)
Under 5 Count (2+)	4.987***	3.621***	8.318***
	(0.646)	(0.474)	(1.643)
Age 5–18 Count (1)	0.846	0.715	1.133
	(0.161)	(0.143)	(0.333)
Age 5–18 Count (2+)	0.900	1.227	1.787*
	(0.165)	(0.241)	(0.485)
Age 18+ Count (2+)	0.965	0.999	0.377***
	(0.217)	(0.249)	(0.110)
Num.Obs.	3250	3087	3210
AIC	2417.6	2232.4	1686.0
BIC	2533.2	2347.1	1801.5
RMSE	0.38	0.39	0.28

*p <0.05, ** p <0.01, *** p <0.001

Table S2: Logistic regression predicting odds of inclusion in the probability sample. Coefficients report odds ratios.

For the probability survey, we do not use survey weights. The probability sample was intended to produce a self-weighting sample, and in the absence of any other high-quality assessment, our probability survey is the most reliable source of population composition estimates available. In settings where high-quality auxiliary data is available, we recommend reweighting the probability sample to account for non-response and other biases.

S2.7 Blended network estimates

We produced separate estimates using reports about neighbor and kin networks. In addition, we use a blended estimator to produce a combined estimate based on both the kin and the neighbor network reports [12]. The advantage of this blended approach is that we expect it to produce smaller mean squared error (MSE) than either the kin or neighbor estimate alone, because the blended estimate is based on more information. But this comes at the cost of additional assumptions; see [12] for a detailed discussion.

The blended estimate is based on averaging together the estimate from each network in a principled way. Suppose we have two estimators for N, $\widehat{N}^A$ and $\widehat{N}^B$. We define the blended estimate with pooling weight θ as:

$$\underbrace{\widehat{N}}_{\text{Blended Estimator}} = \underbrace{\theta \widehat{N}^A}_{\text{Weighted Estimator A}} + \underbrace{(1-\theta)\widehat{N}^B}_{\text{Weighted Estimator B}} \tag{S4}$$

where $\theta \in \mathbb{R}$.

Given estimates of the sampling variance for the two estimates, and assuming that both estimators $\widehat{N}^A$ and $\widehat{N}^B$ are unbiased, we can calculate the weight $\theta^\star$ that minimizes the expected mean squared error as:

$$\theta^\star = \frac{\sigma_B^2 - \sigma_{AB}}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}}, \tag{S5}$$

where σ_A^2 is the sampling variance of estimator $\widehat{N}^A$, σ_B^2 is the sampling variance of estimator $\widehat{N}^B$, and $\sigma_{AB} = cov(\widehat{N}^A, \widehat{N}^B)$ is the covariance of estimator $\widehat{N}^A$ and $\widehat{N}^B$. The blending weights given by Equation S5 are the ones we use to blend estimates in the main text; a full derivation is in Section S2.7.1.

Future studies may have more information about the bias of estimators, perhaps from validation studies. In that case, it would be helpful to have weights that can be used to blend biased estimates together, accounting for the bias. Section S2.7.2 derives another optimal weight in this more general situation:

$$\theta^\star = \frac{\sigma_B^2 - \sigma_{AB} + \beta_B(\beta_B - \beta_A)}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB} + (\beta_A - \beta_B)^2}, \tag{S6}$$

where $\beta_A = \mathbb{E}[\widehat{N}_A - N]$ is the bias of estimator $\widehat{N}_A$ and $\beta_B = \mathbb{E}[\widehat{N}_B - N]$ is the bias of estimator $\widehat{N}_B$.

S2.7.1 Derivation of optimal weight for blended estimator assuming each estimator is unbiased

Here we will derive the blending weight $\theta^\star$ that minimizes the mean squared error (MSE) of our blended estimate (Equation S5). We consider two estimators for N, denoted $\widehat{N}^A$ and $\widehat{N}^B$, which will be combined with a pooling weight θ, as in Equation S4. We assume $\widehat{N}^A$ and $\widehat{N}^B$ are unbiased for N, meaning that

$$\mathbb{E}[\widehat{N}^A - N] = 0 \text{ and } \mathbb{E}[\widehat{N}^B - N] = 0.$$

The blended estimator with blending weight θ has MSE

$$MSE(\hat{N}) = \mathbb{E}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)^2\right]. \tag{S7}$$

In general, for a random variable X, we have $\mathbb{E}[X^2] = \mathbb{E}[X]^2 + \text{Var}[X]$. Applying this relationship to the MSE, we obtain

$$MSE(\hat{N}) = \left(\mathbb{E}\left[\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right]\right)^2 + \text{Var}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)\right]. \tag{S8}$$

By assumption, our estimators $\widehat{N}^A$ and $\widehat{N}^B$ are unbiased for N, so $\mathbb{E}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)\right] = 0$. This leaves us with just the variance term

$$MSE(\hat{N}) = \text{Var}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)\right] = \text{Var}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B\right)\right]. \tag{S9}$$

This can be further simplified using properties of variance. If we let $\text{Var}\left[\widehat{N}^A\right] = \sigma_A^2$, $\text{Var}\left[\widehat{N}^B\right] = \sigma_B^2$, and $\text{Cov}\left[\widehat{N}^A, \widehat{N}^B\right] = \sigma_{AB}$, this will simplify to:

$$
\begin{aligned}
MSE(\hat{N}) &= \text{Var}\left[\theta\widehat{N}^A\right] + \text{Var}\left[(1-\theta)\widehat{N}^B\right] + 2\text{Cov}\left[\theta\widehat{N}^A, (1-\theta)\widehat{N}^B\right] && \text{(S10)}\\
&= \theta^2\text{Var}\left[\widehat{N}^A\right] + (1-\theta)^2\text{Var}\left[\widehat{N}^B\right] + 2\theta(1-\theta)\text{Cov}\left[\widehat{N}^A, \widehat{N}^B\right] && \text{(S11)}\\
&= \theta^2\sigma_A^2 + (1-\theta)^2\sigma_B^2 + 2\theta(1-\theta)\sigma_{AB}. && \text{(S12)}
\end{aligned}
$$

To find the blending weight $\theta^\star$ that is optimal in the sense that it minimizes the $MSE(\widehat{N})$, we will take the derivative of Equation S12 with respect to θ, set it equal to 0, and solve for the optimum, $\theta^\star$.

$$
\begin{aligned}
\frac{\partial MSE(\hat{N})}{\partial\theta} &= 2\theta\sigma_A^2 - 2(1-\theta)\sigma_B^2 + 2\sigma_{AB} - 4\theta\sigma_{AB} = 0\\
&\Leftrightarrow 2\theta\sigma_A^2 - 2\sigma_B^2 + 2\theta\sigma_B^2 + 2\sigma_{AB} - 2\theta\sigma_{AB} = 0\\
&\Leftrightarrow \theta\left(\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}\right) = \sigma_B^2 - \sigma_{AB}\\
&\Leftrightarrow \theta^\star = \frac{\sigma_B^2 - \sigma_{AB}}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}}
\end{aligned}
$$

This derivation goes beyond past results on blended estimates from Feehan et al. [12] by relaxing an important assumption: that both estimates are independent.

S2.7.2 Derivation of optimal weight for blended estimator assuming each estimator is biased

In this section, we provide an expression for the blending weight that minimizes MSE if the estimators are biased. Although we do not apply these results in our study, they may prove useful in future studies where bias has been measured, perhaps using validation study designs. The results in this section nest the results in the previous section in the special case where the bias of the two estimators is 0.

Assume the same setup as in Section S2.7.1, except that $\widehat{N}^A$ and $\widehat{N}^B$ may be biased. Let the bias of $\widehat{N}^A$ be $\beta_A = \mathbb{E}[\widehat{N}^A - N]$ and let the bias of $\widehat{N}^B$ be $\beta_B = \mathbb{E}[\widehat{N}^B - N]$.

Equation S8 showed that the MSE of the blended estimator with blending weight θ can be written

$$MSE(\hat{N}) = \left(\mathbb{E}\left[\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right]\right)^2 + \text{Var}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)\right]. \tag{S13}$$

However, unlike the derivation in the previous section, here we do not assume that the two estimators $\widehat{N}^A$ and $\widehat{N}^B$ are unbiased. This means that the first term of Equation S13 is not zero. Instead, it is the squared bias of the blended estimator:

$$\begin{aligned}
\left(\mathbb{E}\left[\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right]\right)^2 &= \left(\mathbb{E}\left[\theta\widehat{N}^A - \theta N + (1-\theta)\widehat{N}^B - (1-\theta)N\right]\right)^2 && \text{(S14)}\\
&= \left(\theta\mathbb{E}\left[\widehat{N}^A - N\right] + (1-\theta)\mathbb{E}\left[\widehat{N}^B - N\right]\right)^2 && \text{(S15)}\\
&= \left(\theta\beta_A + (1-\theta)\beta_B\right)^2. && \text{(S16)}
\end{aligned}$$

The second part of Equation S13 is the variance of the blended estimator; this term is unchanged. As Section S2.7.1 showed, the variance term can be written

$$\text{Var}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)\right] = \theta^2\sigma_A^2 + (1-\theta)^2\sigma_B^2 + 2\theta(1-\theta)\sigma_{AB}. \tag{S17}$$

Combining the squared bias and the variance using Equation S13, we find that the MSE will be

$$\begin{aligned}
MSE &= \left(\mathbb{E}\left[\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right]\right)^2 + \text{Var}\left[\left(\theta\widehat{N}^A + (1-\theta)\widehat{N}^B - N\right)\right] && \text{(S18)}\\
&= \theta^2\sigma_A^2 + (1-\theta)^2\sigma_B^2 + 2\theta(1-\theta)\sigma_{AB} + \left(\theta\beta_A + (1-\theta)\beta_B\right)^2. && \text{(S19)}
\end{aligned}$$

To find the value of θ that minimizes the MSE, we take the derivative with respect to θ, set it equal to zero, and solve for $\theta^\star$:

$$
\begin{aligned}
\frac{\partial MSE(\hat{N})}{\partial \theta} &= 2\theta\sigma_A^2 - 2(1-\theta)\sigma_B^2 + 2\sigma_{AB} - 4\theta\sigma_{AB} + 2(\theta\beta_A + (1-\theta)\beta_B)(\beta_A - \beta_B)0 \quad \text{(S20)} \\
&= 2\theta\sigma_A^2 - 2\sigma_B^2 + 2\theta\sigma_B^2 + 2\sigma_{AB} - 4\theta\sigma_{AB} + 2\theta\beta_A^2 + 2\beta_A\beta_B - 2\theta\beta_A\beta_B - 2\theta\beta_A\beta_B - 2\beta_B^2 + 2\theta\beta_B^2 \\
&\qquad \text{(S21)} \\
&= 2\theta\left(\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB} + \beta_A^2 - 2\beta_A\beta_B + \beta_B^2\right) - 2\left(\sigma_B^2 - \sigma_{AB} - \beta_A\beta_B + \beta_B^2\right). \\
&\qquad \text{(S22)}
\end{aligned}
$$

Setting this expression equal to 0 and solving for the minimizer, $\theta^\star$, we have

$$
\begin{aligned}
\theta^\star\left(\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB} + \beta_A^2 - 2\beta_A\beta_B + \beta_B^2\right) &= \sigma_B^2 - \sigma_{AB} - \beta_A\beta_B + \beta_B^2 \\
\iff \theta^\star &= \frac{\sigma_B^2 - \sigma_{AB} + \beta_B(\beta_B - \beta_A)}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB} + \beta_A^2 - 2\beta_A\beta_B + \beta_B^2} \\
\iff \theta^\star &= \frac{\sigma_B^2 - \sigma_{AB} + \beta_B(\beta_B - \beta_A)}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB} + (\beta_A - \beta_B)^2}.
\end{aligned}
$$

Note that, plugging in $\beta_A = 0$ and $\beta_B = 0$, we recover the weight derived for unbiased estimators in the previous section.

We can confirm that $\theta^\star$ is a minimum by differentiating Equation S22 again to obtain

$$
\begin{aligned}
\frac{\partial^2 MSE(\hat{N})}{\partial \theta^2} &= 2\left(\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB} + \beta_A^2 - 2\beta_A\beta_B + \beta_B^2\right) && \text{(S23)} \\
&= 2(\beta_A - \beta_B)^2 + 2(\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}). && \text{(S24)}
\end{aligned}
$$

When Equation S24 is greater than 0, $\theta^\star$ will be a minimum. The first term in parentheses, $(\beta_A - \beta_B)^2$ will be greater than zero except in the special case where the bias of the two estimators is identical, i.e., $\beta_A = \beta_B$, which will make the term zero. The second term in parentheses, $\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}$ is equal to the variance of the difference between the two estimators, $\text{var}(\widehat{N}_A - \widehat{N}_B)$. As a variance, this is always greater than or equal to zero, and will equal zero only when $\widehat{N}_A = \widehat{N}_B$. Thus, we conclude that Equation S24 is greater than zero except for the pathological case where $\beta_A = \beta_B$ and $\widehat{N}_A = \widehat{N}_B$.

In our study, the $\theta^\star$ used for blending is always in $[0, 1]$, meaning that it will produce an estimated value in-between $\widehat{N}^A$ and $\widehat{N}^B$. We expect this to be true in most applied settings, but this is not guaranteed; future work could explore when the blending weight will be outside that range.

S2.8 Comparisons with other studies

The Jarrett et al. [9] study took place in the Fizi health zone in South Kivu, which is directly above our focal health zones (Figure S2). The study combined data from both a surveillance program and a retrospective household mortality study. The surveillance program had a recall period of November 1st 2011 to September 30th 2012; the presidential election took place on November 2011, making November a salient reference date. For brevity, we only discuss the mortality estimation component of the study.

The retrospective household survey took place from August 29th, 2012 to September 14th, 2012. The recall period was from November 1st, 2011 until the day of the interview, a period approximately equivalent to the surveillance program recall period. Any discrepancies between the household survey and the surveillance site (i.e., death event reported in one system and not in the other) was investigated in a re-evaluation process. In this re-evaluation, enumerators visited households and asked a series of questions to validate whether a reported event had actually occurred.

The study derived a gold-standard estimate, which used deaths that either (1) matched in both the household survey and surveillance systems or (2) were confirmed as a true death in the re-evaluation stage. The study found 23 true deaths and 38 false positive death reports in the household survey. Of these false positive reports, 12 deaths were outside of recall bounds, 18 deaths were not within the household, and 8 deaths were simply fabricated.

The magnitude of discrepancies is relevant to our study. The respondents sampled here, much like in our focal health zones, may have an incentive to report their situation as being particularly aid-worthy. While the study was conducted approximately 12 years before our study, such overreporting dynamics are also possible in our survey.

Figure S2: This map shows the geographic proximity of our focal health zones to Fizi, the territory considered in Jarrett et al.[9].

Another comparison is the household SMART Survey administered in November 2022 in the Kalemie Health Zone [13]. To facilitate a more direct comparison, we compare our estimates for the Kalemie Health Zone to the SMART Survey in Figure S3.

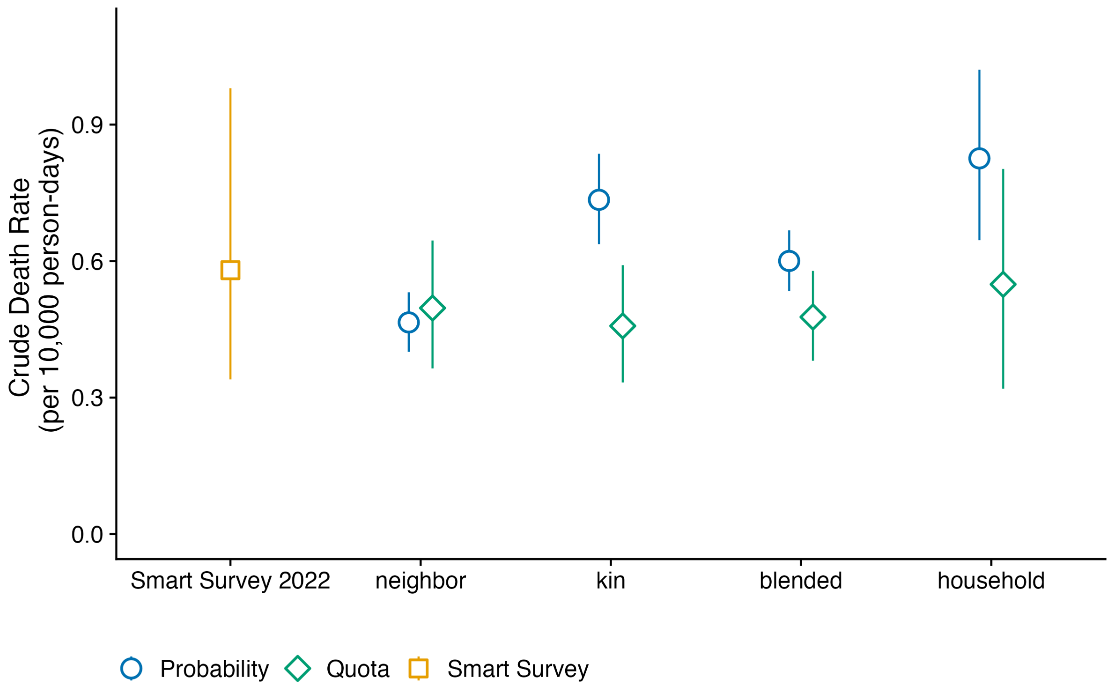

Figure S3: Direct comparison of estimates from our study to external estimates from a 2022 household SMART Survey.

S2.9 Investigating sources of discrepancies

To better understand the potential reasons for the discrepancy between the household and network estimates in our probability sample, we investigate two potential sources of bias: transmission error and strategic overreporting. We compare within the probability sample to control for differences due to sampling design (network-based estimates from probability and non-probability samples are very similar).

Transmission error refers to violations of the perfect visibility assumption—that is, respondents not knowing about a death in the network they are reporting on. In the context of this study, one candidate explanation is differential transmission error: respondents might have more accurate recall for household deaths compared to deaths in their broader social network. Specifically, there would need to be under-reporting of deaths in the neighbor and kin network due to respondents not knowing about deaths that had occurred—but not in their own household, leading to an underestimation of CDRs in the network estimates.

To get a better sense of the extent of transmission error, we compare the household estimate (0.81) to the neighbor estimate (0.40). If we assume the household estimate is correct, respondents would need to miss reporting 51% of the deaths that occurred in their neighbor network. This seems implausible, given that our qualitative research indicates respondents were both (1) confident in their ability to know about and report deaths in neighboring households and (2) expressed no reluctance to report on their neighbors.

Another candidate explanation for this discrepancy is strategic overreporting of deaths in the household. It is possible that respondents in the probability survey are over-reporting deaths in attempts of making their situation appear more aid-worthy. If the kin CDR estimate is correct, how much strategic overreporting (i.e., false positives) of household deaths would be needed to get our household CDR estimate? For every real death, respondents would need to falsely report 1.05 additional deaths, meaning that only 48% of reported deaths actually occurred in the household during the reporting window. While high, this is substantially lower than the 72% false positive rate found in Jarrett et al. 2020 [9]. This suggests that strategic overreporting is plausible in this setting.

These two calculations give a rough sense of how plausible transmission error and strategic overreporting are as explanations for differences between the household and network estimates. We focused on extreme cases in which one factor alone affects one of the estimates at a time. But in reality, a complex combination of factors, including transmission error and strategic overreporting, could lead to errors in either survey-based estimate. Future work should focus on validation designs that compare mortality estimation methods in a setting where gold-standard death rates are available to better understand the properties of both estimators.

S2.10 Ethical considerations of collecting network survival data

In designing our survey, we recognized the ethical challenges associated with reporting deaths of individuals outside the household, given that household surveys typically assume a household member can ethically report on all members within the same household. To address these concerns, we implemented several safeguards to protect respondent ethical compliance:

1. All collected data were anonymized, and no personally identifiable information (e.g.,

names, street addresses) was gathered. Data were securely stored on password-protected laptops and released to the study team only after ensuring anonymity, eliminating any risk of re-identification.

2. Respondents were explicitly informed during the consent process that the survey would be asked about the deaths of neighbors and extended kin, ensuring their awareness and voluntary participation.

3. Our study received ethics approval from the UC Berkeley Institutional Review Board (IRB) and local IRB approval from the University of Kinshasa.

We encourage future researchers and practitioners to adopt similar ethical safeguards when applying the methods introduced in this study.

S3 Validity checks and internal consistency checks

S3.1 Network survival method: internal consistency checks

One advantage of the network method is its potential for partial self-validation. Certain relationships are naturally reciprocal, and we can use this expectation to check for consistency. For example, sibling relationships should be reciprocal. Assuming a perfect probability sample and accurate reporting, we would expect in aggregate, men in our sample would report the same number of connections to sisters as women would report connections to brothers.

As an internal validity check, we compare three relationships in Figure S4 we would expect to be reciprocal: parent child–relationships, sibling relationships by gender, and cousin relationships by gender. We restrict to reported adults over 18, as we only sample adults over aged 18. In our unweighted results, there are small differences: the total number of female reports to brothers are slightly greater than the total number of male reports to sisters. However, our inverse-probability weighted results show nearly perfect reciprocity across all three relationship comparisons. The close alignment between expected reciprocal relationships provides strong evidence for the internal validity and reliability of our network data.

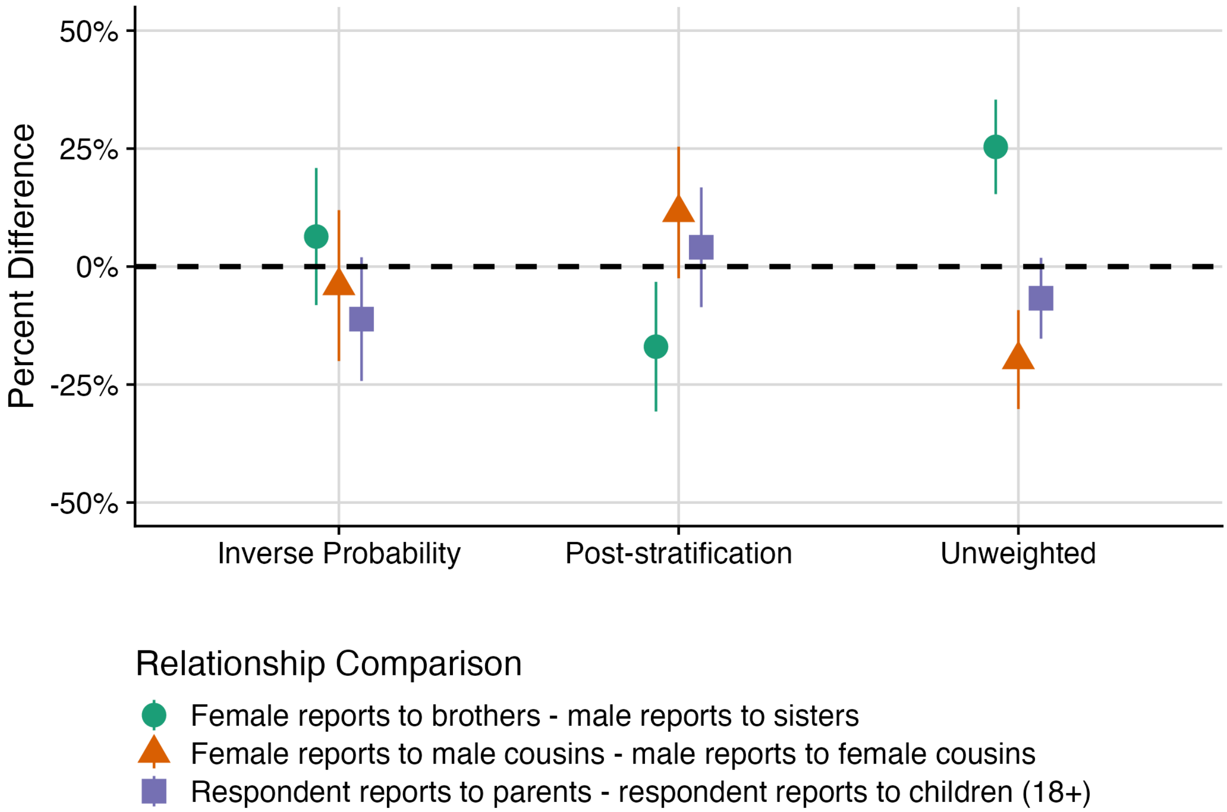

Figure S4: Internal validity checks.

S3.2 Robustness check: age compositions of networks

As a sensitivity check, we investigated the aggregate age composition of each network quota survey respondents report on. As shown in Figure S5, we benchmarked against age composition estimates obtained from the probability-based household rosters, which in this setting are the most reliable estimates available of household composition. We are restricted to the broad age categories of under age 5, 5–17, and 18 and over as we do not collect the exact age for each person the respondent reports on.

Our analysis reveals that, after applying survey weights, the age composition estimates from our quota sample closely match the benchmark data from the household rosters across both household and neighbor networks. This consistency check is reassuring, and suggests that respondents are accurately reporting the age composition of their neighbors. For the kin reports, we see that compared to household or neighbor reports, individuals aged 18 are slightly overrepresented and 5–17 year olds are slightly underrepresented.

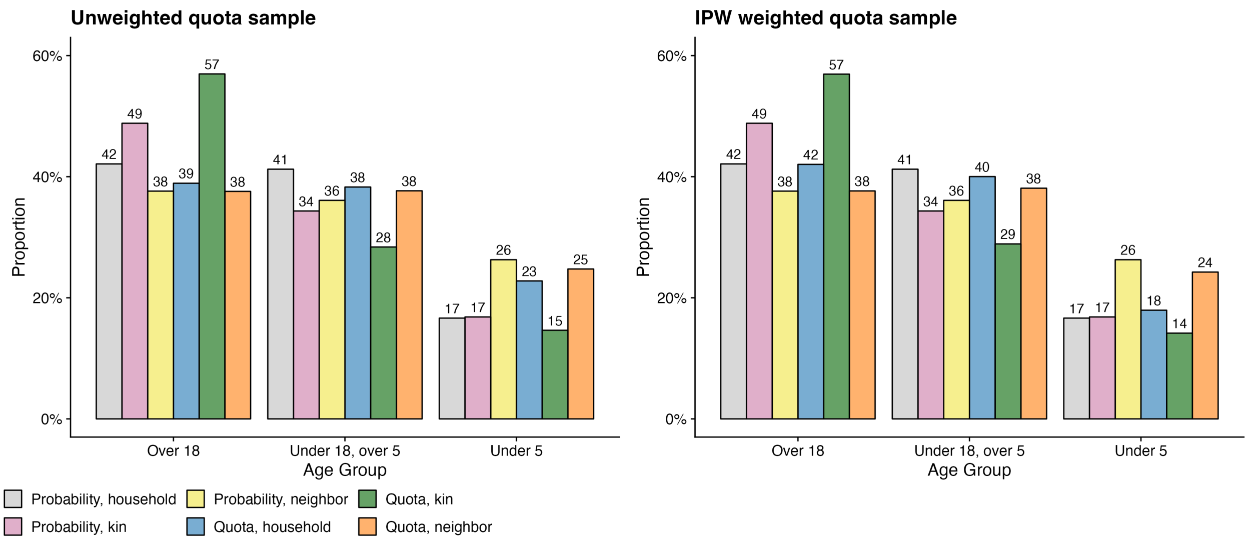

Figure S5: Age composition of different network reports in both quota and probability surveys.

S3.3 Validation check: module randomization

We randomized the order of the kin and neighbor network survival across surveys. As a validation check, we tested whether respondents reported differential network sizes or number of deaths depending on whether a module was administered first or second. Specifically, we wanted to confirm that respondents did not become fatigued taking the survey and report fewer deaths and/or smaller network sizes later in the interview. As shown in Figure S6, we find no statistically significant difference across survey modules for both the quota and the probability survey.

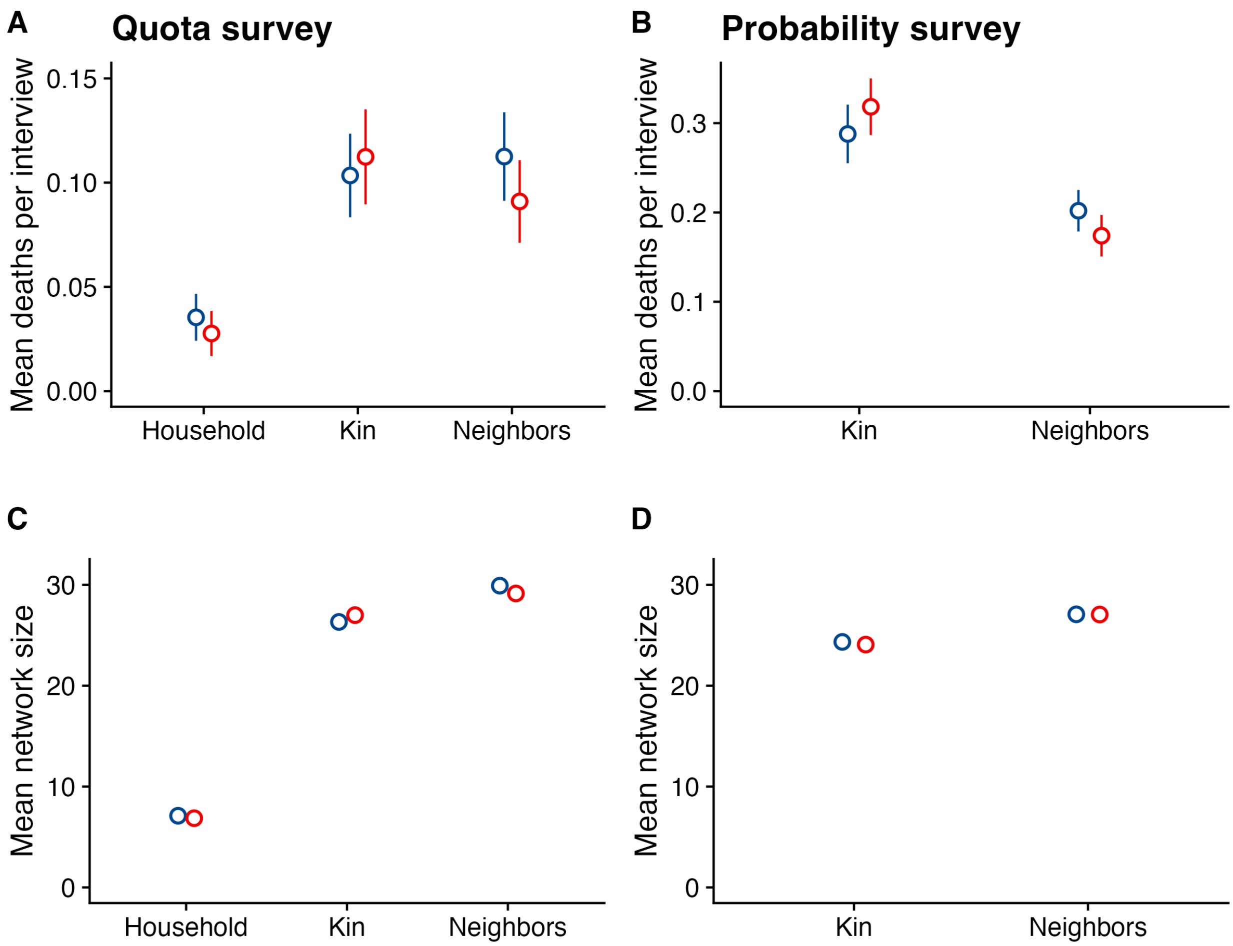

Figure S6: (A) and (B) show the mean deaths reported per interview by tie in the quota sample and probability sample depending on whether a module was randomly administered first (blue) or second (red). (C) and (D) similarly show the average network size by tie depending on randomized module order. We find no statistically significant or otherwise meaningful difference across survey modules for both the quota and the probability survey. **Notes:** In the probability survey, respondents reported on household deaths as part of a separate household module. This was not randomized, and always came before the network module. The probability survey has a larger number of mean reported deaths per interview than the quota survey as it asked about a substantially longer time window.

S4 Data processing

Additional Site in Nyunzu. To facilitate targeting the most remote areas in Nyunzu, we see set up a secondary sampling site in Nyunzu Town. Similar to Kalemie City, Nyunzu Town is an important town that attracts people from nearby villages. Some people from especially remote regions Health Areas of the Nyunzu Health Zone are quite far away from Kalemie City, making this a practical choice to help collect surveys.

This enumerator had little-to-no direct supervision from a field officer, and we still conducted the majority of our Nyunzu interviews (60%+) in Kalemie City. In our main analysis, we present estimates only based on interviews conducted in Kalemie City. As a robustness check, we regenerate weights and calculate our set of network estimates including the Nyunzu enumerator. As shown in Figure S7, there is no statistically significant difference between any estimates including and excluding the Nyunzu enumerator.

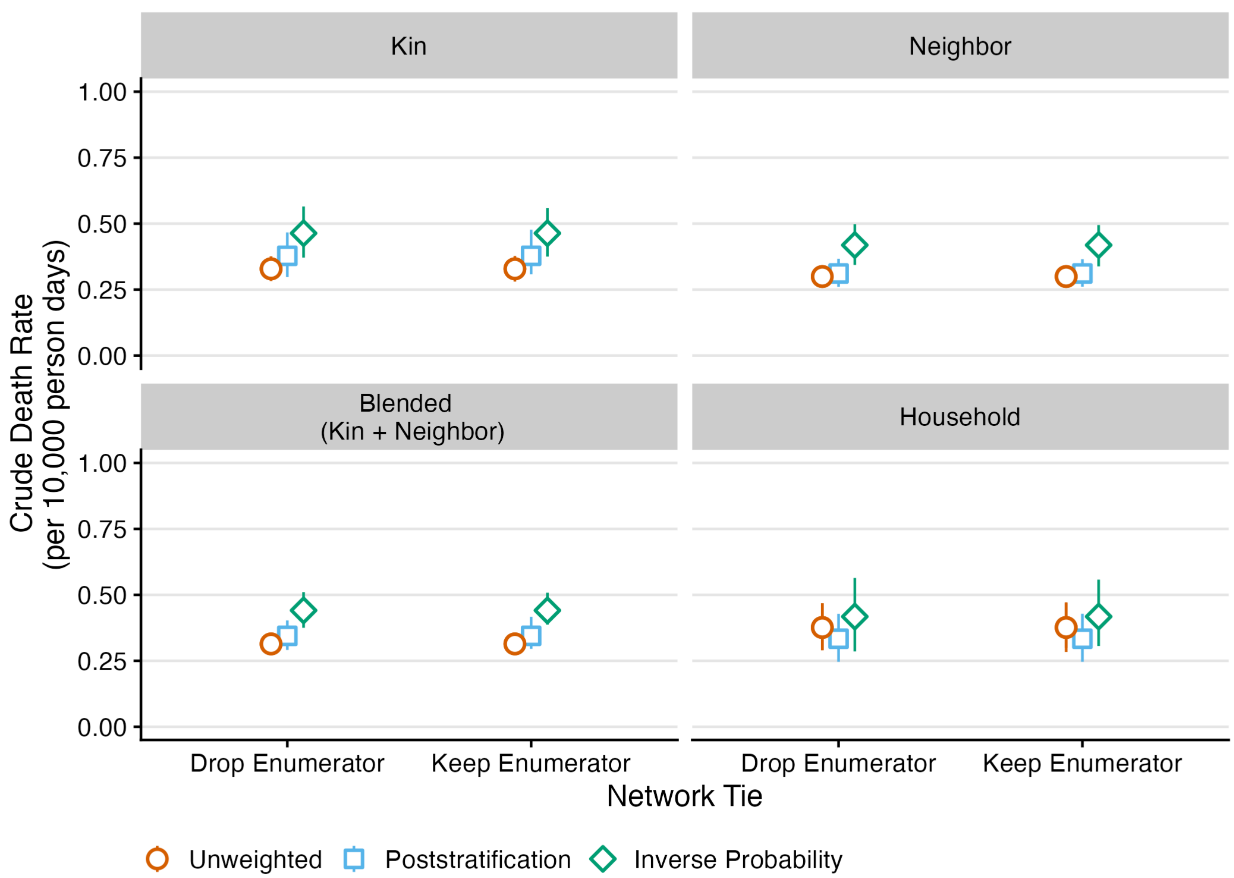

Figure S7: Difference in CDR estimates if Nyunzu Town enumerator is or is not included

S4.1 Missing data

Network method: numerator We drop any respondents with missing values on reported number of deaths. For all deaths, respondents were asked to give an exact date. If the respondents could not provide an exact day of death, they were asked to provide their best guess of the month in which the death occurred. We drop deaths that occurred before the beginning of our observation period, January 1st, 2023 (N = 5).

Network method: denominator We drop respondents who report missingness on questions about the size of their personal networks (N = 18). All respondents were asked to report on the closest five neighboring households by walking distance. In rare cases, respondents could not report accurately on the exact number of household members in all households, especially the fourth and fifth household. When respondents expressed uncertainty about the exact number of household members or gave a range of people living in the household, enumerators instructed the respondent to only report on their closest three households. These respondents were not dropped from the survey.

Weighting variables We drop records with missing values (N = 3) for sociodemographic and weighting variables, including owning a bed, type of cooking fuel, and livelihood. After dropping respondents with missing reported deaths, denominators, and socioedemographic characteristics, we were left with an analytic sample of 2,526 respondents.

S5 Additional results

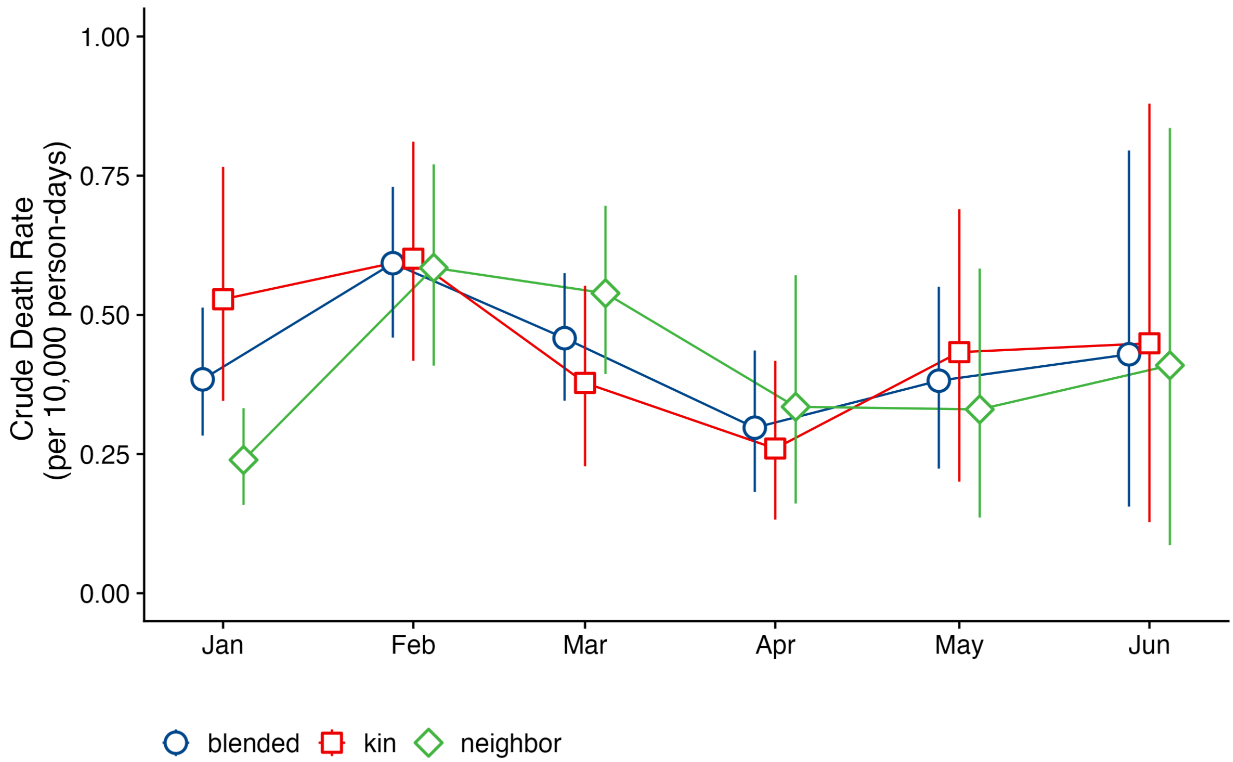

Figure S8: Quota sample estimates over time. Estimates are presented using inverse-probability weights.

Table S3: CDR estimates

Survey	Tie	Health Zone	Month	Weights	Death Rates	Lower	Upper
Probability	Kin	-	-	Unweighted	0.55	0.51	0.60
Probability	Neighbor	-	-	Unweighted	0.40	0.36	0.43
Probability	Blended	-	-	Unweighted	0.48	0.44	0.51
Probability	Household	-	-	Unweighted	0.81	0.71	0.92
Probability	Kin	Kalemie	-	Unweighted	0.73	0.64	0.84
Probability	Kin	Nyemba	-	Unweighted	0.50	0.42	0.58
Probability	Kin	Nyunzu	-	Unweighted	0.46	0.39	0.53
Probability	Neighbor	Kalemie	-	Unweighted	0.47	0.40	0.53
Probability	Neighbor	Nyemba	-	Unweighted	0.34	0.29	0.40
Probability	Neighbor	Nyunzu	-	Unweighted	0.39	0.33	0.44
Probability	Blended	Kalemie	-	Unweighted	0.60	0.53	0.67
Probability	Blended	Nyemba	-	Unweighted	0.42	0.37	0.48
Probability	Blended	Nyunzu	-	Unweighted	0.42	0.37	0.48
Probability	Household	Kalemie	-	Unweighted	0.83	0.65	1.02
Probability	Household	Nyemba	-	Unweighted	0.91	0.72	1.12
Probability	Household	Nyunzu	-	Unweighted	0.72	0.57	0.88
Quota	Kin	-	-	Unweighted	0.33	0.28	0.38
Quota	Neighbor	-	-	Unweighted	0.30	0.26	0.34
Quota	Blended	-	-	Unweighted	0.31	0.28	0.35
Quota	Household	-	-	Unweighted	0.38	0.29	0.47
Quota	Kin	-	-	Poststrat	0.38	0.30	0.47
Quota	Neighbor	-	-	Poststrat	0.31	0.26	0.37
Quota	Blended	-	-	Poststrat	0.34	0.29	0.40
Quota	Household	-	-	Poststrat	0.33	0.25	0.43
Quota	Kin	-	-	IPW	0.46	0.37	0.56
Quota	Neighbor	-	-	IPW	0.42	0.34	0.50
Quota	Blended	-	-	IPW	0.44	0.38	0.51
Quota	Household	-	-	IPW	0.42	0.29	0.56
Quota	Kin	Kalemie	-	IPW	0.46	0.33	0.59
Quota	Kin	Kalemie	-	Unweighted	0.37	0.29	0.46
Quota	Kin	Nyemba	-	IPW	0.30	0.22	0.40
Quota	Kin	Nyemba	-	Unweighted	0.30	0.24	0.38
Quota	Kin	Nyunzu	-	IPW	0.73	0.45	1.09
Quota	Kin	Nyunzu	-	Unweighted	0.30	0.22	0.40
Quota	Neighbor	Kalemie	-	IPW	0.50	0.36	0.65
Quota	Neighbor	Kalemie	-	Unweighted	0.33	0.27	0.41
Quota	Neighbor	Nyemba	-	IPW	0.27	0.20	0.35
Quota	Neighbor	Nyemba	-	Unweighted	0.29	0.23	0.35
Quota	Neighbor	Nyunzu	-	IPW	0.55	0.35	0.76
Quota	Neighbor	Nyunzu	-	Unweighted	0.27	0.20	0.33
Quota	Blended	Kalemie	-	IPW	0.48	0.38	0.58
Quota	Blended	Kalemie	-	Unweighted	0.35	0.30	0.41
Quota	Blended	Nyemba	-	IPW	0.28	0.22	0.36
Quota	Blended	Nyemba	-	Unweighted	0.30	0.25	0.35
Quota	Blended	Nyunzu	-	IPW	0.64	0.45	0.87
Quota	Blended	Nyunzu	-	Unweighted	0.29	0.23	0.35
Quota	Household	Kalemie	-	IPW	0.55	0.32	0.80
Quota	Household	Kalemie	-	Unweighted	0.50	0.34	0.68
Quota	Household	Nyemba	-	IPW	0.39	0.21	0.62
Quota	Household	Nyemba	-	Unweighted	0.42	0.27	0.58

Table S3: Death rate estimates for different health zones and months. *(continued)*

survey	social tie	health_zone	month	weights	death_rate	death_rate_lower	death_rate_upper
Quota	Household	Nyunzu	-	IPW	0.23	0.02	0.57
Quota	Household	Nyunzu	-	Unweighted	0.13	0.04	0.24
Quota	Kin	-	2023-01-01	IPW	0.53	0.35	0.77
Quota	Kin	-	2023-01-01	Unweighted	0.36	0.27	0.45
Quota	Kin	-	2023-02-01	IPW	0.60	0.42	0.81
Quota	Kin	-	2023-02-01	Unweighted	0.42	0.32	0.52
Quota	Kin	-	2023-03-01	IPW	0.38	0.23	0.55
Quota	Kin	-	2023-03-01	Unweighted	0.27	0.20	0.35
Quota	Kin	-	2023-04-01	IPW	0.26	0.13	0.42
Quota	Kin	-	2023-04-01	Unweighted	0.27	0.16	0.38
Quota	Kin	-	2023-05-01	IPW	0.43	0.20	0.69
Quota	Kin	-	2023-05-01	Unweighted	0.27	0.15	0.40
Quota	Kin	-	2023-06-01	IPW	0.45	0.13	0.88
Quota	Kin	-	2023-06-01	Unweighted	0.38	0.14	0.66
Quota	Neighbor	-	2023-01-01	IPW	0.24	0.16	0.33
Quota	Neighbor	-	2023-01-01	Unweighted	0.22	0.17	0.28
Quota	Neighbor	-	2023-02-01	IPW	0.58	0.41	0.77
Quota	Neighbor	-	2023-02-01	Unweighted	0.40	0.31	0.49
Quota	Neighbor	-	2023-03-01	IPW	0.54	0.39	0.70
Quota	Neighbor	-	2023-03-01	Unweighted	0.37	0.29	0.46
Quota	Neighbor	-	2023-04-01	IPW	0.33	0.16	0.57
Quota	Neighbor	-	2023-04-01	Unweighted	0.23	0.16	0.32
Quota	Neighbor	-	2023-05-01	IPW	0.33	0.14	0.58
Quota	Neighbor	-	2023-05-01	Unweighted	0.22	0.13	0.33
Quota	Neighbor	-	2023-06-01	IPW	0.41	0.09	0.84
Quota	Neighbor	-	2023-06-01	Unweighted	0.26	0.10	0.46
Quota	Blended	-	2023-01-01	IPW	0.38	0.28	0.51
Quota	Blended	-	2023-01-01	Unweighted	0.29	0.24	0.35
Quota	Blended	-	2023-02-01	IPW	0.59	0.46	0.73
Quota	Blended	-	2023-02-01	Unweighted	0.41	0.34	0.48
Quota	Blended	-	2023-03-01	IPW	0.46	0.35	0.57
Quota	Blended	-	2023-03-01	Unweighted	0.32	0.27	0.38
Quota	Blended	-	2023-04-01	IPW	0.30	0.18	0.44
Quota	Blended	-	2023-04-01	Unweighted	0.25	0.18	0.32
Quota	Blended	-	2023-05-01	IPW	0.38	0.22	0.55
Quota	Blended	-	2023-05-01	Unweighted	0.24	0.17	0.33
Quota	Blended	-	2023-06-01	IPW	0.43	0.16	0.80
Quota	Blended	-	2023-06-01	Unweighted	0.32	0.16	0.51

S6 Survey Instrument

The full survey instrument for the quota survey is shown below.

Network Method Survey Instrument

Section: Screening Script
I am _____, working for IMPACT Initiatives, a sister organization to ACTED, an international nonprofit organization working in this area. Together with the University of Kinshasa School of Public Health and University of California Berkeley, we are doing research on methods to improve reporting of deaths in the community to better inform the health department on the number and causes of death in this area. This information helps health actors plan and run health services for the population. We are approaching you today because you are coming from, or have information on, hard-to-reach communities in Tanganyika Province. Would you have 10-15 minutes today to answer some questions about births, deaths and other health events that have occurred in your community? If yes, I would like to make sure that you are eligible before I give you more information about our work and invite you to take part in this study. May I ask, which Zone and Aire de Santé are you coming from today? • [Visually assess the sex of the respondent] • [Check against list if coming from a target area] **Is the respondent eligible for the study?** [YES / NO] [If not eligible for interview] Thank you for your time, however we do not need information from you today. [END INTERVIEW]

Section: Informed Consent
[If they are eligible for interview] You are coming from an area where need information on the health situation of the population. Would you have 10-15 minutes to answer some questions for us about births, deaths or other health events that have occurred in your community? If yes, I would like to give you some information about our work and invite you to take part in this study. If there is any part that you don't understand you can ask me to stop and I will take time to explain, or you can ask later. [APPLY INFORMED CONSENT FORM FOR NETWORK METHOD SURVEY] ** Has the respondent consented to participate? ** [YES / NO] [If yes to consent] [Continue to section 1 below]. [If no to consent] Thank you for your time. [END INTERVIEW]

Section 1: Respondent Characteristics		
S/No	**Question**	**Choices**
Q1.1	What Zone de Sante are you coming from today?	[Select one – contextual list]
Q1.2	What Aire de Sante are you coming from today?	[Select one – contextual list]

Q1.3	What Village are you coming from today?	[Select one – contextual list]
Q1.4	Is [village_name] your place of usual residence?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q1.5	What is the sex of the respondent?	1 = Male 2 = Female
Q1.6	What is the age of the respondent (in completed years)	Integer (completed years)
Q1.7	What is the marital status of the respondent?	1 = Single 2 = Married 3 = Divorced 4 = Widowed 5 = Other, please describe: _______________
Q1.8	What is the residency status of the respondent?	1 = Resident 2 = Internally Displaced Person (IDP) 3 = IDP Returnee 4 = Refugee Returnee 5 = Refugee
Q1.9	What is the highest level of education of the respondent?	1 = Pre-primary school 2 = Primary school 3 = Lower Secondary School 4 = Secondary School 5 = Post-secondary school 6 = Trade or professional school 7 = Religious school 8 = Don't know 9 = Prefer not to answer
Q1.10	What does the respondent do to make money or earn food for the household?	[select multiple – contextual list of livelihood activities]
Q1.11	What is the reason for the person's movement through town today?	1 = Transit to another location 2 = Access market 3 = Access health facility 4 = Visiting family or friends 5 = Work related reasons 6 = Other (specify) 8 = Don't know 9 = Prefer not to answer
Q1.12	What is the main material of your home's exterior walls?	1 = No walls 2 = Cane / palm tree / trunks 3 = Earth 4 = Bamboo with mud 5 = Stone with mud 6 = Uncovered adobe / bamboo / wood with mud 7 = Reused wood 8 = Wood 9 = Cement 10= Stone with lime / cement 11= Bricks 12= Cement blocks 13= Coated adobe 14= Wood planks / shingles 15= Other, please describe: _______________

Q1.13	In your household, what type of fuel is primarily used for cooking?	1 =Electricity 2 = Biogas 3 = Kerosene 4 = Coal, ignite 5 = Charcoal 6 = Wood 7 = Straw / shrubs / grass 8 = Agricultural crops 9 = No food cooked in the house 10= Other, please describe: _______________
Q1.14	Does your household have at least one bed?	1 = Yes 2 = No 3 = Don't know 4 = No response
Q1.15	Does your household have at least one radio?	1 = Yes 2 = No 3 = Don't know 4 = No response
Q1.16	Over the last 12 months, what is your occupation, that is, what kind of work do you mainly do	1 = Not currently working 2 = Professional, technical, or managerial worker (salaried) 3 = Clerical worker 4 = Sales worker 5 = Self-employed agricultural worker 6 = Agricultural employee 7 = Household, domestic, or service worker 8 = Skilled manual worker 9 = Unskilled manual worker 10= Armed forces 11= Other, please describe: _______________ 12= Don't know 13= No response

Section 2: Network Method, Household and Neighbor ties

In the following section, we want to know about the number of people you know who are your neighbors or live in your household.

Please think about all the people with your own household. By household, we mean people in most days of the previous week:

- Lived together under the same roof or in the same compound.
- Shared food from the same cooking pot

S/No	Question	Choices
Q2.1	Number of boys < 5 years of age?	Integer (total number)
Q2.2	Number of girls < 5 years of age?	Integer (total number)
Q2.3	Number of boys 5 - 18 years of age?	Integer (total number)
Q2.4	Number of girls 5 - 18 years of age?	Integer (total number)
Q2.5	Number of men 18+ years of age?	Integer (total number)
Q2.6	Number of women 18+ years of age?	Integer (total number)

Q2.7	How many people in your household have died since {recall_event}?	Integer (total number)
Please think of the 5 households closest to your household by walking distance. Please only tell me about the people who usually live in this household. By household, we mean people in most days of the previous week: • Lived together under the same roof or in the same compound • Shared food from the same cooking pot		
Repeat following questions each of the closest 5 closest households by distance, closest household to furthest.		
Q2.8	Number of boys < 5 years of age?	Integer (total number)
Q2.9	Number of girls < 5 years of age?	Integer (total number)
Q2.10	Number of boys 5 - 18 years of age?	Integer (total number)
Q2.11	Number of girls 5 - 18 years of age?	Integer (total number)
Q2.12	Number of men 18+ years of age?	Integer (total number)
Q2.13	Number of women 18+ years of age?	Integer (total number)
Q2.14	How many people in have died in {Neighbor household Num} since January 1st, 2023?	Integer (total number)
Ask the following questions about the respondent's household and 5 closest neighbors combined		
Q2.15	In your household, and your closest 5 neighbours, how many people have **LEFT** their localite or quartier since January 1st, 2023?	
Q2.16	How many births do you know of in your household, and the households of your 5 closest neighbors since January 1st, 2023?	
Q2.17	How many children under-5 years do you know in **your household, and the households of your 5 closest neighbours**, who had **MEASLES** since January 1st, 2023?	

Section 3: Network Method, Extended Kin

We want to know about people you know who: • Reside in the same Zone De Sante as you • You are blood related to • Are still alive today		
Q3.1	How many of *YOUR OWN FEMALE CHILDREN** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+	Integer

Q3.2	How many of *YOUR OWN MALE CHILDREN** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.3	How many of *YOUR OWN FEMALE GRANDCHILDREN** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.4	How many of *YOUR OWN MALE GRANDCHILDREN ** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.5	How many of *YOUR OWN SISTERS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.6	How many of *YOUR OWN BROTHERS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.7	How many of *YOUR OWN FEMALE COUSINS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.8	How many of *YOUR OWN MALE COUSINS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.9	How many of *YOUR OWN PARENTS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer

Q3.10	How many of *YOUR OWN AUNTS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.11	How many of *YOUR OWN UNCLES** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.12	How many of *YOUR OWN MALE GRANDPARENTS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	Integer
Q3.13	How many of *YOUR OWN FEMALE GRANDPARENTS** in {zone_de_sante_name} are: • < 5 years of age • 5–18 years of age • 18+ years of age	
Ask below questions about all above kin relationships pooled		
Q3.14	Within your **EXTENDED FAMILY** which you counted, how many people do you know that have left their localite or quartier since January 1st, 2023 ?	
Q3.15	Within your **EXTENDED FAMILY** which you counted, how many people do you know that have **JOINED** their localite or quartier since January 1st, 2023?	
Q3.16	Within your **EXTENDED FAMILY**, how many births do you know since January 1st, 2023 within your extended family ?	
Q3.17	Within your **EXTENDED FAMILY**, how many older children (5+ years) or adults do you know who had **serious acute watery diarrhoea** since January 1st, 2023 ?	
Q3.18	Within your **EXTENDED FAMILY**, how many children under-5 years of age do you know who had **MEASLES** since January 1st, 2023?	

Section 4: Births

You reported:

- {num_births_neighbours} births from your household and your 5 closest neighbours
- {num_births_kin} births from your extended family

Q4.1	How many total, unique births really happened since January 1st, 2023	Integer
Repeat below questions for each birth reported		
Q4.2	What is your relationship to child #{birth_pos}?	

Q4.3	What is the family relationship?	
Q4.4	Do you know the sex of the child?	1 = Male 2 = Female;
Q4.5	Do you know the day, month, and year of child #{birth_pos} birth?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q4.6	What is the **date of birth** for the child?	Date
Q4.7	If not exact date, can you estimate the **month-year of birth** for the child?	Month-Year
Q4.8	What was the outcome of this birth?	1 = Born, and alive 2 = Born, but now dead 3 = Child not born alive 4 = Don't Know 5 = Other, please describe: _____
Section 5: Suspect Cholera		
You reported: • *{num_awd_neighbours} serious acute watery diarrhoea cases from your household and your 5 closest neighbours* • *{num_awd_kin} serious acute watery diarrhoea cases from your extended family*		
Q5.1	*How many total, unique cases of serious acute watery diarrhoea really happened since January 1st, 2023?*	Integer
Repeat below questions for each unique case reported		
Q5.2	What was the sex of the person ?	1 = Male 2 = Female
Q5.3	What was the age in years of the person?	Integer
Q5.4	Did you observe the person directly when they were sick?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q5.5	Did the person have at least 3 loose stools during a 24hour period?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q5.6	Did the person have any vomiting?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q5.7	Did the person have sunken eyes?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q5.8	Do you know the **day, month, and year** that the person last had symptoms?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q5.9	What is the **last date** that you are aware the person had symptoms?	Date
Q5.10	If not exact date, can you estimate the **month-year** that the person had symptoms?	Month-Year
Q5.11	Did the person seek health care?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer

Q5.12	If so, what place was health care sought?	1 = Govt. hospital 2 = Govt. health center 3 = Govt. health post 4 = Other govt. medical facility 5 = Private hospital 6 = Private clinic 7 = Other private facility 8 = NGO hospital 9 = NGO clinic 10= Other NGO facility 11= Other (please specify) 12= Don't know
Q5.13	What was the outcome of the person's illness?	1 = Person recovered 2 = Person still sick 3 = Person died 4 = Don't know 5 = Other (please specify)
Section 6: Suspect Measles		
You reported: • {num_measles_neighbours} measles cases from your household and your 5 closest neighbours • {num_measles_kin} measles cases from your extended family		
Q6.1	How many children (under-5 years) do you know who had measles since January 1st, 2023?	Integer
Repeat below questions for each person reported		
Q6.2	What was the sex of the child ?	1 = Male 2 = Female
Q6.3	What was the age in years of the child?	Integer
Q6.4	Did you observe the child directly when they were sick?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q6.5	Did the child have a rash on their head and/or neck?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q6.6	Did the child have fever?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q6.7	Do you know the **day, month, and year** that the child had measles?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q6.8	What is the **last date** that you are aware the child had measles symptoms?	Date
Q6.9	If not exact date, can you estimate the **month-year** that the child had measles symptoms?	Month-Year
Q6.10	Did the child seek health care?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer

Q6.11	If so, what place was health care sought?	1 = Govt. hospital 2 = Govt. health center 3 = Govt. health post 4 = Other govt. medical facility 5 = Private hospital 6 = Private clinic 7 = Other private facility 8 = NGO hospital 9 = NGO clinic 10= Other NGO facility 11= Other (please specify) 12= Don't know
Q6.12	What was the outcome of the child's illness?	1 = Person recovered 2 = Person still sick 3 = Person died 4 = Don't know 5 = Other (please specify)

Section 7: Deaths

You reported:

- {num_deaths_hh} deaths from your own household
- {num_deaths_neighbours} deaths from your 5 closest neighbours
- {num_deaths_kin} deaths from your extended family

Q7.1	How many total, unique deaths really happened since {recall_event}?	Integer
Repeat below questions for each person reported		
Q7.2	What was the first name of the deceased individual?	Text
Q7.3	What was the family name of the deceased individual?	Text
Q7.4	Was [name_deceased] known by any other names or nicknames?	Text
Q7.5	What was the sex of [name_deceased]?	1 = Male 2 = Female;

Q7.6	What was the age in completed years of [name_deceased] ?	Integer
Q7.7	Do you know the day, month, and year of [name_deceased] birth?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.8	What is the **date of birth** for [name_deceased]?	Date
Q7.9	If not exact date, can you estimate the **month-year of birth** for [name_deceased] ?	Month-Year
Q7.10	Do you know the day, month, and year that [name_deceased] passed away?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.11	Do you know the exact date that [name_deceased] passed away?	Date
Q7.12	If not, please estimate the month-year of death as close as possible?	Month-Year
Q7.13	What was the main cause of death for [name_deceased]?	1 = Acute disease 2 = Chronic disease 3 = Intentional violence 4 = Accident/trauma 5 = Post-partum (0-42 days) 6 = During pregnancy 7 = During delivery 8 = Other (please specify)

		9 = Don't know
Q7.14	Where did the [name_deceased] pass away?	1 = Current location of residence 2 = Health facility at current location of residence 3 = During migration or displacement 4 = At last place of residence 5 = Health facility at last place of residence
Q7.15	Did [name_deceased] seek health care in the 2 weeks before dying?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.16	If so, what place was health care sought?	1 = Govt. hospital 2 = Govt. health center 3 = Govt. health post 4 = Other govt. medical facility 5 = Private hospital 6 = Private clinic 7 = Other private facility 8 = NGO hospital 9 = NGO clinic 10= Other NGO facility 11= Other (please specify) 12= Don't know
Q7.17	If not, what was the main reason for not seeking care in a health structure/facility?	1 = Immediate death 2 = No money/consultation too expensive 3 = Too sick to seek care 4 = Not sick enough to seek care 5 = Health facility too far away 6 = Went to a traditional healer 7 = No time to go/too busy to go 8 = No trust in the health facility 9 = Safety issue 10= Care was refused at the health center 11= Other please specify 12= Don't know
Q7.18	In your own words, can you provide any other details about the circumstances of [name_deceased]'s death?	[Text description]
Q7.19	Was [name_deceased] a part of your own household?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.20	Was [name_deceased] a membre of the community you currently live in?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.21	If no, what Zone de Sante did [name_deceased] live at the time of death?	[Select one – contextual list]
Q7.22	If no, what Aire de Sante did [name_deceased] live at the time of death?	[Select one – contextual list]
Q7.23	If no, what Village did [name_deceased] live at the time of death?	[Select one – contextual list]

Q7.24	We would like to follow up more closely with the household of [name_deceased] to better understand the causes of their death. This will help us understand the causes of high mortality in Tanganyika Province so the health department and NGOs can better plan their response. We would like to ask your permission to follow up with [name_deceased]'s household directly to better understand the causes of death. We would not disclose your information that you told us about the death, but it would increase the risk of breaching your confidentiality if we discussed with the household. If you are not comfortable with us following up with the household, please tell us. We will only follow up with them if you give us permission to do so. Do we have your permission to follow up with the household of [name_deceased]?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.25	Do you have any phone number you can share for [name_deceased]'s household?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.26	Do we have your permission to follow up with [name_deceased] household with some questions about cause of death?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.27	Phone number	Phone Number
Q7.28	Is there anyone else we could call by phone who could connect us with [name_deceased]'s household?	1 = Yes 2 = No 8 = Don't know 9 = Prefer not to answer
Q7.29	Phone number (alternate):	[Phone Number]
Q7.30	Do you have any other information on how we could reach or contact [name_deceased]'s household?	[Text Description]

References

[1] Dennis M. Feehan, Mary Mahy, and Matthew J. Salganik. The Network Survival Method for Estimating Adult Mortality: Evidence From a Survey Experiment in Rwanda. *Demography*, 54(4):1503–1528, August 2017. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-017-0594-y. URL https://read.dukeupress.edu/demography/article/54/4/1503/167730/The-Network-Survival-Method-for-Estimating-Adult.

[2] Francesco Checchi and Les Roberts. Interpreting and using mortality data in humanitarian emergencies: A primer for non-epidemiologists. Technical report, 2005. URL https://odihpn.org/publication/interpreting-and-using-mortality-data-in-humanitarian-emergencies/.

[3] Christopher McCarty, Peter D. Killworth, H. Russell Bernard, Eugene C. Johnsen, and Gene A. Shelley. Comparing Two Methods for Estimating Network Size. *Human Organization*, 60(1):28–39, 2001. ISSN 0018-7259. URL https://www.jstor.org/stable/44126693.

[4] H Russell Bernard, Tim Hallett, Alexandrina Iovita, Eugene C Johnsen, Rob Lyerla, Christopher McCarty, Mary Mahy, Matthew J Salganik, Tetiana Saliuk, Otilia Scutelniciuc, Gene A Shelley, Petchsri Sirinirund, Sharon Weir, and Donna F Stroup. Counting hard-to-count populations: The network scale-up method for public health. *Sexually Transmitted Infections*, 86(Suppl_2):ii11–ii15, December 2010. ISSN 1368-4973. doi: 10.1136/sti.2010.044446. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3010902/.

[5] Holly E. Reed and Charles B. Keely. Understanding Mortality Patterns in Complex Humanitarian Emergencies. In *Forced Migration & Mortality*. National Academies Press (US), 2001. URL https://www.ncbi.nlm.nih.gov/books/NBK223340/.

[6] Ian M. Timæus. Measurement of Adult Mortality in Less Developed Countries: A Comparative Review. *Population Index*, 57(4):552–568, 1991. ISSN 0032-4701. doi: 10.2307/3644262. URL https://www.jstor.org/stable/3644262.

[7] Bruno Lankoandé, Bruno Masquelier, Pascal Zabre, Hélène Bangré, Géraldine Duthé, Abdramane B. Soura, Gilles Pison, and Sié Ali. Estimating mortality from census data: A record-linkage study of the Nouna Health and Demographic Surveillance System in Burkina Faso. *Demographic Research*, 46:653–680, April 2022. ISSN 1435-9871. doi: 10.4054/DemRes.2022.46.22. URL https://www.demographic-research.org/articles/volume/46/22.

[8] Kenneth Hill, Peter Johnson, Kavita Singh, Anthony Amuzu-Pharin, and Yagya Kharki. Using census data to measure maternal mortality: A review of recent experience. *Demographic research*, 39:337–364, 2018. ISSN 1435-9871. doi: 10.4054/DemRes.2018.39.11. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6903798/.

[9] Prudence Jarrett, Frank J. Zadravecz, Jennifer O'Keefe, Marius Nshombo, Augustin Karume, and Les Roberts. Evaluation of a population mobility, mortality, and birth

surveillance system in South Kivu, Democratic Republic of the Congo. *Disasters*, 44 (2):390–407, April 2020. ISSN 1467-7717. doi: 10.1111/disa.12370.

[10] Michael R. Elliott and Richard Valliant. Inference for Nonprobability Samples. *Statistical Science*, 32(2):249–264, 2017. ISSN 0883-4237. URL https://www.jstor.org/stable/26408228.

[11] Maksym Bondarenko, David Kerr, Alessandro Sorichetta, Andrew Tatem, and WorldPop,. Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints., 2020. URL https://www.worldpop.org/doi/10.5258/SOTON/WP00682.

[12] Dennis M. Feehan, Aline Umubyeyi, Mary Mahy, Wolfgang Hladik, and Matthew J. Salganik. Quantity Versus Quality: A Survey Experiment to Improve the Network Scale-up Method. *American Journal of Epidemiology*, 183(8):747–757, April 2016. ISSN 0002-9262. doi: 10.1093/aje/kwv287. URL https://doi.org/10.1093/aje/kwv287.

[13] Smart Survey. Enquete Nutritionnelle Smart Territoire De Kalemie. Technical report, studfee, 2022.