Estimating death rates in complex humanitarian emergencies using the network method

New York University - Abu Dhabi

Casey F. Breen [1]

[1]University of Oxford | Nuffield College

April 24, 2025

Joint with: Dennis Feehan (UC Berkeley), Saeed Rahman (Reach Initiative), Christina Kay (Reach Initiative), Joeri Smits (Reach Initiative), Steve Ahuka (University of Kinsasha)

Motivation for study

- Death rates are critical for assessing the severity of a crisis and effectively allocating resources

Motivation for study

- Death rates are critical for assessing the severity of a crisis and effectively allocating resources
- How do you best estimate death rates in **humanitarian emergencies**?

Motivation for study

- Death rates are critical for assessing the severity of a crisis and effectively allocating resources
- How do you best estimate death rates in **humanitarian emergencies**?
- Conventional systems and methods fall apart
  - Civil war, earthquake, etc.

We still need new methods....

Emerg Themes Epidemiol. 2007; 4: 9. PMCID: PMC1904216
Published online 2007 Jun 1. doi: 10.1186/1742-7622-4-9 PMID: 17543103

Wanted: studies on mortality estimation methods for humanitarian emergencies, suggestions for future research

Working Group for Mortality Estimation in Emergencies ✉[1]

▾ Author information ▸ Article notes ▸ Copyright and License information PMC Disclaimer

[1]Epicentre, 8 rue Saint-Sabin, 75011 Paris, France
✉Corresponding author.
Working Group for Mortality Estimation in Emergencies : Rebecca.grais@epicentre.msf.org

New Approach

1. Collect a non-probability sample feasible in a humanitarian emergency
2. Use **network survival method** to estimate crude death rates

Overview of Network Survival Method

$$M_\alpha = \frac{D_\alpha}{N_\alpha} \quad (1)$$

where

- M_α is the death rate (for subgroup α)

Overview of Network Survival Method

$$M_\alpha = \frac{D_\alpha}{N_\alpha} \tag{1}$$

where

- M_α is the death rate (for subgroup α)
- D_α is the number of deaths

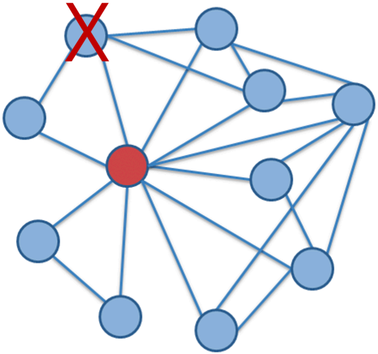
Overview of Network Survival Method

$$M_\alpha = \frac{D_\alpha}{N_\alpha} \tag{1}$$

where

- M_α is the death rate (for subgroup α)
- D_α is the number of deaths
- N_α is the person days of exposure

Insights from social networks



People can report valuable information about mortality among their social network

Network survival method

$$\widehat{M}_\alpha = \frac{D_\alpha}{N_\alpha} \tag{2}$$

$$= \frac{\sum_{i \in s} w_i \; y_{i,D}}{\sum_{i \in s} w_i \; d_i \; E_i} \tag{3}$$

where

- $\sum_{i \in s} w_i \; y_{i,D}$ is the (weighted) total number of people in respondents' personal network who have died in time window

Network survival method

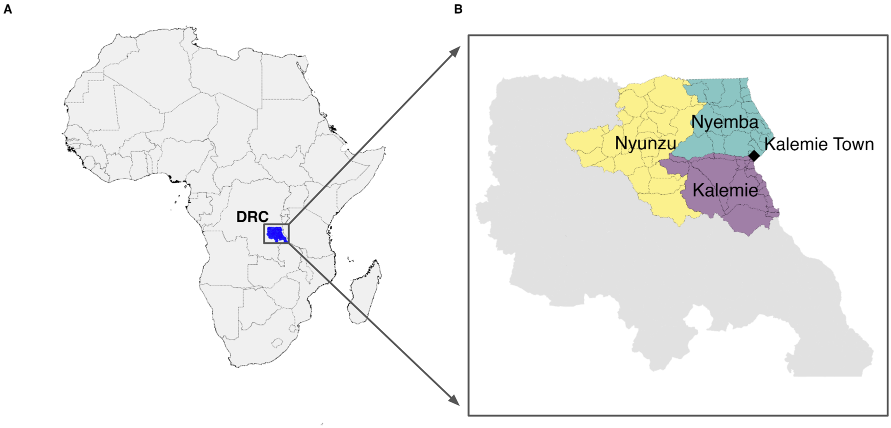
$$\widehat{M}_\alpha = \frac{D_\alpha}{N_\alpha} \tag{2}$$

$$= \frac{\sum_{i \in s} w_i \ y_{i,D}}{\sum_{i \in s} w_i \ d_i \ E_i} \tag{3}$$

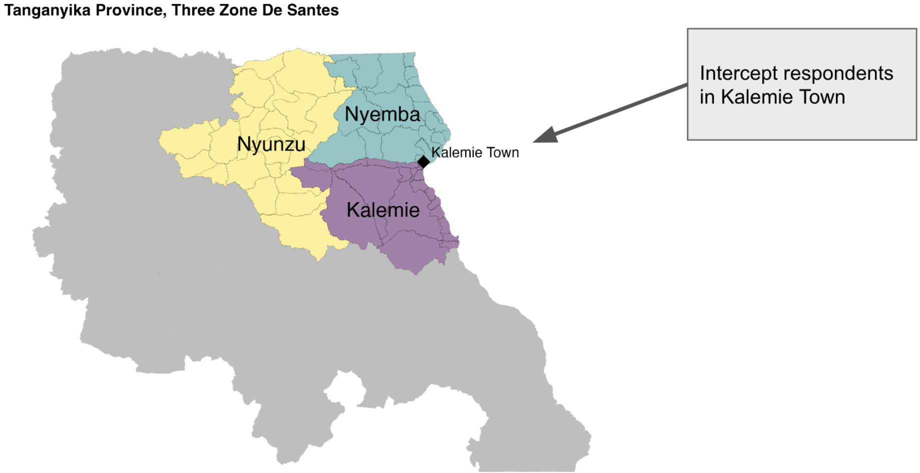
where

- $\sum_{i \in s} w_i \ y_{i,D}$ is the (weighted) total number of people in respondents' personal network who have died in time window
- $\sum_{i \in s} w_i \ d_i \ E_i$ is the (weighted) total amount of exposure reported on in respondents' personal network

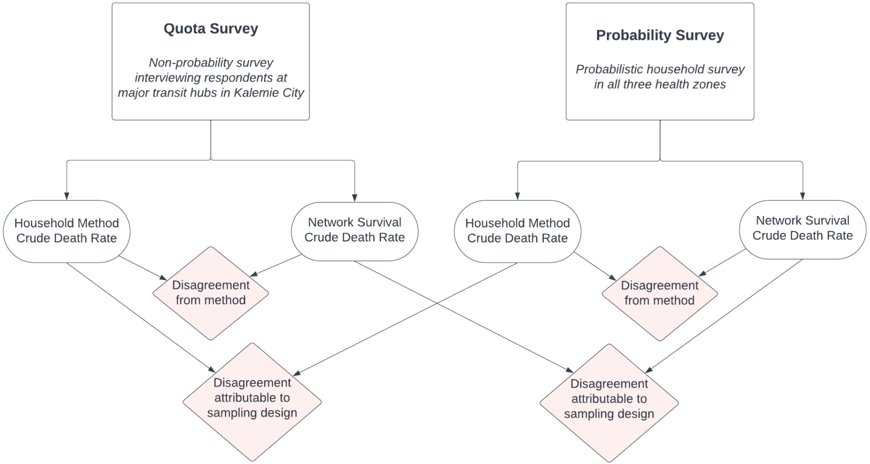
Case Study: Democratic Republic of the Congo



Case Study: Democratic Republic of the Congo



Study design - two data collection efforts

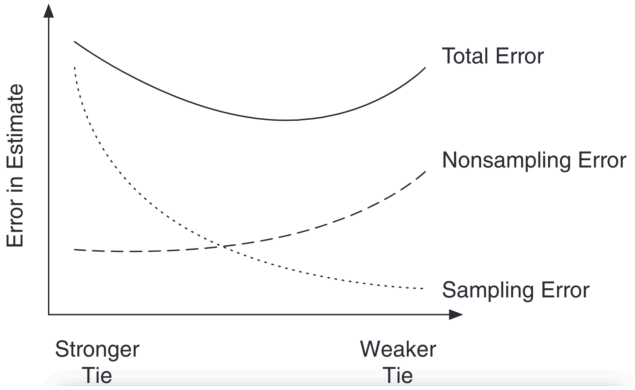


Formative research (qualitative)

- **Goal**:
  - What deaths can respondents report on accurately?
  - What reference date and reporting window should we chose?
  - Identify major transit hubs + selection dynamics
- 8 focus groups + 20 individual interviews

Group	Age	Gender	Location
1	<45	Male	Urban
2	<45	Female	Urban
3	<45	Male	Rural
4	<45	Female	Rural
5	45+	Male	Urban
6	45+	Female	Urban
7	45+	Male	Rural
8	45+	Female	Rural

Key design question – how do we pick a network tie?



Feehan et al. 2016. American Journal of Epidemiology.

Best option — kin and neighbor networks

Module	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

Data Collection, Quota (Partnership with REACH Initiative)

- Interview respondents at major transit hubs in Kalemie Town
  - Ports, markets, taxi stands, foot paths, and health clinics

Data Collection, Quota (Partnership with REACH Initiative)

- Interview respondents at major transit hubs in Kalemie Town
  - Ports, markets, taxi stands, foot paths, and health clinics
- Jan 1st, 2023 – salient reference date
- Continuous data collection for 4 months (March 1st 2023 - July 1st 2023)

Data Collection, Quota (Partnership with REACH Initiative)

- Interview respondents at major transit hubs in Kalemie Town
  - Ports, markets, taxi stands, foot paths, and health clinics
- Jan 1st, 2023 – salient reference date
- Continuous data collection for 4 months (March 1st 2023 - July 1st 2023)
- Quota sample by geographic region + gender (non-probability)

Data Collection

Interviews conducted at transit hubs such as markets, ports, taxi stations, health clinics, etc.

Informed consent on paper form, survey administered on smartphone

Quota sample: Interviews by month ($N = 2{,}650$)

Month	Kalemie	Nyemba	Nyunzu
March	203	198	200
April	201	203	202
May	216	221	232
June	228	237	204

Note: All respondents report on deaths since January 1st, 2023

Probability-based household survey

- Fielded between August 2nd, 2023 and September 2nd, 2023

Probability-based household survey

- Fielded between August 2nd, 2023 and September 2nd, 2023
- 2,970 households in three health zones (Nyunzu, Nyemba, and Kalemie)

Probability-based household survey

- Fielded between August 2nd, 2023 and September 2nd, 2023
- 2,970 households in three health zones (Nyunzu, Nyemba, and Kalemie)
  - Sampling frame was constructed from population data from the Ministry of Health

Probability-based household survey

- Fielded between August 2nd, 2023 and September 2nd, 2023
- 2,970 households in three health zones (Nyunzu, Nyemba, and Kalemie)
  - Sampling frame was constructed from population data from the Ministry of Health
  - Primary sampling units at the village level

Probability-based household survey

- Fielded between August 2nd, 2023 and September 2nd, 2023
- 2,970 households in three health zones (Nyunzu, Nyemba, and Kalemie)
  - Sampling frame was constructed from population data from the Ministry of Health
  - Primary sampling units at the village level
- The household survey asked detailed information about deaths occurring within their household after January 1st, 2023 (same as Quota sample!)

Quota sample re-weighting strategies - 3 different scenarios

	Description	**Auxiliary data**	**Setting**
No weights	No weights. Relies on quota sampling on gender and geography.	**Source**: Ministry of Health population data **Covariates**: Gender, geography (for quotas)	Imitates a setting where limited data are available to establish quotas for sampling.
Poststratification weights	Split sample into cells defined by unique combination of covariates. Weight each respondent within a cell by the inverse of their inclusion probability.	**Source**: Worldpop 100m X 100m unconstrained gridded population estimates **Covariates**: Age, gender, geography	Imitates a setting where no high-quality reference data are available but Worldpop population estimates are available.
Inverse-probability weights	Fit logistic regression model to estimate inclusion probability. Weights generated as inverse of inclusion probability.	**Source**: Our probability survey **Covariates**: Age, gender, household size, household age composition, radio, bed, wall material, fuel type	Imitates a setting where high-quality reference data are available.

More auxiliary data available

Inverse probability weights

Fit a model to estimate inclusion probability:

$$w_i = \frac{1}{\hat{P}(S_i = 1)} \tag{4}$$

where w_i is a weight define as the the inverse probability of being included in the sample $(S_i = 1)$.

Inverse probability weights

Fit a model to estimate inclusion probability:

$$w_i = \frac{1}{\hat{P}(S_i = 1)} \tag{4}$$

where w_i is a weight define as the the inverse probability of being included in the sample $(S_i = 1)$.

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

Blended estimates - combined neighbor and kin estimates

$$\underbrace{\widehat{M}}_{\text{Blended Estimate}} = \underbrace{\theta \widehat{M}^A}_{\text{Weighted Estimator A}} + \underbrace{(1-\theta)\widehat{M}^B}_{\text{Weighted Estimator B}} \tag{6}$$

where θ is a weight $\in [0, 1]$.

Blended estimates - combined neighbor and kin estimates

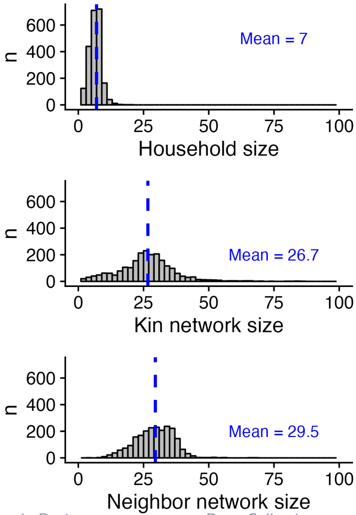
$$\underbrace{\widehat{M}}_{\text{Blended Estimate}} = \underbrace{\theta \widehat{M}^A}_{\text{Weighted Estimator A}} + \underbrace{(1-\theta)\widehat{M}^B}_{\text{Weighted Estimator B}} \tag{6}$$

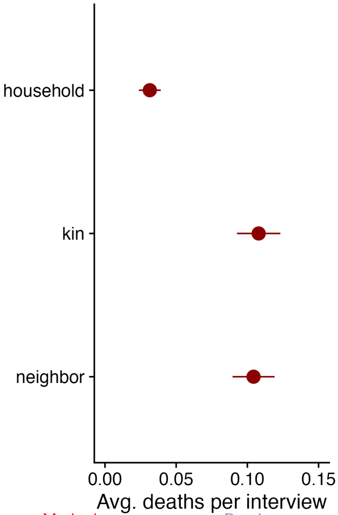
where θ is a weight $\in [0, 1]$.

We can pick the optimal $\theta^\star$ to minimize mean squared error:

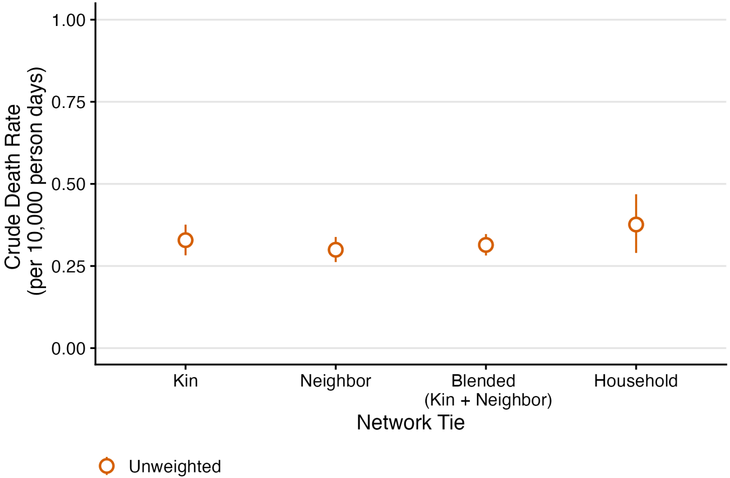
$$\theta^\star = \frac{\sigma_B^2 - \sigma_{AB}}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}}, \tag{7}$$

Network sizes distribution

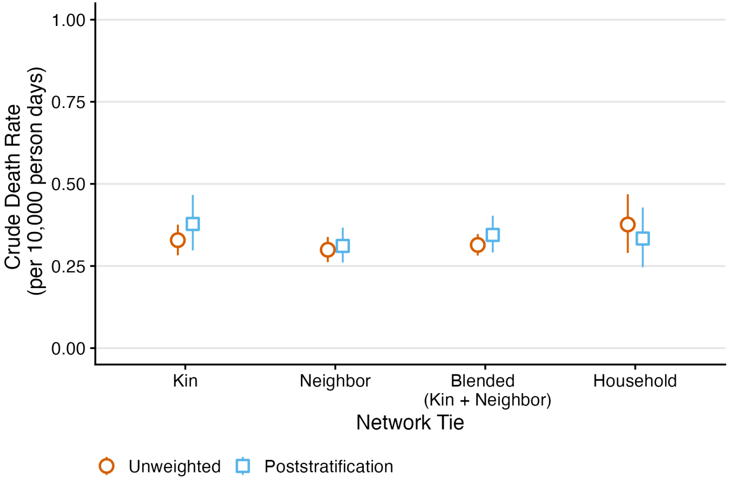




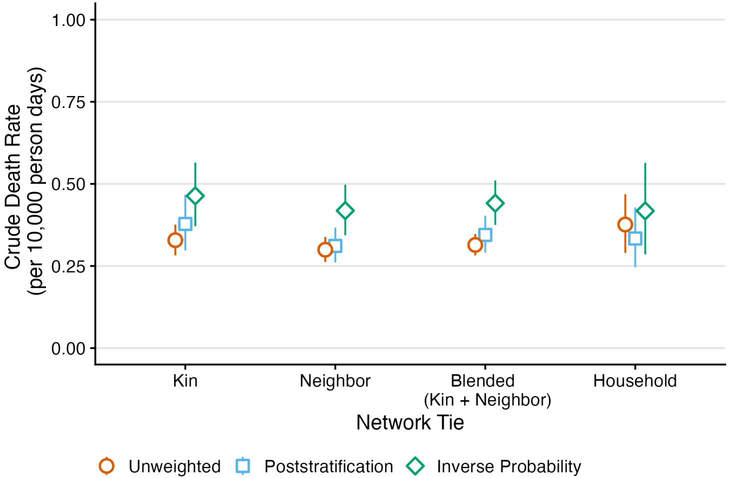
Non-probability network survival results



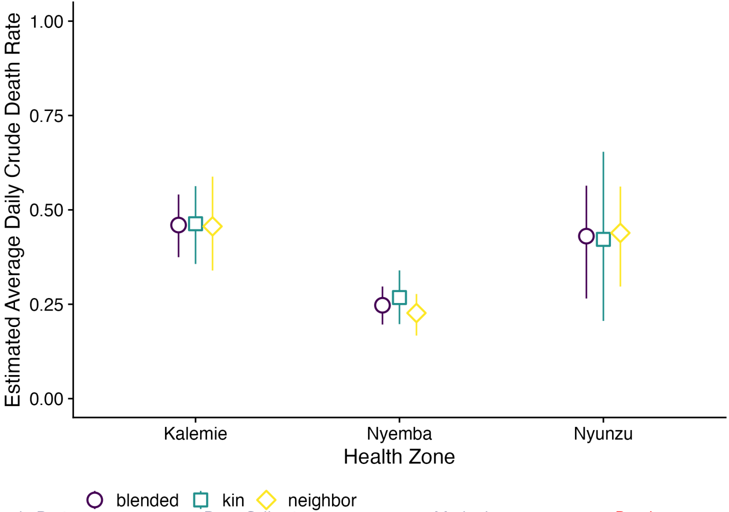
Non-probability network survival results



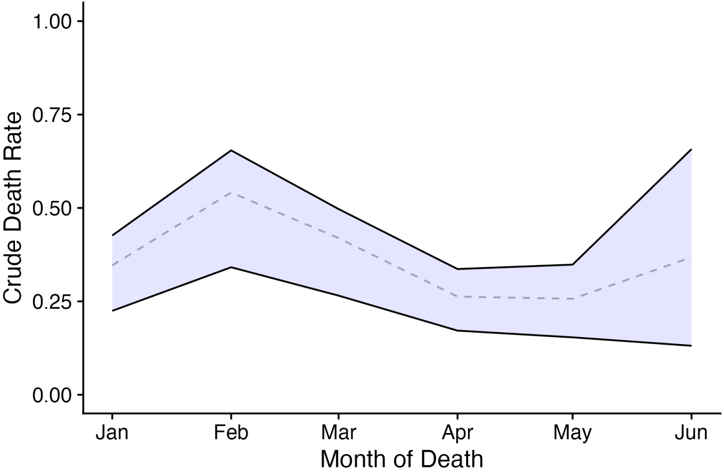
Non-probability network survival results



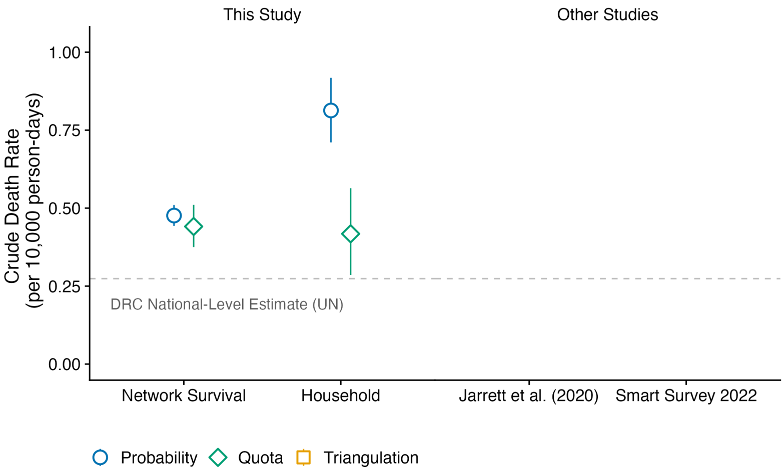
Variation across health zones



Monitoring trends over time



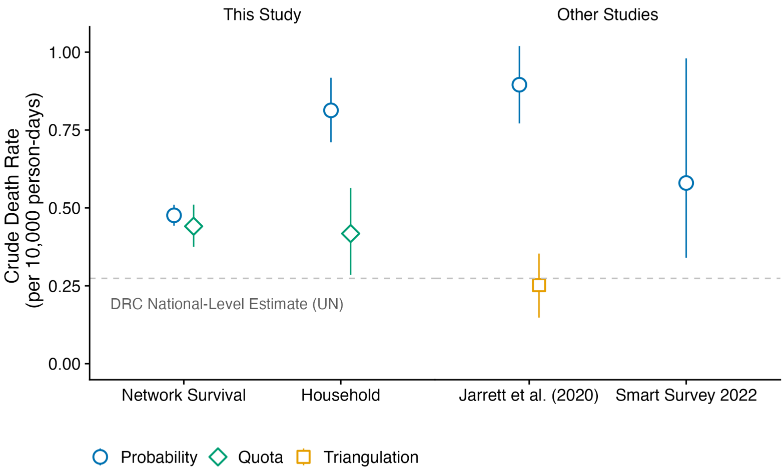
Full comparisons (blended estimates)



Comparator study - Jarrett et al. 2020

- Study conducted in Fizi Province
- Surviellance + household survey
- Verification and reconciliation of each reported death

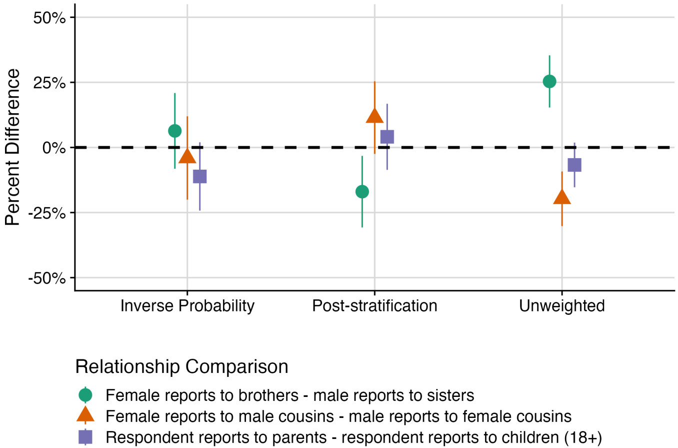
Full comparisons + external studies



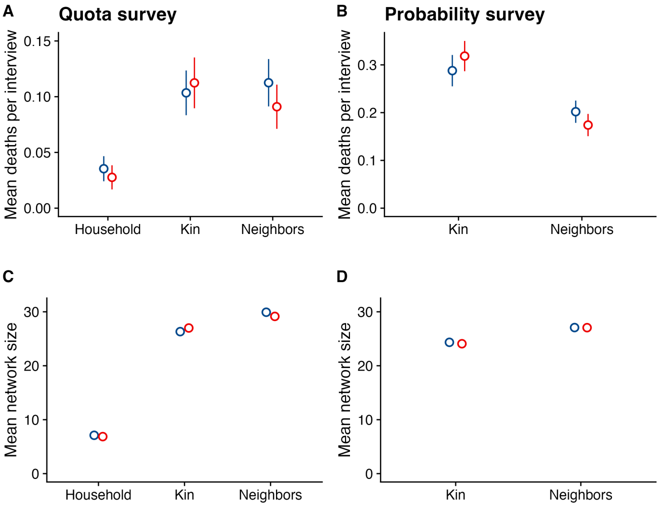
Lots of typical validation checks...

1. Missing under 5 deaths
2. Enumerator effects (no evidence)
3. No recall bias for earlier months
4. Consistent neighbor reports as household get further away...

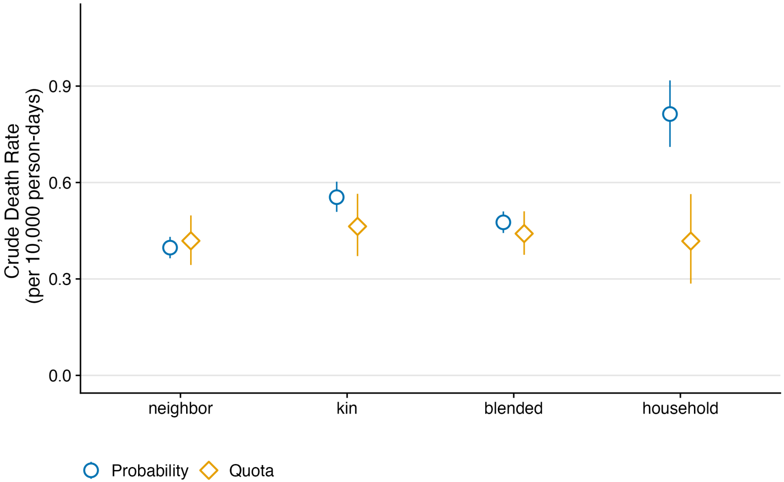
Internal validity checks



Randomized modules - no "speeding"



One estimates stands out...



Potential sources of error

Strategic Overreporting

- Financial incentive to overreport in hopes of larger aid distribution
- If network neighbor estimate is true, respondents would need to overreport by 48%
- High, but lower than 72% found by Jarret et al. (2020)

Transmission Error

- Violations of perfect visibility
- If household estimate is true, respondents would need to miss reporting 51% of deaths of neighbors
- Unlikely based on qualitative research

Summary of study

- Developed and tested a promising **new method** for estimating death rates in humanitarian emergencies

Summary of study

- Developed and tested a promising **new method** for estimating death rates in humanitarian emergencies
  - We need more systematic evaluations

- Highly contextual – **requires** localized knowledge of social networks, diffusion of info about deaths, etc.

Where to next...?

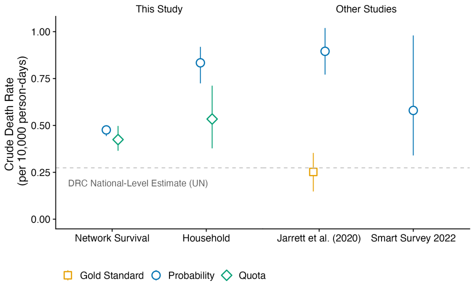
- We need more **rigorous assessment** of both standard household mortality surveys and network survival method in real humanitarian crisis settings

Where to next...?

- We need more **rigorous assessment** of both standard household mortality surveys and network survival method in real humanitarian crisis settings
  - Verification of deaths, network reports, etc.
- Adding network questions onto standard household surveys ("off-the-shelf" module)
- Other **extensions**...?

Thank You

- Questions?



caseyfbreen
casey.breen@demography.ox.ac.uk

Age composition of network members