A Method for Small-Area Estimation of Population Mortality in Settings Affected by Crises

Francesco Checchi (Francesco.Checchi@lshtm.ac.uk)  
London School of Hygiene & Tropical Medicine  
https://orcid.org/0000-0001-9030-5382

Adrienne Testa  
London School of Hygiene & Tropical Medicine

Amy Gimma  
London School of Hygiene & Tropical Medicine

Emilie Koum-Besson  
London School of Hygiene & Tropical Medicine

Abdihamid Warsame  
London School of Hygiene & Tropical Medicine

Research Article

Keywords: Mortality, death rate, crisis, humanitarian, displaced, emergency, war, predictive model, small area estimation, secondary data, method

DOI: https://doi.org/10.21203/rs.3.rs-558991/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
A method for small-area estimation of population mortality in settings affected by crises

Francesco Checchi PhD*
Adrienne Testa MSc
Amy Gimma MSc
Emilie Koum-Besson MSc
Abdihamid Warsame MSc

Department of Infectious Disease Epidemiology
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine

* Corresponding author: Francesco.checchi@lshtm.ac.uk
Abstract

Background

Populations affected by crises (armed conflict, food insecurity, natural disaster) are poorly covered by demographic surveillance. As such, crisis-wide estimation of population mortality is extremely challenging, resulting in a lack of evidence to inform humanitarian response and conflict resolution.

Methods

We describe here a ‘small-area estimation’ method to circumvent these data gaps and quantify both total and excess (i.e. crisis-attributable) death rates and tolls, both overall and for granular geographic (e.g. district) and time (e.g. month) strata. The method is based on analysis of data previously collected by national and humanitarian actors, including ground survey observations of mortality, displacement-adjusted population denominators and datasets of variables that may predict the death rate. We describe the six sequential steps required for the method’s implementation and illustrate its recent application in Somalia, South Sudan and northeast Nigeria, based on a generic set of analysis scripts.

Results

Descriptive analysis of ground survey data reveals informative patterns, e.g. concerning the contribution of injuries to overall mortality, or household net migration. Despite some data sparsity, for each crisis we have used the method in thus far, available predictor data allow the specification of reasonably predictive mixed effects models of crude and under 5 years death rate, validated using cross-validation. Assumptions about values of the predictors in the absence of a crisis provide counterfactual and excess mortality estimates.

Conclusions

The method enables retrospective estimation of crisis-attributable mortality with considerable geographic and period stratification, and can therefore contribute to better understanding and historical memorialisation of the public health effects of crises. We discuss key limitations and areas for further development.

Key words

Mortality, death rate, crisis, humanitarian, displaced, emergency, war, predictive model, small area estimation, secondary data, method
Background

Mortality estimation in crisis-affected populations

In populations exposed to conditions of crisis (armed conflict, food insecurity, natural disasters, etc.), estimation of population death rates and tolls is of paramount importance. The extent to which these conditions result in excess population mortality is an objective criterion for benchmarking the severity of the crisis, at least in public health terms, and therefore a basis on which to predicate an appropriate humanitarian response [1, 2]. Evidence on crisis-attributable mortality is also potentially important for advocacy and historical documentation [3, 4]. Over the past two decades, estimates of mortality have informed war crime prosecution in the former Yugoslavia [5], illuminated the toll of armed conflict in Darfur [6, 7], the Democratic Republic of Congo [8] and Iraq [9, 10], documented the impact of famine in Somalia [11] and, most recently, demonstrated the direct and indirect health impacts of the SARS-CoV-2 pandemic [12–14].

Crisis-attributable mortality is difficult to estimate, even in high-income countries [15, 16]. If the crisis occurs over a large, resource-constrained region, and/or is due to armed conflict and/or political violence, additional challenges [4, 17] arise, including (i) lack of robust, representative routine data on population mortality for some or most of the population and/or period of interest; (ii) unfeasibility of representative primary data collection (e.g. a crisis-wide retrospective survey) to fill information gaps, due to insecurity, lack of authorisations, funding constraints or other factors; and (iii) inability to collect robust retrospective data due to having to elicit recall of demographic events (in the case of a survey) or compose death records (for surveillance or other methods) over a very long period in the past (e.g. > 2y). While there is to our knowledge no evidence on how reliable next-of-kin information becomes as the ‘recall’ period of data collection stretches back to many years prior, such response bias is believed to be a major constraint to survey validity; survival and selection biases caused by households disintegrating due to high mortality or migration are a further concern [17]. Even where data become available, establishing a counterfactual (i.e. non-crisis) death rate presents a further challenge, particularly in very protracted crises (e.g. Afghanistan or the eastern Democratic Republic of Congo) where such a baseline has been unobservable for decades.

Scope of this paper

Here, we describe the design and implementation of a method for circumventing the above challenges, and estimating crisis-attributable death rates and death tolls based on previously collected data. Past applications of the method for estimation of famine and food insecurity-attributable mortality in Somalia (2010-2012) [11], and deaths due to armed conflict in South Sudan (2013-2018) [18], have been published elsewhere, albeit reflecting less advanced iterations. Detailed results of implementation in Somalia (2014-2018), Nigeria and the Democratic Republic of the Congo will be the subject of separate papers. South Sudan, Somalia and Nigeria examples are however used here to illustrate the application and constraints
of the method. We first outline the general method, then detail the implementation of key steps, and lastly discuss ways forward.

Methods

Why a small-area estimation approach?

The methods we set out below are broadly classifiable as a small-area estimation approach, first developed in the United States to efficiently estimate characteristics of interest, e.g. smoking prevalence or poverty levels, for small geographical units (e.g. counties) without having to conduct primary data collection within each such unit [19]. Our method consists entirely of secondary data analysis, and is designed to deliver estimates for small geographical and time strata while exploiting existing data from different sectors of humanitarian response, insofar as these relate to population health and thus mortality.

General framework

Suppose a population of size $N$ is exposed to a crisis during a given period $T$. The rate and number of excess deaths $D$ that may be attributed to it can be defined conceptually as the difference between the number of deaths that has actually occurred during the crisis and the number that would have occurred in the absence of the crisis, or

$$D_E = D_A - D_C$$

where $E$ denotes the excess (i.e. crisis-attributable), $A$ what actually happened and $C$ the counterfactual, i.e. what would have happened, had there been no crisis.

Figure 1 illustrates a hypothetical example, in which, notably, excess mortality persists beyond the end of the crisis period: this could happen (i) if the crisis negates secular improvements in mortality that might have been expected to accrue in its absence, effectively returning the population to a ‘higher’ baseline than pre-crisis, or (ii) if certain effects manifest with delay (e.g. increased tuberculosis mortality due to higher transmission of $M. \text{tuberculosis}$ when people lived in displacement camps years earlier).

[FIGURE 1 HERE]
Suppose we wish to estimate $D_E$ over $N$ and $T$. Suppose moreover that the population is distributed into $K$ geographical units. These could be administrative level 2 entities such as counties or districts; they could also however be geographical units whose boundaries may correlate more closely with mortality risk, e.g. settlements for internally displaced persons (IDPs) or areas characterised by a dominant economic activity, e.g. pastoralism or sedentary agriculture (such areas are commonly known as livelihood zones; see below). The extent of crisis-attributable mortality may vary considerably over time and across these geographic units: information on where and when mortality is highest may be useful for humanitarian responses (e.g. to identify gaps) or advocacy and documentation purposes (e.g. to better understand the dynamics of an armed conflict). We may therefore also wish to estimate $D_E$ for specific sub-populations or periods of interest: more generally, we may wish to estimate $D_{E,k,t}$, i.e. for each possible $k, t$ stratum of analysis (e.g. district-month).

The total death toll $D$ is the product of $y$, the mean death rate per capita and per time unit $t$, the population $N$ experiencing this death rate and the period of interest $T$:

$$\text{Eq. 2} \quad D = yNT$$

where the quantity $NT$ approximates the epidemiological concept of ‘person-time at risk’. Rewriting this relationship for our problem, and for any stratum $kt$, we have

$$\text{Eq. 3} \quad D_{E,k,t} = D_{A,k,t} - D_{C,k,t} = y_{A,k,t}N_{A,k,t} - y_{C,k,t}N_{C,k,t}$$

Equation 3 shows that four quantities need to be estimated: the actual and counterfactual death rates, and the corresponding population sizes (in a counterfactual no-crisis scenario, population size might have evolved differently from actuality, for example because forced displacement associated with the crisis would likely not have happened). Once all four quantities are estimated, we can aggregate results for any strata into a partial death toll (e.g. over a region of interest or a specific sub-period in the crisis timeline), or compute overall excess mortality as

$$\text{Eq. 4} \quad D_E = \sum_{k=1}^{K} \sum_{t=1}^{T} D_{E,k,t}$$

We may also wish to estimate mortality within a specific age group $a$ (for example, children aged under 5y). This requires knowledge of the death rate $y_a$ and population size $N_a$ falling within the age group of interest. If we know the proportion $p_a$ of the population who is within the age group, then $N_a = Np_a$. Similarly, estimating the proportion of deaths due to intentional injuries, i.e. the direct effects of violence
(\nu), requires an estimate of the cause-specific death rate \( y_{\nu} \), or alternatively the proportion \( p_{\nu} \) of all deaths that are due to intentional injuries, since \( y_{\nu} = y p_{\nu} \). We will not explore age- and cause-specific mortality estimation further here, as they are straightforward extensions of our general approach, provided the above parameters are known.

**Estimation steps**

Our adaptation of small-area estimation consists of using available data to develop a statistical model (specific to each crisis) that predicts the death rate as a function of several ‘predictor’ variables; and applying this model to project mortality under actual (observed) and assumed counterfactual conditions, including for locations and periods for which no ground mortality data are available. Excess mortality is then the difference between actual and counterfactual estimates.

Table 1 summarises the steps involved in the full application of the method. The details of data management steps required to prepare and link the different datasets are omitted here, and are instead annotated on R statistical scripts (see Declarations and Additional File). We also omit step 2, namely reconstructing population denominators for each stratum of analysis, taking into account variability in available demographic estimates and displacement occurring as a result of the crisis: for brevity, we will detail this potentially cumbersome step in a separate paper.

### Table 1. Summary of estimation steps.

| Step | Description | Sub-steps | Data requirements | Depends on |
|------|-------------|-----------|------------------|------------|
| 1    | Identify existing ground mortality data and prepare them for analysis | Identify all available estimates. Extract meta-data for each estimate. Clean and re-analyse datasets. Grade estimate quality. Describe data coverage and crude patterns in key demographic indicators. | Raw datasets of surveys or other estimation exercises Survey reports Official administrative data, shape files for geographic boundaries | Data collection and management steps |
| 2    | Reconstruct population denominators [not presented in this paper] | Identify and curate alternative population datasets. Grade their robustness. Identify and curate displacement data. Make appropriate assumptions on population and displacement dynamics. Reconstruct population for each \( k, t \) stratum as an average of alternative estimates. | Population datasets Remote sensing estimates Internal and refugee displacement data Explanatory accompanying documents and reports | Steps 1-2 |
| 3    | Capture predictor variable data and prepare them for analysis | Identify possible sources of data based on a conceptual framework. Capture and curate predictor datasets. Ascertain missingness and perform any appropriate imputation. | Predictor datasets Explanation of variable meanings / |
| Step | Description | Sub-steps | Data requirements | Depends on |
|------|-------------|-----------|------------------|------------|
| 4    | Fit a statistical model to predict the death rate as a function of the predictors | ▪ Explore correlation among predictors.  
▪ Do univariate analysis.  
▪ Fit alternative multivariate models and select the most appropriate one. | Extensive contextual knowledge.  
▪ Mortality and predictor data for periods as long as possible before the crisis (optional). | Steps 1-3 |
| 5    | Apply the model to estimate excess mortality while propagating known sources of error | ▪ Specify a set of counterfactual scenarios:  
▪ Agree on what key deviations from normal define the crisis being analysed.  
▪ Arbitrarily define alternative (e.g. most likely, best-case, worst-case) scenarios for what values the model predictors would have taken in the absence of a crisis.  
▪ Construct counterfactual predictor datasets accordingly.  
▪ Apply counterfactual death rates and assumptions on displacement to reconstruct corresponding counterfactual population denominators.  
▪ Set up statistical simulation that implements Eq. 3 for each k, t stratum while drawing from known error distributions of each parameter.  
▪ Compute excess death toll estimates overall and for sub-populations/periods of interest (Eq. 4). | Extensive contextual knowledge.  
▪ Mortality and predictor data for periods as long as possible before the crisis (optional). | Steps 2-4 |
| 6    | Conduct sensitivity analyses of interest | ▪ Explore how possible bias or uncertainty in key parameters affect the estimates, by running the analysis with alternative data or assumptions. | Step 5 |

**Defining the analysis person-time and strata**

Specifying the population and period for which estimates are sought, and the granularity with which these may be computed, determines most of the subsequent steps, but itself can only be undertaken once the analysis question is clarified, and the availability of mortality, predictor and demographic data is established. The boundaries of N and T may be defined to reflect what is known about the geographical distribution of the population affected by the crisis, and the timespan of the crisis itself. In some scenarios, this will be straightforward (e.g. an entire country or a specific region is affected by armed conflict with a clear start and end date). In other cases, the analysis may be conducted to estimate mortality up to a certain time point in the crisis.

The definition of ‘crisis’ also needs to be made explicit: for example, Somalia has experienced 30 years of armed conflict; against this backdrop, drought and flooding emergencies have repeatedly occurred. Our analyses to date in Somalia have aimed to estimate mortality attributable to exceptional food insecurity events (2010-2012, 2017-2018) [20] triggered by drought, i.e. above and beyond any excess deaths caused by the protracted conflict alone. Accordingly, we have defined the period of analysis as that over which key food security indicators and other markers of crisis conditions were reported to be unusually poor. In Nigeria, we wished to estimate mortality attributable to the armed conflict between the government and Boko Haram, which affects three states (Borno, Yobe, Adamawa) in the northeast: this is a more straightforward scenario in which a relatively recent baseline of no conflict precedes the crisis. Refugees
IDPs who leave the crisis-affected region may also be considered within the study population. However, this bears several complexities: for example, refugees will be exposed to different risk factors and may paradoxically experience lower mortality than if they had remained in their country of origin, implying a negative excess mortality: this has been documented for South Sudanese refugees in Uganda [21], and could plausibly apply to the large Syrian refugee population now living in Europe.

In practice, the population and time period of analysis may be constrained by data availability: the method requires data on predictors and population to be consistently available across person-time, at the chosen level of stratification. Furthermore, it is advisable to (i) start the timespan of data collection at least 6-12 months prior to the start of \( T \), so as to allow exploration of lagged effects of predictors on mortality; and (ii) if the actual geographical and timespan of the crisis is uncertain, extend \( N \) and \( T \) outward to include a ‘buffer’, which can later be omitted when presenting estimates, and may be useful as ‘baseline’ person-time to set counterfactual values for the predictors (see below).

Generally, it is useful to divide the analysis person-time into the smallest strata \( k,t \) for which consistent data are available: predictive models are likely to be more accurate as input data increase in size and variability, and estimates may have greater utility if end-users can observe patterns over small time and geographical intervals, and map these to known developments in the crisis. In practice, the geographical structure of available data will determine the level of stratification. As detailed below, sample surveys conducted for routine assessment and monitoring by various humanitarian actors are the commonest source of mortality ground data. In a Somalia study (2010-2012) [11] we conducted in the aftermath of a severe famine using a small-area approach for the first time, nearly all mortality surveys had as their sampling universe the intersection of regional and livelihood zone boundaries: for example, within Gedo region some surveys were designed to represent communities that predominantly relied on pastoralism, while other surveys covered IDPs and riverine agriculturalists. Most of the predictors and demographic estimates were also collected at or could be aggregated to this stratification level. Our chosen \( k \) was thus regional livelihood zones (Table 2).

### Table 2. Analysis strata, Somalia (2010-2012) [11].

| Region            | Number of strata, by livelihood type | Total strata |
|-------------------|-------------------------------------|--------------|
|                   | Pastoralist | Agro-pastoralist | Riverine | Urban | IDP |               |
| Bakool            | 1          | 1                | 0        | 0     | 1   | 3             |
| Banadir (Mogadishu)| 0          | 0                | 0        | 1     | 1   | 2             |
| Bay               | 1          | 1                | 0        | 1     | 1   | 4             |
| Galgaduud        | 1          | 1                | 0        | 0     | 1   | 3             |
| Gedo             | 1          | 1                | 1        | 0     | 1   | 4             |
| Hiraan           | 1          | 1                | 1        | 0     | 1   | 4             |
| Lower Juba       | 1          | 1                | 1        | 1     | 1   | 5             |
| Middle Juba      | 1          | 1                | 1        | 0     | 1   | 4             |
| Mudug            | 1          | 1                | 0        | 0     | 1   | 3             |
| Lower Shabelle   | 1          | 1                | 1        | 1     | 1   | 5             |
| Middle Shabelle  | 1          | 1                | 1        | 1     | 1   | 5             |
| Totals           | 10         | 10               | 6        | 5     | 11  | 42            |
In more recent work (South Sudan, 2013-2018; Somalia, 2014-2018), available data were mainly representative of level 2 administrative units (counties and districts respectively). In Nigeria (2015-2018), where the main administrative levels consist of states (level 1) and Local Government Areas (LGAs, level 2), most available surveys use two-stage cluster sampling and have as their sampling universe contiguous aggregations of LGAs known as ‘domains’ (i.e. the three states of interest are divided into 12 domains and 65 LGAs). By contrast, predictors and population data are consistently stratified to LGA level. As survey raw datasets and reports were also available to us, we were able to identify the LGA from which each survey cluster was drawn, and thus undertake analysis at the LGA-month level, ignoring domains altogether.

Results

Data collection and management steps

Mortality data

Ground mortality observations are required to train and validate a predictive model. Mortality data may arise from any robust estimation method, as long as the number of deaths and the population and period within which they have occurred is known, and data are representative of a known population/area (in most crisis settings, mortality figures collected only within health facilities would not constitute a robust source, as many people die outside of healthcare settings). Different types of sources (e.g. surveys and vital events surveillance) can probably be combined, though this would likely require solutions to deal with the different statistical error processes behind each source.

The Standardised Monitoring of Relief and Transitions (SMART) initiative [22] has developed a globally applicable protocol for rapid, relatively small-site surveys that primarily aim to estimate the prevalence of acute malnutrition, but often also include a questionnaire module that elicits information from sampled households on their demographic experience over a retrospective ‘recall’ period, typically 3-6 months long [23]. SMART surveys are highly standardised and conducted routinely in most humanitarian responses [24], albeit not usually with wide geographical coverage. Their design and analysis are automated by Emergency Nutrition Assessment (ENA) software, reducing the potential for surveyor error [25]. Hundreds such surveys are conducted worldwide every year by various humanitarian actors, both to investigate specific locations and to generate seasonal estimates across the crisis-affected region. They typically rely on two-stage cluster sampling, though some, e.g. in IDP camps, are exhaustive or use systematic random selection. Sample sizes of 300-1000 households and 20-30 clusters are typical, i.e. sampled households are only a small fraction of the total. We identified 205 analysis-eligible SMART surveys in Somalia (2010-2012), 210 in South Sudan (2013-2018), 91 in Somalia (2014-2018) and 70 in Nigeria (2016-2018). Despite
these substantial numbers, geographic and period data coverage can be sparse, as illustrated in Figure 2 for South Sudan.

Either aggregate or individual mortality questionnaires are employed by SMART surveys: this in turn determines which demographic indicators can be estimated, and how household person-time should be computed (see Additional File). The individual questionnaire is recommended and increasingly utilised [23]. After cleaning datasets to resolve errors (e.g. values out of the allowed range), the crude death rate (CDR), under 5 years death rate (U5DR or CDR among the population aged under 5y), crude birth rate, in-, out- and net migration rate, and, for individual questionnaire surveys only, cause- and gender-specific death rates may be computed. The CDR and U5DR in particular are widely used by humanitarian actors to benchmark the severity of a crisis in health terms [1]. We use the survey package in R software [26] to calculate point estimates and 95%CIs for each indicator, assuming a Poisson distribution of deaths and adjusting standard errors for cluster sampling design. Inspection of crude patterns in survey indicators may be informative: for example, in South Sudan the highest CDRs were recorded in Unity State, while many surveys also indicated very high injury-attributable death rates and high relative risks of dying among males, compared to females (Figure 3).

Given challenges with humanitarian survey quality [27, 28], it is worthwhile to appraise the robustness of available data. We adopt the overall quality score \( w_{A,s} \) for anthropometric data, produced by ENA software, to compute a mortality quality score \( w_{Q,s} (\in [0,1]) \) for each survey (see Additional File), for which no method has been proposed by SMART. As the same team of surveyors collects anthropometric and mortality data, it is plausible that the quality of the two datasets would be correlated. An alternative is to systematically review survey reports for information suggesting quality problems: an algorithm for this has been developed [27], and an Excel template for applying the algorithm is available from the authors. Unfortunately, SMART survey reports do not systematically report quality issues; any available documentation should nonetheless be scrutinised to identify potential biases, particularly any restriction of the effective sampling frame to only a fraction of the intended sampling universe, due for example to insecurity or inaccessibility.

Once each survey’s quality has been reviewed, a decision can be taken to exclude surveys falling below a certain quality threshold, or, more informatively, each survey observation can be attributed a quality weight. We apply a composite weight \( w_s = w_{Q,s}w_{B,s} \), where \( w_{B,s} \), a representativeness weight (\( \in [0,1] \)), is the approximate fraction of the sampling universe that was actually included in the sample, as per the survey’s report (for example, if a report states that the sampling frame excluded 3 out of 5 districts, we set \( w_{B,s} = 0.4 \)); where an unspecified number of sampling units are excluded from the sampling frame, we assume \( w_{B,s} = 0.5 \).
Predictor data

If the statistical objective of analysis is merely to predict the death rate, any set of predictor variables that does so accurately, whatever their causal relationship with mortality, may be appropriate. However, choosing predictors that are causally related to mortality, or proxies for mortality risk determinants, is likely to enhance predictive power and help assess the model’s internal validity. To this end, we have defined a causal framework of factors leading to mortality (Additional File), which we believe is broadly applicable to different crises. Accordingly, at least some of the selected predictors should be related to plausible drivers of excess mortality risk: for example, in a drought-triggered food security crisis these might include rainfall, food purchasing power, burden of malnutrition and the incidence of epidemics (cholera, measles); in an armed conflict, the intensity of violence and disruptions to public health services might be more relevant. Identifying such ‘crisis-sensitive’ predictors is critical, as the method defines no-crisis scenarios by specifying counterfactual values for these very predictors.

In armed conflict settings and humanitarian responses, data collection is often unsystematic and disrupted by a variety of challenges [29]. In our experience to date, data are available for only few causal factors, and negotiation with agencies and humanitarian coordination mechanisms holding non-public datasets occupies a large share of analyst time. Such datasets generally have poor integrity; they are typically entered onto spreadsheet software without standardisation of geographical nomenclature, value cell or formula protections, variable dictionaries or automatic error checking – thus necessitating extensive curation. Missingness is a common problem, as illustrated for Somalia in the Additional File. We arbitrarily retain potential predictor datasets by applying an arbitrary “70-70-70” rule, namely ≥ 70% complete for ≥ 70% of $k$ and ≥ 70% of $t$. Remaining missingness is resolved through imputation, either statistical or manual (e.g. based on contextual knowledge, Somali districts with missing market staple price values are attributed a weighted mean of values from other districts, nearby districts receiving a higher weight). In order to reduce the influence of outliers (some of which may be data entry errors), where appropriate we apply moderate smoothing or running means to time series. Details of predictors considered are presented in crisis-specific papers; Table 3 shows predictors included in the final models for each of the crises studied thus far.

Table 3. Predictors included in the final models of CDR, by crisis.
Analysis steps

Predictive model fitting

If the raw datasets of mortality surveys are mostly unavailable, only stratum-level regression is feasible: this is discussed in the Additional File, and is sub-optimal. If raw data for most mortality surveys are available, household-level regression may be undertaken. Each survey \( s \) contributes data for individual households \( i \), grouped within survey sample clusters \( j \), each cluster falling within a given stratum \( k \).

Household mortality data are collected over the survey’s retrospective recall period \( T_{r,s} = \{t_{r,s,min}, t_{r,s,max}\} \), at least 3 months long. Survey data of interest include the number of deaths per household \( d_{i,j,k} \) and the person-time \( n_{i,j,k} \sum_{s=1}^{N_{i,j,k}} t_{r,s,n_{i,j,k}} \) contributed by household members \( \sum_{s=1}^{N_{i,j,k}} t_{r,s,n_{i,j,k}} \) during the recall period. Given values of predictors \( \mathbf{x}_{i,j,k} \) and the person-time \( \mathbf{n}_{i,j,k} \), we compute their weighted mean over \( T_{r,s} \), with weights equal to the proportion of each month that the recall period covers. The data structure is partly longitudinal: in Nigeria, five consecutive survey rounds took place during 2016-2019. While each survey round drew an independent sample, most LGAs hosted survey clusters during each round. In Somalia, some surveys were only representative of IDP settlements or urban areas within districts: we assume simplistically that district-wide predictor values also apply to these populations.

We use a generalised linear model with weights \( w_s \) (see above) and a quasi-Poisson distributional assumption to account for overdispersion in the count outcome. The predictive model thus takes the following form:
\[
\text{Eq. 5} \quad \log d_{i,j,k,T_{rs}} = (x_{1,k,T_{rs}} \beta_1 + x_{2,k,T_{rs}} \beta_2 + x_{3,k,T_{rs}} \beta_3 + \ldots) + (u_j + u_k) + \log \Pi_{i,j,k,T_{rs}} + \epsilon_{i,j,k}
\]

Where \( d_{i,j,k,T_{rs}} \) is the number of household deaths during the recall period; \( \beta_1, \beta_2, \beta_3 \) etc. are the fixed-effect linear coefficients of predictors \( x_1, x_2, x_3 \) etc.; \( u_j \) and \( u_k \) are, respectively, random effects for cluster \( j \) and stratum \( k \), capturing a plausible hierarchy of data as well as the repeated nature of observations; \( \log \Pi_{i,j,k,T_{rs}} \) is an offset to account for varying household person-time at risk; and \( \epsilon_{i,j,k} \) is the residual error not explained by the model. We validate candidate models for out-of-sample prediction through k-fold cross-validation (CV; partition of data into folds must be at the \( k, T_{rs} \) level if predictors have no variability below this). We use the mean Dawid-Sebastiani score (DSS) \([30]\) as a proper scoring rule appropriate for count outcomes to evaluate model fit on the training data and on CV. After exploratory analysis, where possible we select between maintaining the continuous version of the predictor or categorising into bins, as well as alternative lags, based on the lowest DSS_{CV}, and screen out predictors that are not significantly better-fitting than the null model based on an F-test p-value threshold. We fit each possible combination of remaining predictors (\( X \) predictors = \( 2^X \) possible combinations) and shortlist candidate models whose DSS is within a given bottom quantile. We select the final set of predictors based on DSS_{CV}, plausibility considerations and whether they are crisis-sensitive (see above). We test for plausible interactions and, lastly, add random effects, retaining the mixed model if its DSS_{CV} improves on the fixed-effects alternative. Thresholds for short-listing at each stage may vary depending on the number of predictors available. In practice, a mixed model may be of limited utility if most prediction happens for person-time with new levels of the random effect (e.g. in geographic strata not covered by any survey used to train the model on).

As an example, we provide in Table 4 model coefficients and performance metrics for South Sudan; predictive accuracy on cross-validation is shown in Figure 4. Aside from its moderately good performance, the observed associations support model validity: mortality increases with insecurity and where measles epidemics are present, but decreases if people are living in Protection of Civilians camps (in South Sudan, these places afforded relative safety and more intense humanitarian services) and as purchasing power improves.

Table 4. Final model to predict crude death rate, South Sudan (2013-2018). Note that the predictors and values below differ from the original model presented in the study report, as they arise from an improved fitting procedure. Random effects are omitted.

| Fixed effect  | Relative rate | 95% CI          | p-value |
|---------------|---------------|-----------------|---------|
| Intercept     | 0.00014       | 0.00008 to 0.00022 | < 0.001 |
| Region        |               |                 |         |
| Northeast     | [ref.]        |                 |         |
| Northwest     | 0.54          | 0.41 to 0.72    | < 0.001 |
| Southern      | 0.80          | 0.51 to 1.25    | 0.326   |
| Main livelihood type |           |                 |         |
| Agriculturalists | [ref.]       |                 |         |
| agro-pastoralists | 0.82       | 0.55 to 1.22    | 0.329   |
| Fixed effect | Relative rate | 95% CI     | p-value |
|--------------|---------------|------------|---------|
| Pastoralists displaced to Protection of Civilians camps | 1.24 | 0.69 to 2.23 | 0.478 |
| Rate of insecurity events (per 100,000 people per month, lag = 4 months) | 0.52 | 0.34 to 0.81 | 0.004 |
| 0.01 to 0.99 | 1.16 | 1.02 to 1.32 | 0.021 |
| ≥ 1.00 | 1.32 | 1.08 to 1.62 | 0.008 |
| Uptake of measles vaccine (doses administered per 100,000 people per month) | 0.52 | 0.34 to 0.81 | 0.004 |
| 0.01 to 0.99 | 1.16 | 1.02 to 1.32 | 0.021 |
| ≥ 1.00 | 1.32 | 1.08 to 1.62 | 0.008 |
| Rate of violent incidents affecting humanitarian staff (per 100,000 per month, lag = 4 months) | 0.992 | 0.987 to 0.996 | < 0.001 |
| Incidence rate of confirmed or probable measles cases (per 100,000 per month) | 0.992 | 0.987 to 0.996 | < 0.001 |

Model performance metric | Value | Notes |
|--------------------------|-------|-------|
| Akaike Information Criterion | 20104.7 | (observed - predicted)^2/variances + 2 × log(variances) |
| Dawid-Sebastiani score (internal prediction) | 26.9 | based on 10-fold cross-validation (CV) |
| Dawid-Sebastiani score (out-of-sample prediction) | 29.2 | predicted − observed/observed |
| Relative bias (on CV) | -0.064 | observed − predicted |
| Relative 95% precision (mean across strata on CV) | 1.011 | 0.5 × (upper 95%CI − lower 95%CI) |
| Coverage of 80% confidence intervals (on CV) | 0.754 | proportion of stratum observations falling within the confidence interval of the prediction |
| Coverage of 95% confidence intervals (on CV) | 0.901 |

Excess mortality estimation

In our framework, excess mortality estimation requires projecting the death toll in counterfactual scenarios of what would have happened in the absence of a crisis. These scenarios should define counterfactual values for all crisis-sensitive predictors included in the final models, and for the population denominators. Several approaches to set counterfactual values may be used: (i) in the absence of a crisis, it may be assumed that certain predictors or types of displacement would have taken a zero value: for example, epidemics (e.g. cholera, measles) that are known to be associated with extreme food insecurity crises might not have occurred; similarly, no war-related displacement would have happened; (ii) pre-crisis values of the predictors, if available, may be adopted as counterfactuals: for some predictors (e.g. market prices), we use the local average (e.g. the district median prior to the crisis’ start); for others (e.g. rainfall), seasonality should also be considered; (iii) if no pre-crisis data are available, levels from reasonably comparable regions within the country that are not affected by the crisis may instead be considered. Table
5 shows ‘most likely’ counterfactual assumptions for the South Sudan analysis we previously conducted. To portray uncertainty in what are inevitably arbitrary assumptions, we also define reasonable best- and worst-case scenarios.

Table 5. Most likely scenario counterfactual assumptions, South Sudan (2013-2018).

| Variable                          | Counterfactual assumptions                                                                                                                                                                                                 | Notes                                                                                                                                                                                                 |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Proportion of IDPs               | The proportion of IDPs in each county would have been equal to the mean total across South Sudan in Jan 2012-Nov 2013, multiplied by the county’s mean percent share of total IDPs during Dec 2013-Apr 2018. | Assume that the relative scale of internal displacement during the war reflects each county’s general potential for displacement. Accordingly, in the counterfactual denominator IDPs are “returned” to their counties of origin pro rata to the assumption. |
| Incidence of armed conflict events | Mean of 2012-2013 level within each county, or actual level, whichever is lower.                                                                                                                                              | Assume conflict in Pibor County would have continued, as this pre-dated the current civil war. Pre-crisis baseline.                                                                                   |
| Incidence of attacks against aid workers | Mean of 2012-2013 level within each county, or actual level, whichever is lower.                                                                                                                                              | Pre-crisis baseline.                                                                                                                                                                              |
| Terms of trade purchasing power index | Mean of 2012-2013 levels per state.                                                                                                                                                                                           | Pre-crisis baseline.                                                                                                                                                                              |
| Uptake of measles routine vaccination | On an annual basis, no lower than the mean of 2012-2013 levels per county.                                                                                                                                                   | Assumption preserves any improvements in vaccination coverage observed during the crisis period in any county.                                                                                 |
| Measles incidence                | Mean of 2012-2013 level within each county, or actual level, whichever is lower.                                                                                                                                               | Pre-crisis baseline.                                                                                                                                                                              |

To propagate error in the model predictions of $y_{A,k,t}$ and $y_{C,k,t}$ into final estimates, we can set up a bootstrap simulation that, for each $z$ of $Z$ total runs, draws from error distributions to generate an excess death toll estimate:

$$\text{Eq. 6} \quad D_{E,k,t,z} = y_{A,k,t,z}N_{A,k,t} - y_{C,k,t,z}N_{C,k,t}$$

where $y_{A,k,t,z} \sim e^{\text{N}(\ln \bar{y}_{A,k,t}, \text{SE}(\ln \bar{y}_{A,k,t})^2)}$ and $y_{C,k,t,z} \sim e^{\text{N}(\ln \bar{y}_{C,k,t}, \text{SE}(\ln \bar{y}_{C,k,t})^2)}$

For each $z$, we then aggregate run outputs $\{D_{E,k,t,1}, D_{E,k,t,2}, D_{E,k,t,3}, \ldots, D_{E,k,t,Z}\}$ over $K$ and $T$ as desired, and compute point estimates and 95% confidence intervals by taking the mode, 2.5th and 97.5th percentiles of the resulting distribution of $Z$ run outputs. Partial aggregations (e.g. by region or year) can also be done. Note that if counterfactual population denominators are considerably different from the actuals (e.g. if large-scale displacement outside the region of interest has occurred), comparing actual and counterfactual mortality is fraught: we therefore scale excess death rates to the actual population denominators.
Sensitivity analyses

While a number of sensitivity analyses may be conducted to explore variability in the results as a function of parameter uncertainty, we focus here on two particularly important issues.

**Population denominator uncertainty.** Most displacement data in crisis settings do not arise from statistically robust estimation methods. Over-reporting of population figures may occur if population counts are perceived as registration for relief allocation [31]. Conversely, insecurity and lack of connectivity may result in undetected population movements. These issues compound the discrepancy in available demographic estimates. We thus explore combinations of sensitivity values for both displacement and demographic estimates (as a ratio of true to reported values, where values < 1 indicate over-reporting, and vice versa), and re-run analysis accordingly.

**Under-estimation of mortality in surveys.** In previous South Sudan work, possible under-estimation of deaths among children under 5y has been noted, as indicated by a low ratio of under 5y to all-age deaths and proportion of infant deaths (Table 6). Similar concerns have been raised in Yemen [32]. Under-reporting of infant and particularly neonatal deaths is plausible, due to stigma and/or emotional trauma associated with losing a young child or insufficient probing during questionnaire administration. The true death rate within the survey sample is equal to \( y = \frac{d_o + d_u}{n_o + n_u} \), where \( d \) is deaths, \( n \) is person-time, \( o \) means observed and \( u \) unobserved (i.e. not reported). If we assume that \( p_u \) of all deaths are not reported, \( d = d_o + d_u = d_o + dp_u \), i.e. \( d_u = d_o \frac{p_u}{1-p_u} \). It follows that \( n_u = d_u T_r f_{r,u} \), where \( f_{r,u} \) is the fraction of the recall period that unobserved deaths spend on average within the surveyed household (we assume that \( f_{r,u} = 0.25 \): if most unobserved deaths are among infants and neonates, many of these might have died shortly after birth and thus contributed little to overall person-time). For alternative values of \( p_u \), we compute \( D_{u,s} \) for each survey \( s \), and from this the mean \( \bar{d}_{u,s,i} \) per household \( i \) in the survey sample. We then simulate a large number of datasets by generating a random value of \( d_{u,s,i} \) from a Poisson distribution with mean \( \lambda_{u,s,i} \), and calculate the total \( d_{s,i} \) and \( n_{s,i} \) for both children under 5y and all ages. We use these augmented datasets to carry out further estimation steps and observe median results of the simulations.

Table 6. Average survey-estimated crude death rate, under 5 years death rate and proportion of infant deaths, by country. Numbers are the median of point estimates among available surveys, and, in parenthesis, the range of point estimates and number of surveys the statistics are based on.

| Characteristic                                      | Nigeria (2016-2018) | Somalia (2014-2018) | South Sudan (2012-2018) |
|-----------------------------------------------------|----------------------|----------------------|-------------------------|
| Eligible surveys (N)                                | 70                   | 97                   | 181                     |
| Crude death rate                                    | 0.55 (0.17 to 1.58, 70) | 0.43 (0.00 to 1.61, 97) | 0.67 (0.04 to 4.22, 181) |
| Under 5 years death rate                            | 1.14 (0.23 to 4.46, 70) | 0.66 (0.00 to 2.48, 97) | 0.72 (0.00 to 3.94, 181) |
| Proportion of <5yo deaths that were among infants <1yo| 0.35 (0.00 to 1.00, 70) | 0.43 (0.00 to 1.00, 59) | 0.33 (0.00 to 1.00, 145) |
Discussion

Advantages of the method

The approach we have described can efficiently reconstruct the evolution of mortality across long retrospective periods and large areas, including where ground data collection would be unfeasible due to inaccessibility; in South Sudan, a setting with virtually no vital events registration, our application of the method generated evidence supporting a large excess death toll (about 380,000, half attributable to intentional injuries) attributable to 5 years of war, that might otherwise have evaded historical documentation forever. Predictive models underlying the estimates have quantifiable external validity. While predictive power is ultimately their most important attribute, observing the directionality of coefficients can help to appraise internal validity, particularly if dose-response associations are noted. A known challenge of crisis-attributable mortality estimation is defining an appropriate counter-factual: our method achieves this by generating non-crisis death tolls through the same statistical processes that result in the estimate of actual mortality, yielding meaningful confidence intervals. It explicitly links the definition of the crisis with the choice of counterfactual predictors and values, drawing upon a causal framework of how excess mortality comes about and contextual understanding of the crisis itself.

Known limitations

The method's main limitations reflect sources of unknown error in input data: (i) error in the predictor data, for example arising from differences in the way predictors are measured over time or in different locations; random error would result in underestimation of associations between predictors and mortality, or ‘regression dilution’ in predictive terms; bias could cause over- or underestimation; (ii) bias in mortality data, e.g. due to problems with under-ascertainment of deaths (see above), which quality weights may reduce but not eliminate; (iii) non-parametric uncertainty around population and displacement estimates; (iv) demographic projections based on inaccurate assumed growth rates (both (iii) and (iv) are discussed in an accompanying paper); (v) inappropriate assumptions on counterfactual conditions; and (vi) omission of excess mortality among people who migrate out of the affected region (e.g. refugees), or due to long-term impacts (for example, a high frequency of sexual violence among women during wartime would lead to many unplanned births; these newborns would in turn be at higher risk of poor life outcomes over the ensuing decades). These limitations imply that estimates should be interpreted with caution, with reference to confidence intervals and after thorough exploration of uncertainty through alternative counterfactual scenarios and sensitivity analyses.

Data requirements

The method's applicability is limited by the following data requirements: (i) at least some ground mortality information arising from a population-based method of recognised validity, e.g. a survey or prospective
surveillance system. Such data should be granular in nature, i.e. representative of small geographic units and time periods (alternatively, one could use large-area surveys as long as the location of surveyed communities is reported in the dataset). Some documentation (e.g. survey reports) should be available to scrutinise methods; (ii) data covering the entirety or most of the person-time of interest for at least a few variables that may plausibly be expected to predict mortality. The system for measuring these predictors should have remained consistent over time. The pattern of data missingness should be mostly random: missingness clustered in specific areas or periods (particularly at the start or end of the time series, or where mortality data are also least available) makes imputation harder and more bias-prone; (iii) reasonable demographic estimates based on a census or similarly robust data collection exercise, performed no more than a few years prior to the analysis; in addition, data on displacement (including both the geographic unit of origin and that of arrival) covering most or all of the person-time should be available, or composable from existing reports and databases.

Minimal data requirements, e.g. how many ground surveys or predictor variables are needed, are difficult to establish \textit{a priori}: the predictive power of the model is a function not just of the amount of data, but also of the extent to which these data capture population variability and the local strength of association between predictors and mortality. As such, an additional limitation of the method is that the precision, and thus interpretability, of estimates arising from it may only become clear \textit{a posteriori}.

Computational implementation

With the exception of step 2 (population denominator reconstruction), for which only crisis-specific analysis methods appear feasible, we have developed generic R analysis scripts that implement estimation steps for any crisis setting and generate output datasets, tables and graphs (see Additional File and https://github.com/francescochecchi/mortality_small_area_estimation). The analyst interacts with these scripts through Microsoft Excel spreadsheets containing input datasets and various parameters to control the analysis.

Conclusions

We are currently testing an extension of the method for forecasting mortality over short time horizons of 3-6 months: this could provide an efficient means to do real-time estimation across the crisis-affected region, thereby generating information for decision-makers tasked with allocating humanitarian resources. Key requirements for such an application would be immediate predictor data sharing and standing capacity to implement analysis.

Other improvements to the method are worth exploring. As instances of its use accumulate, a Bayesian estimation framework specifying informative priors for key predictor coefficients (e.g. armed conflict intensity) may be attractive. Improvements to model fitting could include data mining techniques or
Bayesian model averaging. Further development of the method, however, will require dedicated resources and buy-in from humanitarian stakeholders who hold access to key input data.

List of abbreviations

| Abbreviation | Definition                                      |
|--------------|------------------------------------------------|
| CDR          | Crude death rate                               |
| CV           | Cross-validation                               |
| DSS          | Dawid-Sebastiani score                         |
| ENA          | Emergency Nutrition Assessment                 |
| IDP          | Internally displaced person                    |
| LGA          | Local Government Area (Nigeria)                |
| SMART        | Standardised Monitoring of Relief and Transitions initiative |
| U5DR         | Under 5 years death rate                       |

Declarations

Ethics approval and consent to participate

All data were previously collected for routine humanitarian response and/or health service provision purposes, and were either in the public domain or shared in fully anonymised format. The study was approved by the Ethics Committee of the London School of Hygiene & Tropical Medicine (ref. 15334), with amendments to cover implementation in different countries. Country-specific approvals came from the Nigerian Institute of Medical Research Institutional Review Board (ref. IRB/18/065) and the Research and the Ethics Review Committee of the Ministry of Health and Human Services, Somali Federal Republic (ref. MOH&HS/DGO/1944/Dec/2018). We applied to the Ethics Review Committee of the South Sudan Ministry of Health (6 Apr 2018), but did not receive a response despite repeated inquiries.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from various United Nations and non-governmental agencies, but restrictions apply to the availability of these data, which were used under
license for the current study, and so are not all publicly available. Data are however available from the authors upon reasonable request and with permission of the above agencies. Furthermore, we have uploaded curated R scripts and a set of dummy datasets on https://github.com/francescochecchi/mortality_small_area_estimation (also see Additional File). These materials should enable independent replication of all our analysis steps. Data will be made available to the extent possible as part of the publication of country-specific papers.

Competing interests
The authors declare that they have no competing interests.

Funding
This document is an output from a project funded by the UK Foreign, Commonwealth and Development Office (FCDO; formerly Department for International Development) through the Research for Evidence Division (RED) for the benefit of developing countries. However, the views expressed and information contained in it are not necessarily those of or endorsed by FCDO, which can accept no responsibility for such views or information or for any reliance placed on them. The analyses described in this report were partly funded by UK Research and Innovation as part of the Global Challenges Research Fund, grant number ES/P010873/1 (South Sudan, Somalia 2014-2018), the United States Institute of Peace (South Sudan), the UN Food and Agriculture Organisation (Somalia 2010-2012) and the Famine Early Warning Systems Network (Somalia 2010-2012).

Authors' contributions
FC designed the method, did statistical analysis and wrote this paper. AT, AG, EKB and AW contributed to study design, collected and managed data. All authors read and approved the final manuscript.

Acknowledgments
We are indebted to Dr Sandrine Foldvari Tobelem and Prof Nicholas Jewell for statistical advice and encouragement. We are also grateful to Anna Carnegie for project management support and Chris Jarvis for geospatial analysis advice.

Disclaimer
Geographical names and boundaries presented in this report are used solely for the purpose of producing scientific estimates, and do not necessarily represent the views or official positions of the authors, the
London School of Hygiene and Tropical Medicine, any of the agencies that have supplied data for this analysis, or the donor. The authors are solely responsible for the analyses presented here, and acknowledgment of data sources does not imply that the agencies or individuals providing data endorse the results of the analysis.

References

1. Checchi F. Estimation of population mortality in crisis-affected populations: Guidance for humanitarian coordination mechanisms. Geneva: World Health Organization; 2018. https://www.who.int/health-cluster/resources/publications/FLSHTM-Mortality-Estimation-Options-oct2018.pdf.

2. Heudtlass P, Speybroeck N, Guha-Sapir D. Excess mortality in refugees, internally displaced persons and resident populations in complex humanitarian emergencies (1998-2012) - insights from operational data. Confl Health. 2016;10:15.

3. Checchi F, Roberts L. Documenting mortality in crises: what keeps us from doing better. PLoS Med. 2008;5:e146.

4. Checchi F, Warsame A, Treacy-Wong V, Polonsky J, van Ommeren M, Prudhon C. Public health information in crisis-affected populations: a review of methods and their use for advocacy and action. Lancet. 2017;390:2297–313.

5. Ball P, Betts W, Scheuren F, Dudukovich J, Asher J. Killings and Refugee Flow in Kosovo, March - June 1999: A Report to the International Criminal Tribunal for the Former Yugoslavia. Washington, DC: American Association for the Advancement of Science; 2002. http://www.icty.org/x/file/About/OTP/War_Demographics/en/s_milosevic_kosovo_020103.pdf. Accessed 6 Apr 2021.

6. Degomme O, Guha-Sapir D. Patterns of mortality rates in Darfur conflict. The Lancet. 2010;375:294–300. doi:10.1016/S0140-6736(09)61967-X.

7. Depoortere E, Checchi F, Broillet F, Gerstl S, Minetti A, Gayraud O, et al. Violence and mortality in West Darfur, Sudan (2003-04): epidemiological evidence from four surveys. Lancet. 2004;364:1315–20.

8. Coghlan B, Ngoy P, Mulumba F, Hardy C, Bemo VN, Stewart T, et al. Update on mortality in the Democratic Republic of Congo: results from a third nationwide survey. Disaster Med Public Health Prep. 2009;3:88–96.
9. Hagopian A, Flaxman AD, Takaro TK, Esa Al Shatari SA, Rajaratnam J, Becker S, et al. Mortality in Iraq associated with the 2003-2011 war and occupation: findings from a national cluster sample survey by the university collaborative Iraq Mortality Study. PLoS Med. 2013;10:e1001533.

10. Iraq Family Health Survey Study Group, Alkhuzai AH, Ahmad IJ, Hweel MJ, Ismail TW, Hasan HH, et al. Violence-related mortality in Iraq from 2002 to 2006. N Engl J Med. 2008;358:484–93.

11. Checchi F, Robinson, Courtland. Mortality among populations of southern and central Somalia affected by severe food insecurity and famine during 2010-2012 - Somalia. Nairobi: Food and Agriculture Organization; 2013. https://reliefweb.int/report/somalia/mortality-among-populations-southern-and-central-somalia-affected-severe-food. Accessed 11 Jan 2021.

12. Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: A county-level analysis. PLoS Med. 2021;18:e1003571. doi:10.1371/journal.pmed.1003571.

13. Koum Besson ES, Norris A, Bin Ghouth AS, Freemantle T, Alhaffar M, Vazquez Y, et al. Excess mortality during the COVID-19 pandemic: a geospatial and statistical analysis in Aden governorate, Yemen. BMJ Glob Health. 2021;6:e004564. doi:10.1136/bmjgh-2020-004564.

14. Imperial College COVID-19 Response Team, Watson OJ, Alhaffar M, Mehchy Z, Whittaker C, Akil Z, et al. Leveraging community mortality indicators to infer COVID-19 mortality and transmission dynamics in Damascus, Syria. Nat Commun. 2021;12:2394. doi:10.1038/s41467-021-22474-9.

15. Toulemon L, Barbieri M. The mortality impact of the August 2003 heat wave in France: Investigating the ‘harvesting’ effect and other long-term consequences. Population Studies. 2008;62:39–53. doi:10.1080/00324720701804249.

16. Sandberg J, Santos-Burgoa C, Roess A, Goldman-Hawes A, Pérez CM, Garcia-Meza A, et al. All Over the Place?: Differences in and Consistency of Excess Mortality Estimates in Puerto Rico After Hurricane Maria. Epidemiology. 2019;30:549–52.

17. Working Group for Mortality Estimation in Emergencies. Wanted: studies on mortality estimation methods for humanitarian emergencies, suggestions for future research. Emerg Themes Epidemiol. 2007;4:9.

18. Checchi F, Testa, Adrienne, Warsame, Abdihamid, Quach, Le, Burns, Rachel. Estimates of crisis-attributable mortality in South Sudan, December 2013- April 2018: A statistical analysis - South Sudan. ReliefWeb. 2018. https://reliefweb.int/report/south-sudan/estimates-crisis-attributable-mortality-south-sudan-december-2013-april-2018. Accessed 11 Jan 2021.
19. Rao JNK, Molina I. Small Area Estimation: Rao/Small Area Estimation. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2015. doi:10.1002/9781118735855.

20. Seal A, Checchi F, Balfour N, Nur A-R, Jelle M. A weak health response is increasing the risk of excess mortality as food crisis worsens in Somalia. Conflict and health. 2017;11:12.

21. Komakech H, Atuyambe L, Orach CG. Integration of health services, access and utilization by refugees and host populations in West Nile districts, Uganda. Confl Health. 2019;13:1. doi:10.1186/s13031-018-0184-7.

22. Standardised Monitoring and Assessment of Relief and Transitions (SMART). Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: SMART Methodology. https://smartmethodology.org/. Accessed 14 Feb 2021.

23. Cairns KL, Woodruff BA, Myatt M, Bartlett L, Goldberg H, Roberts L. Cross-sectional survey methods to assess retrospectively mortality in humanitarian emergencies. Disasters. 2009;33:503–21.

24. Altare C, Guha-Sapir D. The Complex Emergency Database: a global repository of small-scale surveys on nutrition, health and mortality. PLoS One. 2014;9:e109022.

25. Erhardt J. Emergency Nutrition Assessment (ENA) Software for SMART. 2020. https://smartmethodology.org/survey-planning-tools/smart-emergency-nutrition-assessment/.

26. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. https://www.R-project.org/.

27. Prudhon C, de Radiguès X, Dale N, Checchi F. An algorithm to assess methodological quality of nutrition and mortality cross-sectional surveys: development and application to surveys conducted in Darfur, Sudan. Popul Health Metr. 2011;9:57.

28. Grellety E, Golden MH. Change in quality of malnutrition surveys between 1986 and 2015. Emerg Themes Epidemiol. 2018;15:8.

29. Maxwell D, Gottlieb G, Coates J, Radday A, Kim J, Venkat A, et al. Humanitarian Information Systems: Anticipating, Analyzing, and Acting in Crisis. Tufts - Feinstein International Center. https://fic.tufts.edu/research-item/the-constraints-and-complexities-of-information-and-analysis/. Accessed 14 Feb 2021.

30. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 2007;102:359–78. http://www.jstor.org/stable/27639845. Accessed 22 May 2021.
31. Harrell-Bond B, Voutira E, Leopold M. Counting the Refugees: Gifts, Givers, Patrons and Clients. J Refugee Stud. 1992;5:205–25. doi:10.1093/jrs/5.3-4.205.

32. Maxwell D, Hailey P, Spainhour Baker L, Kim JJ. Constraints and Complexities of Information and Analysis in Humanitarian Emergencies: Evidence from Yemen. Feinstein International Center, Tufts University and Centre for Humanitarian Change; 2019. https://fic.tufts.edu/publication-item/famine-and-analysis-in-yemen/. Accessed 6 Apr 2021.
Figure 1. Illustration of actual and counterfactual mortality during and after a hypothetical crisis.

Figure 2. Coverage of SMART mortality surveys, by region, district and month in Somalia. Heat colours denote months falling within the recall period of one or more surveys, with increasing colour intensity proportional to the precision of estimates.

Figure 3. Trends in selected survey-estimated indicators, South Sudan, 2013-2018. Each dot-line segment denotes the recall period of one survey. All indicators include all age groups except for the U5DR graph (top right panel).

Figure 4. Predicted versus observed numbers of deaths per stratum (county), South Sudan, 2013-2018, based on 10-fold cross-validation. The red line indicates perfect fit.
Illustration of actual and counterfactual mortality during and after a hypothetical crisis.
Figure 2

Coverage of SMART mortality surveys, by region, district and month in Somalia. Heat colours denote months falling within the recall period of one or more surveys, with increasing colour intensity proportional to the precision of estimates.
Figure 3

Trends in selected survey-estimated indicators, South Sudan, 2013-2018. Each dot-line segment denotes the recall period of one survey. All indicators include all age groups except for the U5DR graph (top right panel).
Figure 4

Predicted versus observed numbers of deaths per stratum (county), South Sudan, 2013-2018, based on 10-fold cross-validation. The red line indicates perfect fit.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- checchimortalitysasmethodsadditionalfile.docx