Measuring the Impact of COVID-19 on Mortality: rethinking sampling strategies

An overview of sampling strategies in major household surveys

Prof. Dr. A. Sinan Türkyılmaz
Head of Social Research Methodology Department
Hacettepe University, Institute of Population Studies, Ankara, Turkey

Tuesday, 5 July 2022
Introduction and Outline

• The aim of this presentation is to briefly mention about sampling strategies of major large-scale demographic surveys where basic child and adult mortality indicators that can be derived.

• The Outline of presentation
  • Major large-scale demographic surveys
  • Basic Childhood and Adult Mortality Indicators and Brief explanation on calculation procedures and data needs
  • General sample design and sample size calculation issues on major household surveys
  • Discussion on sample designs under consideration of mortality estimates
    • Mortality is a rare event for sample designs
    • Estimates on Disaggregation and/or sub-populations
  • Ideas on sampling issues with for countries with declined(ing) mortality/fertility rates
    • Is over-sampling strategy possible for household surveys to pre-determine the households with recent deaths?
  • Impact on Covid-19 on mortality estimates
Some of major household sample survey programs those use Infant and child mortality estimation processes and the adult mortality estimation processes

- Demographic and Health Surveys (DHS)
- Multiple Indicator Cluster Surveys (MICS)
- Reproductive Health Surveys (RHS)
- World Health Survey (WHS)
- Living Standards Measurement Studies (LSMS)
- Pan Arab Programme on Family Health (PAPFAM/PAPCHILD)
- Gulf Health Surveys (GHS)
- Contraceptive Prevalence Surveys (CPS)
- World Fertility Surveys (WFS)
Childhood mortality indicators and direct estimates

- **Under-five mortality (U5MR or 5q0):** the probability of dying between birth and exact age 5 years;
  - **Child mortality (CMR or 4q1):** the probability of dying between exact ages 1 and 5 years;
- **Infant mortality (IMR or 1q0):** the probability of dying between birth and exact age 1 year;
  - **Neonatal mortality (NN):** the probability of dying between birth and exact age 1 month;
  - **Post neonatal mortality (PNN):** the difference between infant and neonatal mortality.

- **Direct estimation methods:** Observed or reported data are used on survivorship of individual children, including data on their:
  - date of birth,
  - survival status
  - and date of death or age at death for those who have

- **Indirect estimation methods:** Modelling techniques were applied to cumulative data
Adult mortality indicators

- **Adult mortality to age 60 (45q15):** the probability of dying between exact ages 15 and 60 years;

- **Adult mortality to age 50 (35q15):** the probability of dying between exact ages 15 and 50 years.

- **Maternal mortality rate** (reported per 1,000 women aged 15-49) and the **maternal mortality ratio** (MMR, reported as the number of maternal deaths among women aged 15-49 per 100,000 live births).
Adult mortality estimation methods

• Sibling History(Direct Estimates): The sibling histories collect information for each birth or pregnancy that the respondent’s mother ever had, and usually include at a minimum:
  • Sex of the sibling;
  • Survival status of the sibling – alive or dead;
  • Age of sibling for surviving siblings;
  • Age at death for siblings who died;
  • Years since death or year of death for siblings who died.

• Recent Household Deaths(Direct Estimates): The data requirements are quite simple. For each person that died in the time period, the following information is elicited:
  • Sex of the person;
  • Age at death;
  • Date of death.

• Indirect Estimation: The concept of survivorship of relatives
  • Orphanhood (or survivorship of parents),
  • Widowhood (or survivorship of spouse)
  • Sibling survivorship.
General sample design and sample size calculation issues on major household surveys

Sample size determination: The size of the sample is perhaps the most important parameter of the sample design, because it affects the precision, cost and duration of the survey more than any other factor. Sample size must be considered both in terms of the available budget for the survey and its precision requirements. (MICS Sampling Manual)

To calculate the sample size, using the appropriate mathematical formula, requires that several factors be specified and values for others be assumed or taken from previous or similar surveys. These factors are:

- The level of precision required, measured by the sampling error, (RME)
- The level of confidence desired (t value, $\alpha=0.05$)
- The estimated (or known) proportion of the population in the specified target group ($pb$)
- The predicted or estimated rate, or prevalence, for the specified indicator ($r$)
- The sample design effect due to complex sample design ($deff$)
- The average household size ($Avesize$)
- An adjustment for potential loss of sample households due to non-response. ($RR$)

Confidence limits (95% confidence):
Upper: $r \times (1 + RME)$
Lower: $r \times (1 - RME)$

Sample size:
$$n = \frac{4 \times r \times (1-r) \times deff}{(RME \times r)^2 \times pb \times AveSize \times RR}$$

Standard error (se):
$$(r \times RME) / 2$$
Defining And Choosing The Key Indicator To Calculate Sample Size and Mortality is a rare event for sample designs

- According to the MICS Sampling Manual, the recommended strategy for calculating the sample size is to choose an important indicator that will yield the largest sample size. This will mean first choosing a target population that comprises a small proportion of the total population (r in the above formula).

- However, the infant mortality rate (IMR) or the maternal mortality ratio (MMR) are not mentioned as candidates for the key indicator. This is because the sample sizes that would be necessary to measure these indicators are much too large – in the tens of thousands – and it would be impractical to consider them.

- This does not necessarily mean that such indicators should not be measured in the survey, but rather that the sample size for the survey should not be based on them. The survey results for these indicators will have larger sampling errors and, hence, wider confidence intervals than the other indicators. Because the sample sizes are relatively small, the sampling errors around child mortality estimates are quite large. DHS and MICS reports provide sampling errors for child mortality rates in the appendices of each final report.
Reference dates, Disaggregation and/or sub-populations, Missing data

• In order to increase the sample size and reduce the sampling errors, direct mortality estimates generally calculated for periods (mostly 5 years period for national and 10 years for sub-national). In other words, mortality estimates refers a date which is 2.5 years prior to survey date, if 5 years period estimates were used.

• Disaggregating the results into smaller groups, such as regions, education groups even urban/rural increases the sampling errors further. Even a survey provides a sample size that is reasonable for national estimates of child mortality for a five-year period, but the sampling errors quickly grow as the sample size is reduced or the data are disaggregated.

• Several other dimensions can be considered as potential variables for the disaggregation of mortality rates, including socio-economic characteristics such as urban/rural residence, region of residence, level of education, household income, household wealth quintiles, ethnicity, language, religion or occupation.
Reference dates, Disaggregation and/or sub-populations, Missing data

• All variables in survey datasets may potentially contain missing or other special values. Decisions concerning the handling of missing data and special values essentially fall into two categories:
  • Exclude cases with missing data or other special values from analysis;
  • Replace the missing data or other special values with an in-range value, based on a particular set of rules how to replace or impute a new value for the values with missing or other special responses.

• Certain rules have been used in some survey programs to set minimum sample sizes for displaying mortality rates. DHS uses the following rule concerning the presentation of mortality estimates: “Rates based on 250 to 499 unweighted exposed persons should be shown in parentheses. Rates based on fewer than 250 unweighted exposed persons should not be shown (*) and appropriate footnotes should be added to the tables if either or both of these cases exist.”
## Sample Size and precision scenarios for different Infant Mortality levels

### Target sample size (Number of household)

<table>
<thead>
<tr>
<th>Country</th>
<th>r (IMR)</th>
<th>RME</th>
<th>n</th>
<th>Upper</th>
<th>Lower</th>
<th>se</th>
<th>Number of household members</th>
<th>Number of women aged 15-49 years</th>
<th>Number of children aged 0-4 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.005</td>
<td>0.15</td>
<td>147,407</td>
<td>0.004</td>
<td>0.006</td>
<td>0.000</td>
<td>530,667</td>
<td>132,667</td>
<td>53,067</td>
</tr>
<tr>
<td>B</td>
<td>0.020</td>
<td>0.15</td>
<td>36,296</td>
<td>0.017</td>
<td>0.023</td>
<td>0.002</td>
<td>130,667</td>
<td>32,667</td>
<td>13,067</td>
</tr>
<tr>
<td>C</td>
<td>0.050</td>
<td>0.15</td>
<td>14,074</td>
<td>0.043</td>
<td>0.058</td>
<td>0.004</td>
<td>50,667</td>
<td>12,667</td>
<td>5,067</td>
</tr>
<tr>
<td>D</td>
<td>0.100</td>
<td>0.15</td>
<td>6667</td>
<td>0.085</td>
<td>0.115</td>
<td>0.008</td>
<td>24,000</td>
<td>6,000</td>
<td>2,400</td>
</tr>
<tr>
<td>F</td>
<td>0.200</td>
<td>0.15</td>
<td>2963</td>
<td>0.170</td>
<td>0.230</td>
<td>0.015</td>
<td>10,667</td>
<td>2,667</td>
<td>1,067</td>
</tr>
</tbody>
</table>

### Design effect (deff) pb AveSize RR

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.5</td>
<td>0.100</td>
<td>AveSize</td>
<td>0.9</td>
</tr>
</tbody>
</table>

### Confidence limits (95% confidence): Upper: \( r \cdot (1 + \text{RME}) \)
Lower: \( r \cdot (1 - \text{RME}) \)

### Sample size:

\[
n = \frac{4 \cdot r \cdot (1-r) \cdot \text{deff}}{(\text{RME} \cdot r)^2 \cdot \text{pb} \cdot \text{AveSize} \cdot \text{RR}}
\]

### Standard error (se):

\[
(se) = \frac{r \cdot \text{RME}}{2}
\]

Please note that: Calculations are based on 5 years period estimates. Proportion of under-five, number of women aged 15-49, average household size and response rate also decreases when mortality and fertility rates are decreasing or already low.
An idea on sampling issues with for countries with declined(ing) mortality/fertility rates

• Low and/or decreasing mortality and fertility rates are challenging from statistical perspective to have adequate number of observations on classical sampling approaches.

• There are already over-sampling strategies be used in Global surveys such as in MICS to increase number of under-five children: “Oversampling Of Children Under-five In Low Fertility Settings (Megill D., Khan SM., and Hancioglu A. (2018).”
  • Oversampling of households with under-fives follows a similar approach to the standard MICS approach, wherein sample sizes and the number of EAs are first calculated, although it is necessary to consider the oversampling strategy on the number of under-fives expected in the sample. Sampling specialists first examine the number of under-fives that a usual sample would yield and then examine various oversampling rates to determine the gains of the strategy.

• After the census enumeration areas are selected (i.e. the first stage sampling), for the listing of the households, the listing form is designed to identify which households have children under-five. Then, the surveys select these households with a higher sampling rate compared to the households without children under-five.
Is over-sampling strategy possible for household surveys to pre-determine the households with recent deaths?

- Can MICS over-sampling be adopted to household surveys to measure adult and/or overall mortality?

  - A similar screening during household listing prior to the data collection can be done and households be asked whether any deceased member in recent period (last two years? Last year?, ??).

  - This may be more effective for adult mortality. If regular MICS strategy be applied, cases of Under-five (and in-directly households with 15-49 aged women) should be already higher.

  - The MICS strategy of oversampling households with children under-five is recommended only if a country has a relatively low fertility rate and/or a small average household size. So, this can be examined in such countries. However, over-sampling of households with deceased members can also be used in moderate, even high mortality countries since adult mortality in general a rare event.
Is over-sampling strategy possible for household surveys to pre-determine the households with recent deaths?

• Risks? Several or additional risk as observed MICS over-sampling strategy including:
  • Since it is necessary to identify households with deceased members in recent period in the listing, the listing operation is more complex, taking more time to solicit more information from respondents in the households. It may not be possible to contact some households during the listing, and the secondary source of data, such as neighbors, may be unreliable.
  
  • Further, when listings are done much in advance of the fieldwork, deaths happened out of the specified age groups and membership issues. These can result in misclassification, which can impact the variability of weights and thus, increase the design effects and sampling errors.

  • As the sample design increases the proportion of one age/sex group(s) (deceased people), the overall effect may be a decrease or increase in other age groups selected for the survey. This may be especially true for sexes/ages with higher mortality. Such changes in the distribution of the sample population can create imbalances in the overall age structure of the unweighted sample, and weights must be calculated based on the different sampling rates to compensate.

  • While the oversampling strategy will improve the sample size for mortality indicators, at the same time the strategy will increase the variability in the weights, which is marked by an increase in the design effects and the sampling errors for other indicators.
General Mortality trends and Impact on Covid-19 on mortality estimates

• Since 1990, the global under-5 mortality rate has dropped by 60%, from 93 deaths per 1,000 live births in 1990 to 37 in 2020. Children in sub-Saharan continued to have the highest rates of mortality in the world at 74 (68-86) deaths per 1000 live births.

• COVID-19 disease has a modest direct impact on child mortality, however there are on-going concerns about the indirect impact on child mortality. In 2020, increased under-5 deaths were anticipated from the repercussions of strained and under-resourced health systems, limitations on care-seeking and preventative measures like vaccination and nutrition supplements, or socioeconomic strains on households resulting from job losses, economic contractions or even deaths of parents due to COVID-19.

• Challenge: How to adopt household surveys to measure declining mortality as well as excess mortality due to pandemic?
References

- Population Division Technical Paper, No. 2011/2, “Mortality estimates from major sample surveys: towards the design of a database for the monitoring of mortality levels and trends”.
- [https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death#:~:text=Since%201990%2C%20the%20global%20under,1%20in%202027%20in%202020.](https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death#:~:text=Since%201990%2C%20the%20global%20under,1%20in%202027%20in%202020.)
- [https://data.unicef.org/topic/child-survival/under-five-mortality/](https://data.unicef.org/topic/child-survival/under-five-mortality/)
Thanks!

sinanturkyilmaz@yahoo.com

Prof. Dr. A. Sinan Türkyılmaz
Head of Social Research Methodology Department
Hacettepe University, Institute of Population Studies, Ankara, Turkey

Tuesday, 5 July 2022