Developing Data Appraisal and Assessment Criteria for Reviewing Data Sources

Dr. Aoife A. Kervick, Dan Ryan, and Dr. Kiran Sarma

Draft submitted: 22/02/2017

Table of Contents

1. Background

Low- and Middle-Income Countries (LMICs) are subject to disproportionately high rates of drowning mortality (e.g. Hyder et al., 2008; Peden & McGee, 2010). Amongst a range of key objectives recommended by the World Health Organisation (WHO) to effectively address this global health issue, is the development of National Water Safety Plans (WHO, 2014). In order to develop such a plan, accurate, timely data, detailing the true nature and extent of drowning rates and risk is necessary. Likewise, should drowning prevention policies and interventions be implemented as part of this plan, high-quality mortality data are then needed to critically evaluate their impact.

At present, a series of multi-country, mortality datasets, featuring drowning fatality data, are available for use. Amongst others, key datasets include those provided through the World Health Organisation (WHO) database[1], the Global Burden of Disease (GBD) study[2], INDEPTH Network[3], and the International Disaster Database (EM-DAT)[4].

The data provided by such sources, and particularly data from LMICs, are subject to certain limitations however (e.g. see WHO, 2010; 2014). Insufficient resources in LMICs for example, may result in fatalities data being modeled and estimated for certain countries, rather than collected and documented by a civil registration unit. Similarly, some data sources may draw from media reports (e.g. the EM-DAT) in place of definitive vital statistics to populate their multi-country dataset. Insufficient or inaccurate data may lead to ineffective or inappropriate drowning prevention measures.

As such, before drawing from these data sources to inform a National Water Safety Plan, it is crucial to identify the strengths and limitations of each dataset, and acknowledge and account for these in any future work. In the absence of a formal evaluative tool to guide this process, the objective of the current exercise is to devise prospective data appraisal and assessment criteria for mortality data sources, featuring drowning data.

2. Methodological Approach

In order to create data appraisal and assessment criteria for reviewing drowning data sources, a four-stage process was adopted (see Figure 1 below).

[1] WHO Global Mortality Database: http://www.who.int/healthinfo/mortality_data/en/
[2] Global Burden of Disease Study: http://www.healthdata.org/gbd
[3] INDEPTH Network: http://www.indepth-network.org/about-us
[4] The International Disasters Database: http://www.emdat.be/

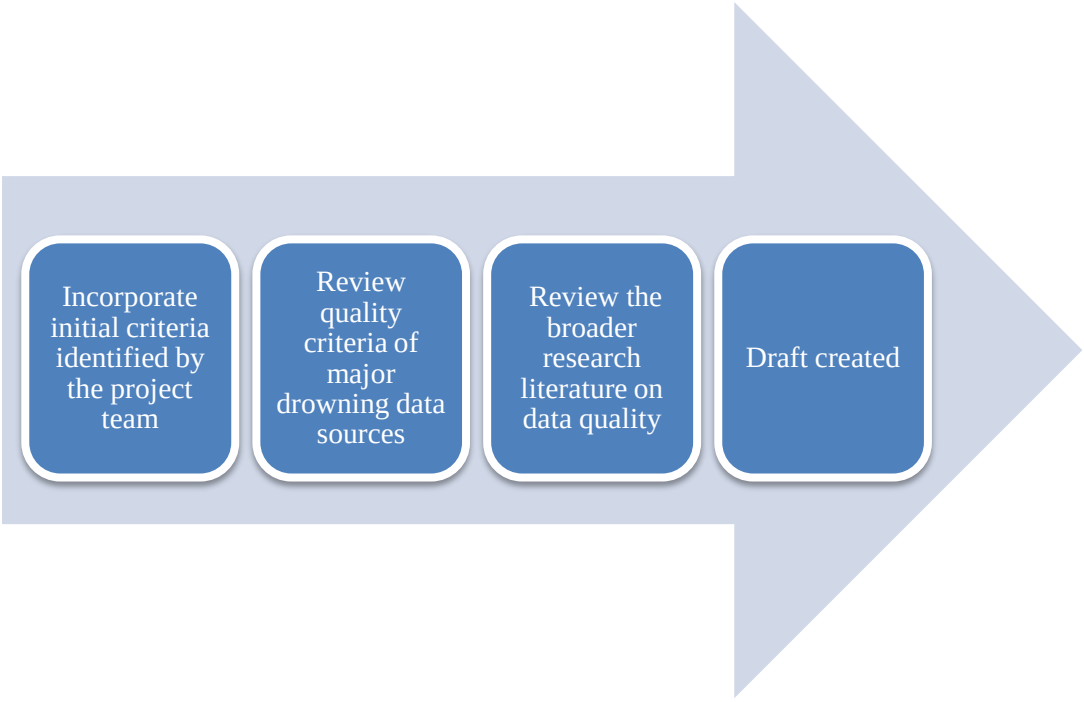

Figure 1. *Approach to creating data appraisal and assessment criteria for review drowning data sources.*

Stage 1: Institutional knowledge of project team

Based on a consultative process within the RNLI and NUI Galway, an initial set of quality assessment criteria were identified, and specified in the guiding objective for this body of work, including the following:

1. Is the dataset drowning-specific?
2. How is drowning defined in the context of the dataset?
3. How has the data been collected?
4. What is the geographic coverage of the dataset?
5. What is the time series of the data?
6. What demographic stratification is provided?
7. What key metadata can be gleaned from the source?

Stage 2: Review of quality criteria of major mortality data sources, featuring drowning data

A review of any data quality criteria stipulated by four, major, multi-country data sources featuring drowning data (the WHO, GBD study, INDEPTH network and EM-DAT) was then conducted, with key assessment items/criteria extracted primarily from online material provided by their host websites, or linked webpages. The importance of the use of error detection software when entering fatalities data was incorporated from reading about the WHO's ANACoD ('Analysing mortality level

and cause-of-death') electronic tool for example, in order to minimise inaccuracies in data entry.

Stage 3: Review of academic literature

Publications from academic journals, conferences, or technical reports on drowning, and broader injury data quality assessments were then examined, following a systematic search of this research literature. Once more, potentially relevant data appraisal and assessment criteria were identified and extracted from this body of reviewed research. Key considerations here included the importance of assessing the amount of ill-defined or 'dump/garbage' codes (see ICD-10, Chapter XVIII) within datasets using ICD-10 coding for example, to attest to the overall quality of cause of death reporting (Philips et al., 2014).

Stage 4: Draft appraisal and assessment criteria created

Following these reviews, extracted data appraisal and assessment criteria for potential user consideration were synthesised as key themes. These consist of: Metadata, Drowning Data Characteristics, Data Coverage and Completeness, Data Quality, and Data Accessibility (described in greater detail in Section 3 below). For each theme, a series of key, probative questions designed to evoke the suitability of the data source being examined for the user, are provided. Depending on the needs and criteria of the particular user (e.g. they may accept estimate-based mortality data, whether others may not), the answers to these questions should provide them with sufficient information to judge how suitable the data being assessed are, for their use.

3. Data Appraisal and Assessment Criteria

A. Metadata

Metadata are included in the data appraisal and assessment criteria, as, while not necessarily indicative of quality, they provide additional information about the data origins, and essential characteristics, which should be noted in any formal data evaluation.

A1. Who owns these data?

A2. Who is the author(s) of these data?

A3. Who funded/funds this dataset?

A4. What is the time series of these data?

A5. When were these data initially published/made available?

A6. When were these data most recently updated?

B. Drowning Data Characteristics

The characteristics of the drowning data featured in the mortality dataset is critical when assessing different sources for use. The ability to isolate drowning specific fatalities, and compare how these are defined and categorised, depending on the needs of the user, is of key importance here.

B1. Is it possible to isolate drowning fatalities within this dataset?

B2. What definition is used by the source for to define a drowning fatality?

B3. Do these data feature formal codes for drowning? [e.g. ICD-10 codes]

- If so, what codes are used? [e.g. list all ICD-10 codes used]
- If not, does it distinguish between different types of drowning fatalities? And, if so, what definitions are used to describe these different drowning incidents?

B4. If available, what proportion of drowning fatalities are categorised as undetermined/unspecified?

C. Data Coverage and Completeness

The extent to which the mortality data being assessed provides data for the country/region(s) the user is seeking to examine (referred to as 'data coverage') is a key consideration when assessing mortality data for use. Likewise, the extent to which the data provided for the user's area(s) of interest, captures all of the drowning deaths that occurred during a certain timeline (i.e. 'data completeness'), should inform data assessments. Both provide indications of the accuracy and representativeness of the drowning-related mortality data captured.

C1. What is the geographic coverage of this dataset, relative to the user's area(s) of interest?

- Clarify for which countries/regions complete versus estimated data are available.

C2. Is population data available for each country/region covered, for comparison with mortality data?

- If not, are estimates of this available?

C3. How complete is the mortality data? That is, to what extent are the deaths that occur in the country/region documented by a civil registration system/data capture tool.

- If this information is available, how was this calculated (e.g. the Bennett-Horiuchi method[5])?

D. Data Quality

The quality of the mortality data provided by the source is another key consideration when assessing data for use. Of particular importance, the tools and methods used to collect/generate mortality data should be examined, including the protocol surrounding the assessment of raw, source data incoming to (or being reviewed within) the database, when assessing mortality data quality for use.

Data Collection and Entry

D1. How are data collected for this dataset? [e.g. from demographic surveys, verbal autopsy reports, calculated estimates etc.]

- Are these standardised (i.e. established, or validated) measures/tools?

D2. Are data estimates included in this dataset?

- If so, how are these estimates calculated?
- If so, are uncertainty/error intervals calculated for the estimated data? And, if so, how are these calculated?

D3. Who assigns the definitive cause of death to the mortality data documented in this dataset?

- What level of training do they have?
- If using ICD-10 codes, is the full (i.e. not summary) list, and native language version, used?

D4. Is software used to detect errors when entering the data by the source? [e.g. missing data, extreme outliers, inconsistent trends]

D5. Is a data dictionary/glossary provided for the different variables included in the dataset?

D6. Are key mortality variables (which also facilitate demographic stratification, such as age, gender) tabulated/recorded in this dataset?

D7. What is the dataset's 'timeliness', i.e. what is the span of time from the reporting period, to the dissemination of the dataset?

[5] The Bennett-Horiuchi method is based on the premise that for any population, the entry rate minus the growth rate must be equal to the exit or death rate. Differences between the entry rate and the growth rate (which is a residual estimate of the exit rate calculated from census age distributions), and the exit rate (calculated from information on deaths by age) identify differences in the reporting of population and of deaths. The magnitude of the inconsistency can be interpreted as a measure of completeness of death reporting relative to population reporting.

Data Quality Assessments

D8. Is there a quality assessment protocol for evaluating the data received by the source?

- If so, describe this tool/process etc. [e.g. are trends compared to previous/additional datasets to assess consistency?]

D10. Has the dataset been reviewed and/or updated at any point/regularly?

- If so, describe this process, including any review/update time schedules.

D11. Has the quality of the source's cause of death coding been evaluated?

- If so, describe this process, and whether it was found to be satisfactory.

D12. What proportion of the data are deaths assigned to ill-defined cause of death or 'dump/garbage' codes (see ICD-10, Chapter XVIII) for unspecified/undetermined fatalities?

D13. Are any limitations of these data/the dataset listed by the source?

- If so, summarise these.

E. Accessibility

A final consideration when assessing mortality data for drowning prevention users is how accessible it is. How easy the data are to access, interact with, and use will ultimately influence how useful it is for those seeking to use it.

E1. Can the selected data be accessed by the user?

E2. Can the selected data be accessed within the timeframe required by the user?

E3. Is guidance documentation provided to facilitate the download/analysis process?

E4. Does the data require expertise (e.g. experience using SPSS or Stata) to manage and analyse?

References

Hyder, A. A., Borse, N. N., Blum, L., Khan, R., El Arifeen, S., & Baqui, A. H. (2008). Childhood drowning in low- and middle-income countries: Urgent need for intervention trials. *Journal of Pediatrics and Child Health, 44*(4), 221- 217. Doi: 10.1111/j.1440-1754.2007.01273.x.

Peden, M. M., & McGee, K. (2003). The epidemiology of drowning worldwide. *Injury Control and Safety Promotion, 10*(4), 195-199. Doi: 10.1076/icsp.10.4.195.16772.

Philips, D. E., Lozano, R., Naghavi, M., Atkinson, C., Gonzalez-Medina, D., Mikkelsen, L., Murray, C. J. L., & Lopez, A. D. (2014). A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Population Health Metrics, 12*(14), 1-16. Doi: 10.1186/1478-7954-12-14.

World Health Organisation (2014). *Global report on drowning – Preventing a leading killer*. Retrieved from: http://www.who.int/violence_injury_prevention/global_report_drowning/en/

World Health Organisation (2010*). Improving the quality and use of birth, death and cause-of-death information: Guidance for a standards-based review of country practices.* Retrieved from: http://apps.who.int/iris/handle/10665/44274