POPULATION HEALTH METRICS

RESEARCH

Open Access

Information–processing methods for mortality surveillance in the presence of varying levels of completeness and ill–defined codes of causes of death – the case of Brazil

Antony Stevens[*], Maria Inês Schmidt and Bruce B Duncan

Abstract

Background: The World Health Organization has developed proposals on how efforts to reduce non–communicable diseases (NCD) in low– and middle–income countries may be monitored over time. One of the proposed indicators is the unconditional probability of death between the ages of 30 and 70 from any of the four main groups of non–communicable diseases – cardiovascular disease, cancer, chronic respiratory disease, and diabetes. Our objective is to describe information–processing methods developed to facilitate this monitoring of mortality over time for Brazil.

Methods: We developed an IPython Notebook which incorporates mortality records, population sizes, estimates of sub notification, redistribution of ill–defined causes of death, international disease codes, and world standard population weights for five–year age group, gender, state, and year strata. The approach permits flexibility in the incorporation of different estimates of sub–notification and ill–defined causes of death. The main output is a "Basic Sheet", where each line provides corrected deaths by disease categories and denominators for a given stratum. This sheet is then used to generate desired statistics.

Results: This collection of shareable computer code and data organizes the approach necessary for calculations, making the data available to interested parties for the remaining relatively simple calculations. The mortality statistic suggested by the World Health Organization is derived from this sheet.

Conclusions: The approach developed is an additional step toward rapid and accurate reporting of Brazilian NCD mortality data. The code is available and may be adapted by others facing similar tasks.

Keywords: Chronic disease, Surveillance, Mortality, Cardiovascular diseases, Diabetes mellitus, Respiratory tract diseases, Neoplasms

Background

The growing disease burden due to the non–communicable diseases (NCD) [1] has led the World Health Organization (WHO) [2-4] and the United Nations [5] to propose an international public health focus on the prevention and control of NCDs [6-8]. Within these proposals, emphasis is placed on establishing and monitoring NCD programs in low– and middle–income countries, where the disease burden is rapidly rising and less awareness of the magnitude of the problem exists.

The WHO has emphasized the necessity of strengthening disease–and risk–factor–monitoring systems as a major part of this public health effort [2] and is in the process of finalizing recommendations for indicators by which to monitor the success of NCD prevention and control measures in individual countries. Trends in mortality are central to any evaluation of NCD control efforts, and one of the proposed indicators relates to mortality – the unconditional probability of dying between 30 and 70 years due to any of the four main groups of NCDs – cardiovascular disease, cancer, chronic respiratory disease, and diabetes [9].

Yet major inadequacies are present in the mortality registry systems of most low– and middle–income

*Correspondence: antony.stevens@gmail.com
Postgraduate Program in Epidemiology, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

countries, currently making it difficult, if not impossible, to accurately measure trends in NCD death rates. To adequately monitor mortality, these systems will need to be improved, and means developed to estimate trends in the face of the two major problems in the quality of mortality reporting – sub–notification of deaths and ill–defined causes of reported deaths.

Brazil has invested in improving the quality of its mortality reporting over the past decade. Ill–defined causes (ICD–10 chapter XVIII) have fallen from 14% in 2000 to 7% in 2010. Sub–notification, though more difficult to accurately quantify, has fallen from an estimated 14% in 2000 to 6% in 2010 [10]. Thus mortality rates and trends are being reported with greater precision [11-13].

Within this latter effort, we at the Ministry's Collaborative Center in Surveillance of Diabetes, Cardiovascular and Other Chronic Diseases, have been engaged over the past few years in the creation of tools capable of providing more accurate and timely reports of NCD mortality levels and trends. The objective of this report is to describe information–processing methods developed to facilitate this monitoring of mortality over time.

Methods

The study of mortality rates and/or probabilities of death requires several different types of data:

- A mapping of the causes of death into disease groups.
- For each of the years of the study:
  - Mortality records that permit the calculation of frequencies of different kinds of death at the level of individual strata.
  - Estimates of the completeness of the mortality information system that may be applied at the level of strata. By *completeness* we mean the proportion of all deaths which are registered in the population covered by the vital registration system.
  - Population figures for each of the strata.

To obtain a report, the separate sets of data have to be organized and can then be joined in a Basic Sheet from which the reports may be obtained.

Figure 1 shows the major steps employed in processing the information required for calculating the desired mortality statistics. The steps converge on the Basic Sheet. This sheet can then be used as a starting point for generation, through relatively simple programs, of most of the outputs desired.

The steps are outlined in detail in the IPython Notebook. See Additional file 1 for the notebook, see Additional file 2 for a rendering of the notebook in HTML and see Additional file 3 for the Python routines that are required to run the notebook.

Preliminary steps in the programming

We first determined the granularity, that is to say, the degree of detail of the strata to be used. We defined strata by sex/age group/administrative unit/year combinations. The administrative unit chosen was the state. We chose to analyze five–year age groups. As we had reliable population estimates starting with the year 2000, we started our series at that year.

Next, as our objective was to characterize mortality from chronic diseases, aside from categorizing the major disease groupings – communicable and maternal and child; chronic; injuries; and ill–defined causes – we also categorized major sub-groupings of chronic diseases as suggested by the WHO [14] and defined by Mathers et al. [15].

Individual Python functions, one for each disease group, were used to establish whether or not an ICD code belongs to a disease group.

Obtaining and organizing the mortality records

In Brazil, mortality records are in the public domain, available for download as compacted XBASE files. Individual files exist for each state for each year. We downloaded and incorporated the files available for the 26 states and the federal district for each of the 11 years contemplated, 2000 to 2011.

Records with implausible values for age and sex were excluded.

Obtaining counts of the population for each stratum

The strata used in this study were sex/age group/state/year. The population counts and estimates that are distributed in Brazil have the state as the administrative unit and are available for the years 2000–2030.

Distributing ill–defined causes of death among the defined causes

"Ill–defined," in the current programming, refers only to chapter XVIII of the ICD–10. Deaths reported as being due to ill defined causes were redistributed among the natural causes (which exclude the external causes) in the proportion in which they were recorded, following international recommendations [14] at the level of each individual stratum, using the following formula:

Correction factor for Ill–Defined Causes

$$\frac{Total\ Number\ of\ Deaths - Deaths\ Due\ to\ External\ Causes}{(Total\ Number\ of\ Deaths - Deaths\ due\ to\ External\ Causes) - Deaths\ due\ to\ Ill - Defined\ Causes}$$

Figure 1 Systems chart. Systems chart for the main steps in mortality surveillance information–processing.
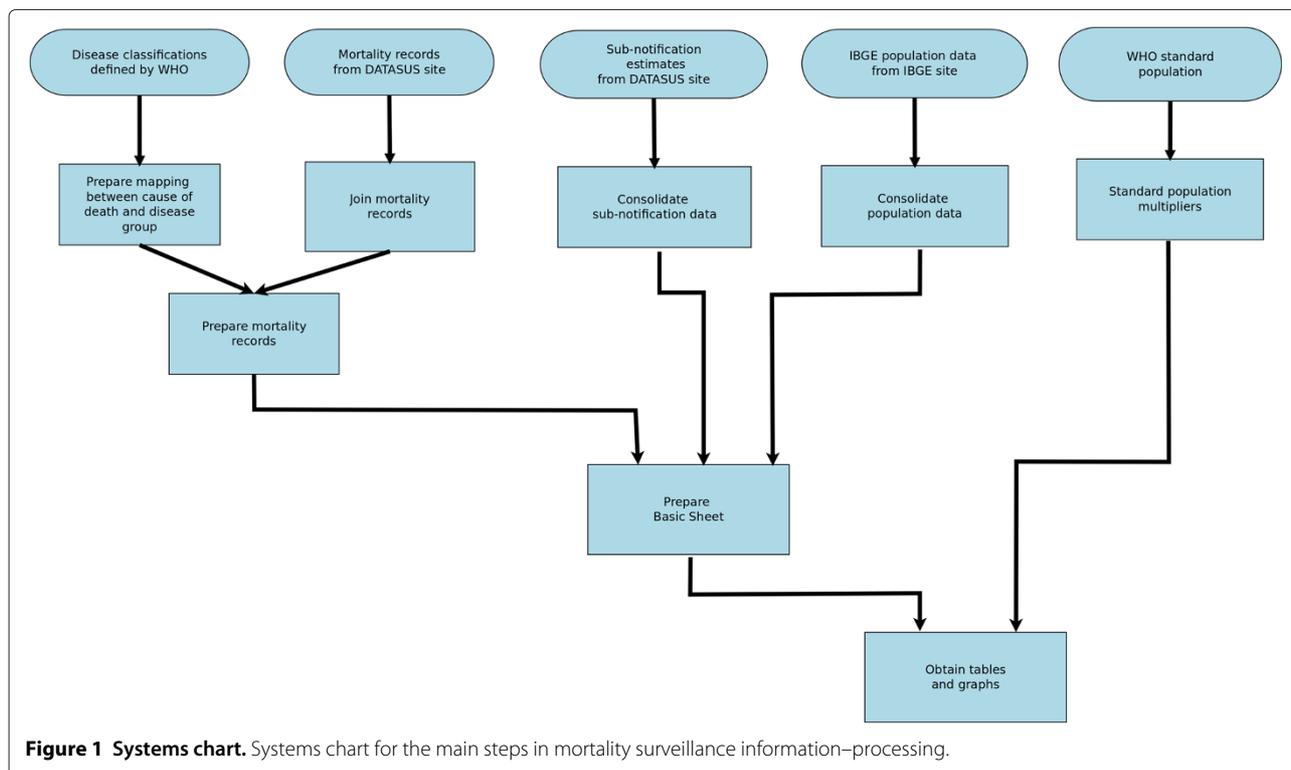
Determining correction factors for sub–notification at each stratum

In the work described in this paper, the term "sub–notification" means "the underreporting of deaths due to weaknesses in the mortality information system".

Estimates of sub–notification are provided by the Ministry of Health and are the results of data collected in the field. It must be noted that further studies should produce estimates of differential completeness. For the present, the rather simpler assumption must be made that the causes of death on the records that are missing are distributed in the same way as those that are known.

Preparing the Basic Sheet

The frequency counts, the sub–notification data, and the population counts are joined to form a "Basic Sheet". Table 1 shows values for three of the lines of this sheet (200–202).

Using the example of the first line (200):

Mortality surveillance analyses may be performed using this Basic Sheet as a starting point.

The disease groups included in this Basic Sheet vary according to the aims of the study.

Obtaining tables and graphs

The final step is to undertake the specific analyses desired, starting from the Basic Sheet. These analyses select the ICD code grouping desired (e.g., NCD sub-groupings such as cancer and diabetes). Different years, states or regions, and genders can be selected for analysis. Different statistics can be programmed, e.g., uncorrected and corrected number of deaths, crude and adjusted mortality rates, and the unconditional probability of dying from a given cause or group of causes.

When the results need to be standardized the sheet can be merged, by age group, with a table of coefficients.

$$Corrected\ Injury\ Deaths = Injury\ Deaths/sub-notification * 100$$
$$82.469369 = 70/84.88 * 100$$
$$Correction\ Factor\ Ill-Defined = \frac{Total\ Deaths - Injury\ Deaths}{(Total\ Deaths - Injury\ Deaths) - Ill-Defined\ Deaths}$$
$$1.17037037 = \frac{228 - 70}{(228 - 70) - 23}$$
$$Cancer\ Deaths\ After\ Sub-notification = \frac{Cancer\ Deaths}{sub-notification} * 100$$
$$36.522149 = \frac{31}{84.88} * 100$$
$$Cancer\ Deaths\ After\ Both\ Corrections = Cancer\ Deaths\ After\ sub-notification * Correction\ Factor\ Ill-Defined$$
$$42.744441 = 36.522149 * 1.17037037$$

**Table 1 The values from three of the lines of a Basic Sheet (200–202)**

Line	200	201	202
uf	11	11	11
Year	2005	2005	2005
Sex	1	1	1
Age group	11	12	13
Sub–notification	84.88	84.88	84.88
Ill–defined deaths original	23	20	33
Ill–defined deaths after correction for sub–notification	27.097078	23.562677	38.878417
Ill–defined deaths after both corrections	0	0	0
Total deaths original	228	221	266
Total deaths after correction for sub–notification	268.614515	260.367578	313.383600
Total deaths after both corrections	268.614515	260.367578	313.383600
Cancer deaths original	31	33	54
Cancer deaths after correction for sub–notification	36.522149	38.878417	63.619227
Cancer deaths after both corrections	42.744441	43.590952	74.385558
Injury deaths original	70	36	38
Injury deaths after correction for sub–notification	82.469369	42.412818	44.769086
Injury deaths after both corrections	82.469369	42.412818	44.769086

The values for cancer are affected by both correction for sub–notification and the redistribution of records with ill–defined causes. The values for total deaths and injuries are only modified by sub–notification.

Results

Figure 2 is an example of the kind of result that may be obtained with relative ease from the Basic Sheet. It is one of the examples in the IPython Notebook, see Additional file 1. The figure shows standardized mortality, both corrected and uncorrected, from the four main NCD groups.

To generate this output, additional programming created two new variables in the strata, the sum of deaths corrected, for ill–defined causes and sub–notification for the four NCD groups, and the sum of deaths without correction. The crude and corrected numbers of deaths and corresponding population sizes (for use as denominators
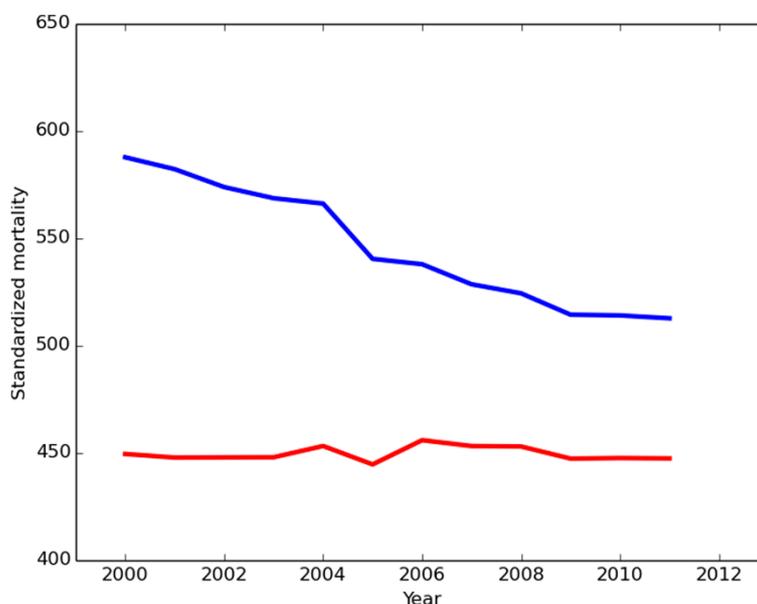
Figure 2 Standardized mortality due to NCDs. Long-term trends in standardized mortality due to NCDs for all ages, Brazil, 2000 - 2011, shown with (upper line) and without (lower line) correction for sub–notification of deaths and ill–defined causes of death.

in mortality rate calculations) are then summed across the age group/year combinations. Next, mortality rates are calculated, adjusted for the proportion of men and women, and standardized to the world population. Finally, the program produces the graph.

Figure 2 illustrates the major change seen in the trend once corrections are applied, highlighting the importance of an efficient way of applying the corrections when mortality trends are generated directly from mortality reporting rather than census–based indirect methods.

Figure 3 shows the unconditional probability of death metric proposed by the WHO as one of the main country targets for the monitoring of NCD prevention and control efforts [9]. This is also one of the examples in the IPython Notebook, see Additional file 1. In this case, for each year the number of corrected deaths in the four main NCD groupings and the respective population size are summed across gender and states for each of the age group strata corresponding to the ages 30 to 69 (eight age groups). The formula for the unconditional probability [9] is then applied directly, and the program outputs numerical data (not shown) and the graph. The graph shows the application of the formula to both corrected and uncorrected data.

Discussion

We have presented a practical method, organized in a series of steps, of combining mortality and population data along with estimates of sub–notification to produce a basic spreadsheet from which a variety of mortality surveillance reports may be obtained.

The changes seen with the corrections in Figure 2 demonstrate the importance of applying such methods in the presence of incomplete crude mortality data. Even countries with almost complete registration of deaths and a low percentage of ill–defined causes use corrections to produce data on current mortality and mortality trends. Countries with very poor mortality registration systems – due to unreported deaths taking place at home, the lack of medical attendance for many deaths, the lack of training of physicians on proper certification of causes of death, and issues related to the accurate coding of causes of death, including, most importantly, the selection of the underlying cause of death – will have to continue to use indirect methods to estimate their levels of mortality. However, as time passes, an increasing number of countries will find themselves in situations similar to that of Brazil - having mortality registration systems which, while far from perfect, permit, with the application of adequate corrections, a more reliable picture than one derived from indirect methods. The ability to apply these corrections is especially important for the analysis of mortality trends in countries with rapidly improving mortality registration systems. For these countries, the availability of information–processing systems like that described here will facilitate the rapid generation of mortality statistics.

Sub–notification and ill–defined codes are not the only factors that affect our ability to carry out accurate mortality surveillance. A recent comprehensive study [16] of
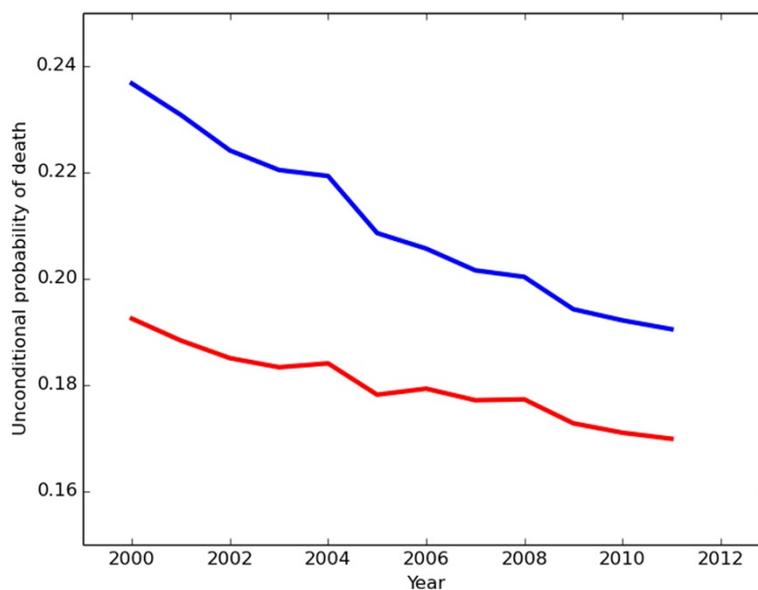
Figure 3 Unconditional probability of death due to the four main NCD groupings. Unconditional probability of death due to the four main NCD groupings (cardiovascular, cancer, chronic respiratory, and diabetes) between the ages of 30 and 70, Brazil, 2000 - 2011, shown with (upper line) and without (lower line) correction for sub–notification of deaths and ill–defined causes of death.

the factors a affecting the quality of mortality information proposes several factors that are not included in our approach. These include the quality of age and sex reporting, internal consistency, and the level of cause–specific detail. However, as pointed out by the authors of this study, completeness (sub–notification) is the most important determinant of mortality reporting quality, and the quality of cause of death reporting is another major factor.

Correction for these factors would imply a modification of individual records before the processing described here. As far as cause–specific detail is concerned, we have to admit that since our original interest was in the broad groupings of the chronic diseases we felt that corrections of detail would not greatly affect the overall result. Nevertheless, it is important to appreciate that new and more detailed approaches will be developed in the future.

It should be noted that corrected values may reduce bias but they do not reduce uncertainty. If shaded regions were to represent the uncertainty around the trend lines based on the original data, we may expect that these regions will be wider in corrected trends because of the uncertainty in the accuracy of the corrections. This is an area for further investigation.

We believe that one of the merits of our approach is that the programs are to be run with data that are available in the public domain. This makes the results reproducible by any interested party. This also means that different ideas about how to deal with ill–defined causes and sub–notification can be shared by making specific changes to the program code without the necessity to share large data files. For example, more detailed programming to include the redistribution of "garbage codes" could be added to the existing code we have created.

The separation of the complicated correction procedures from the generation of reports means that users will be able to obtain the basic spreadsheet and use other software for their analyses.

Future developments include creating more flexibility for defining disease groupings, adding code redistributing intermediate or "garbage" codes such as heart failure or septicemia, and the development of basic sheets for other geographic groupings. The definition of the disease groups has to be fixed before producing the basic spreadsheet. If the number of disease groups gets very large the spreadsheet becomes unwieldy. A clear improvement will be to permit the user to choose disease groups for his spreadsheet. We are also considering developing spreadsheets for other strata that are defined with other administrative criteria such as state capitals, micro–regions, health authorities and urban agglomerations. Other disease groups and combinations of age groups may also be implemented. For example, studies could be restricted to other sets of age groups such as 60 years and above.

Conclusion

The strategy of information–processing presented here separates the handling of the primary concerns in mortality surveillance. On the one hand, the complicated process of combining the data and implementing the correction methodologies is dealt with prior to creating the Basic Sheet. On the other, the Basic Sheet, once created, permits undertaking simpler tasks for report generation, which should be the end users' main concern.

Additional files

Additional file 1: IPython Notebook. An **ipynb** file which the user can run on his own computer.

Additional file 2: IPython Notebook rendered as html. The same notebook as Additional file 1 rendered as an **html** file so that the user can study it with an inernet browser.

Additional file 3: Ancillary programs. A **zip** file containing a number of ancillary functions and definitions required for when the notebook is in use.

References
1. Bloom DE, Cafiero ET, Jané–Llopis E, Abrahams–Gessel S, Bloom LR, Fathima S, Feigl AB, Gaziano T, Mowafi M, Pandya A, Prettner K, Rosenberg L, Seligman B, Stein AZ, Weinstein C: **The Global Economic Burden of Non-communicable Diseases.** 2011. [http://www.weforum.org/EconomicsOfNCD]
2. **WHO | 2008–2013 Action plan for the global strategy for the prevention and control of noncommunicable diseases.** [http://www.who.int/nmh/publications/9789241597418/en/]
3. **WHO | Global status report on noncommunicable diseases 2010.** [http://www.who.int/nmh/publications/ncd_report2010/en/]
4. **WHO | Preventing chronic diseases: a vital investment.** [http://www.who.int/chp/chronic_disease_report/en/]
5. Beaglehole R, Bonita R, Alleyne G, Horton R, Li L, Lincoln P, Mbanya JC, McKee M, Moodie R, Nishtar S, Piot P, Reddy KS, Stuckler D: **UN high-level meeting on non-communicable diseases: addressing four questions.** *Lancet* 2011, **378**(9789):449–455. [http://www.ncbi.nlm.nih.gov/pubmed/21665266]. [PMID: 21665266]
6. Alleyne G, Binagwaho A, Haines A, Jahan S, Nugent R, Rojhani A, Stuckler D, Matswama M: **Embedding non-communicable diseases in the post-2015 development agenda.** *Lancet* 2013, **381**(9866):566–574.
7. Mendis S: **The policy agenda for prevention and control of non-communicable diseases.** *Br Med Bull* 2010, **96:**23–43.
8. Maher D, Harries AD, Zachariah R, Enarson D: **A global framework for action to improve the primary care response to chronic**

   non-communicable diseases: a solution to a neglected problem.
   BMC Public Health 2009, **9:**355.
9. **Mortality from NCDs Target - World Health Organization.** *Available on–line* 2012. [www.who.int/nmh/events/2012/note_20120426.pdf]
10. RIPSA: **Indicadores e Dados Básicos – Brasil – 2011.** March. [http://tabnet.datasus.gov.br/cgi/idb2011/matriz.htm#mort]
11. Jorge MHPdM, Laurenti R, Gotlieb SLD: **[Quality analysis of Brazilian vital statistics: the experience of implementing the SIM and SINASC systems].** *Ciência & Saúde coletiva* 2007, **12**(3):643–654. [http://www.ncbi.nlm.nih.gov/pubmed/17680121]. [PMID: 17680121].
12. Brasil. Ministério da Saúde: *A experiência brasileira em sistemas de informação em saúde; The brazilian experience in health information systems. Volume 1.* Brasil: Textos Básicos de Saúde, Produção e disseminação de informações sobre saúde no; 2009. [http://www.saude.gov.br/bvs]
13. Brasil. Ministério da Saúde: *A experiência brasileira em sistemas de informação em saúde; The brazilian experience in health information systems. Volume 2.* Brasil: Textos Básicos de Saúde, Produção e disseminação de informações sobre saúde no; 2009. [http://www.saude.gov.br/bvs]
14. **WHO | Global Burden of Disease (GBD).** [http://www.who.int/healthinfo/global_burden_disease/en/index.html]
15. Mathers CM, Bernard C, Iburg KM, Inoue M, Ma–Fat D, Shibuya K, Stein C, Tomijima N, Xu H: **Global burden of disease in 2002: data sources, methods and results.** Tech. rep., World Health Organization 2003 (revised 2004). [www.who.int/healthinfo/paper54.pdf]
16. Phillips DE, Lozano R, Naghavi M, Atkinson C, Gonzalez-Medina D, Mikkelsen L, Murray CJ, Lopez AD: **A composite metric for assessing data on mortality and causes of death: the vital statistics performance index.** *Popul Health Metr* 2014, **12:**14.