Report of the Workshop on Hospital Mortality Data Analysis

Estimating Causes of Death from Biased Datasets

Vevey, Switzerland

15-16 May 2008

World Health Organization, Geneva, Switzerland
Institute for Health Metrics and Evaluation, Seattle, WA, USA
Health Metrics Network, Geneva, Switzerland
Table of Contents

Proceedings of the Workshop ........................................................................................................ 3
Appendix 1. Workshop Agenda ...................................................................................................... 9
Appendix 2. Introductory Presentation ............................................................................................. 11
Appendix 3. Step-by-step Instructions .............................................................................................. 17
Appendix 4. Group Presentations .................................................................................................... 30
Appendix 5. List of Participants ...................................................................................................... 37
Appendix 6. Organization Overviews .............................................................................................. 46
Proceedings of the Workshop

Information on the causes of death that affect a population is a critical input to public health planning. Despite the importance of population-based data on causes of death, in many countries their availability is limited. Therefore, new and innovative ways to determine population cause-specific mortality fractions (CSMFs) are needed. A method has been developed that generates relatively accurate CSMFs using cause-of-death data from in-hospital deaths, which are more likely to be complete and of high quality than civil registration death records.

World Health Organization (WHO), in collaboration with the Institute for Health Metrics and Evaluation (IHME) at the University of Washington, Seattle and the Health Metrics Network (HMN), jointly organized this workshop to apply the method to countries' hospital data.

The workshop's objectives were four-fold:
1. to disseminate the methodology and test it using country datasets;
2. to draw inferences about population cause-of-death patterns based on the results;
3. to assess the quality of both the hospital data and the cause-of-death data from the routine civil registration system; and
4. to identify the gaps in mortality data and areas of future work to improve the cause-of-death information at country level.

To fulfil the objectives of the workshop, participants brought hospital death records and/or civil registration death records with the place of death recorded (i.e., hospital or not). The participants analysed their own datasets with technical support from the meeting organizers.

Thirty countries participated in the workshop. They were selected on the basis of regional distribution, data availability and quality of hospital records, and willingness to participate and share their datasets. Participants, in collaboration with WHO and IHME, did extensive data cleaning and formatting in preparation for the workshop, which was essential for participants to carry out analysis by the end of the workshop (see Appendix 1 for the workshop agenda). A summary of the proceedings follows.

Day 1: Introduction and application of the Hospital Mortality Method

Introduction to a new method for hospital data analysis

Christopher Murray (IHME) presented the new hospital mortality method (Appendix 2). He first introduced three problems that exist in civil registration systems that record cause-of-death information in developing countries: 1) failure to record all deaths (i.e., low completeness), 2) insufficient information at the time of death to correctly assign a cause of death, in particular for deaths outside health facilities, and 3) frequent use of "garbage" (ill-defined) codes despite sufficient information to correctly assign a cause of death. The impact of these three factors is that deaths for which a cause is correctly assigned represent a biased sample of all deaths. He hypothesized that cause-of-death assignment is more complete and of higher quality in hospitals, and though that data alone is biased due to selection bias, they can be used to estimate causes of death for deaths occurring in other locations. To do so, an estimate of the proportion of deaths that occur in a hospital by age, sex, and cause is needed.

Because the natural history of a disease strongly affects the proportion of deaths occurring in hospital for an age-sex-cause group, it may be possible to transfer these values from a place
where cause-of-death certification and coding are of high quality to a place where data do not exist to calculate these proportions. Several country studies were carried out to validate the method and test whether the proportions could be transferred from other countries or regions. It was suggested that the proportions would be transferable at least when using data from four countries used in the analysis (US, Mexico, South Africa and Iran). However, this requires further validation given the large variations in the cause-of-death patterns across regions. A logistic regression model is being tested that predicts the probability of in-hospital death at the individual level. Because this model uses information on a country’s level of development as well as age, sex, and cause in its predictions, it produces better estimates than when proportions are transferred from another country. CSMF estimates obtained using the model may also be useful in identifying coding quality issues in those countries with vital registration. He concluded that the error in estimating CSMFs for the hospital mortality method are small compared to the error when using verbal autopsy methods.

**Step-by-step approach to hospital data analysis presented by IHME staff**

Jeanette Kurian, with support from Dennis Feehan and Rafael Lozano, reviewed and demonstrated the hospital mortality method. They described the variables needed to apply the method and reviewed the mathematics in detail, and then applied the method to Mexico’s data using Stata, using the step-by-step methodology prepared for the workshop and presented in Appendix 3.

**Application of the method to country data**

The participants were divided into six groups based on country as follows:

1. African group: Ethiopia, Ghana, Kenya, Mozambique, Tanzania
2. American group 1: Argentina, Brazil, Colombia, Costa Rica, Ecuador, El Salvador, Nicaragua
3. American group 2: Barbados, Guyana, Suriname
4. Asian group: Malaysia, Mongolia, Myanmar, Philippines, Sri Lanka, Thailand
5. Eastern Mediterranean group: Algeria, Egypt, Oman, Saudi Arabia
6. European group: Belarus, Georgia, Kazakhstan, Poland, Turkey

During the afternoon, participants followed the instructions in Appendix 3 to apply the method to their data. Each group was assisted by staff members from WHO and IHME to help them in applying the method to their data. By the end of the first day all participants had estimated the cause-specific mortality fractions using their own hospital or/civil registration data or with the probability of dying in hospital per age-sex-cause group from another country (Iran, Mexico, South Africa or US).

**Day 2: Sensitivity analysis, presentation of results and discussion of next steps**

**Sensitivity analyses**

During the morning, the participants repeated the above exercise by using proportions of deaths occurring in hospital from different countries in order to test the sensitivity of their results. For countries where individual-level data were available, they also compared their results to the output of a logistic regression that predicted proportions for their respective countries from pooled datasets of 4 countries (US, Mexico, South Africa, and Iran) after adjusting for the levels of GDP. A further refinement of this model requires the individual-level data from more countries.

The participants were encouraged to critically review the estimates of the cause-specific mortality fractions thus obtained in the light of the epidemiological situation of their country.
In addition, they explored the sources of "garbage" codes in their data by looking at the detailed ICD codes used in, for example, "other cardiovascular disease".

**Presentation of results**

The participants discussed their results with other participants in their group. Each group prepared a brief presentation of their results, conclusions and key issues for future work. A summary of region-specific issues follow, and issues that were common across groups as well as proposed next steps are described in the discussion section.

**African group**: No member of this group had a continuously operating civil registration system, and generally hospital data had been collected for very few years. Hospital deaths were frequently aggregated into age groups at the level of the hospital. For this group, the cause list used was inadequate as malaria was not analysed separately from other infectious diseases.

**American group 1**: All participants in this group had access to three to six years of hospital and civil registration death records. In most cases, civil registration data were used as the underlying cause of death was not recorded in the hospital data. Within this group, countries generally either had low use of garbage codes and moderate coverage, or high coverage and higher use of garbage codes.

**American group 2**: This group was made up of small countries, and the total number of observations in the datasets were small despite using between four and six years of data. The need for large datasets hampered application of the method for these participants.

**Asian group**: This group was quite heterogeneous, with some countries using hospital data only and others using civil registration data. In some of the countries (Thailand, Malaysia, Sri Lanka and Myanmar) ill-defined conditions was the leading cause of death in their primary analysis.

**Eastern Mediterranean group**: In general, the data used by these participants had very high use of garbage codes. Therefore, results were often difficult to interpret.

**European group**: These countries generally had high civil registration coverage and had access to both civil registration data and hospital data, but had gaps in their data collection (in terms of variables in their datasets) or weaknesses in terms of cause-of-death coding.

**Discussion and conclusions**

1. **Garbage codes**

   For some countries, hospital data had a very high use of garbage coding, which limited utility of the method in those cases. This was also contrary to the expectations of the workshop organizers, who expected substantially lower use of garbage codes in hospital deaths. Because many participants applied proportions from countries where few hospital deaths are ill-defined and most ill-defined deaths occurred out of hospitals (for example, in Mexico and South Africa, around 10% of ill-defined deaths occurred in hospitals, vs. 40% in the U.S.), the method predicted a large proportion of ill-defined deaths in the general population based on a moderate proportion of ill-defined deaths in hospitals. When use of garbage codes was high for in-hospital deaths, results were sensitive to the sources for proportions (i.e., Mexico and South Africa vs. the U.S.). Therefore much of the discussion focused on use of garbage codes.
There are two general ways to deal with garbage coding: in the short term, garbage codes may be reassigned based on researchers’ understanding of when specific garbage codes are used. This approach requires an algorithm to redistribute garbage codes to an appropriate set of codes, which is being developed by IHME in collaboration with WHO. Once it is ready, participants can redistribute ill-defined deaths prior to applying the hospital mortality method. Because the use of garbage codes can vary by country, empirical work (e.g., chart reviews) is urgently needed to validate the redistribution algorithm for different settings and modify it where appropriate.

In the long term, use of garbage codes must be addressed by the countries during data collection by either suppressing their use or improving the mechanism for selecting underlying cause of death on the death certificate. Rather than discussing disease classification systems (such as ICD), the discussion explored the process of reaching a particular underlying cause of death given a set of information available to those who certify the death. As shown below, who assigns the cause of death depends on where the death occurred:

<table>
<thead>
<tr>
<th>Location of death</th>
<th>Hospital</th>
<th>Other health facility</th>
<th>Home</th>
<th>Outside home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who certifies the death?</td>
<td>Medical doctor</td>
<td>Nurse, community health worker</td>
<td>Varies by country</td>
<td>Police / forensic institute</td>
</tr>
<tr>
<td>Other sources of information</td>
<td>Patient records (signs and symptoms with clinical test results, diagnostic imaging, etc.)</td>
<td>Patient records (signs and symptoms with some clinical test results)</td>
<td>Verbal autopsy (signs and symptoms)</td>
<td></td>
</tr>
</tbody>
</table>

When a death occurs in a hospital it is the responsibility of a medical doctor to certify the cause of death. In other health facilities it is generally the nurse or community health worker who certifies the cause of death. Who certifies deaths occurring at home varies widely from country to country; it also varies for deaths occurring outside of homes though typically either the police or forensic institute certifies the cause of death. In principle, the process of reaching a specific diagnosis from a set of information (signs and symptoms, lab tests, diagnostic imaging, etc.) should be the same regardless of where the death occurs - only the amount of available information differs substantially.

However, physician practice plays a large role in how cause of death is assigned in hospital deaths. That is, given identical clinical history, physicians in different settings will consistently assign different causes of death. Determining the effect of physician culture on how a specific diagnosis is assigned based on sign and symptoms is a key step to understanding the garbage code problem. It was suggested that one way to address garbage coding is to return to the signs, symptoms, and laboratory test results as recorded at the hospital. If test results could be probabilistically associated with causes of death, a distribution could be obtained with likely causes of death, from which physicians can choose the underlying cause of death. This would minimize the variations due to subjective judgement of physicians.

Some participants argued that it would not be possible to improve on physician judgement, and that it would be better to educate physicians about the importance of filling out the death
certificate correctly. In addition it would be difficult to access patient records, especially in private hospitals. In some cases, a patient is transferred to a long-term care facility to another hospital just prior to death; the medical records at the long-term care facility would be more relevant than those at the hospital that reported the death.

It was also noted that the majority of deaths coded to garbage codes are coded to a few specific ICD codes, which may vary by country. One suggestion was to distribute a list of the ten most commonly used garbage diagnoses for each country to physicians, and ask physicians to avoid using those diagnoses unless they were strictly indicated.

For deaths occurring outside of hospitals, verbal autopsy can be used to obtain cause of death information if it is not recorded, or validate recorded or predicted cause of death information.

2. Country-specific estimate of proportion of deaths in hospitals
Developing proportions of deaths that occur in hospitals for each country was discussed. For a country-specific estimates to be calculated, complete vital registration with high-quality cause-of-death coding and location of death (hospital or elsewhere) is needed for at least one geographic area within the country. For age, sex, and cause-specific proportions to be stable, the number of observed deaths must be quite large. Few participants were able to calculate proportions with their own data. In some cases, country-specific proportions could be calculated with improved data collection, considering a longer time-series. An alternative was using the logistic regression model that incorporates country-specific characteristics, which could be improved by adding more countries to the regression.

The country characteristics on which proportions borrowed from other countries should be matched, or which should be included in a final regression model were discussed. Factors to consider include the epidemiological characteristics of the country, the level of development of the country, and the overall proportion of deaths that occur in-hospital. One issue to consider when using this method on hospital data from a government hospital system (instead of from vital registration) is how deaths in government hospitals may differ from deaths in private hospitals, and whether that affects transferability of proportions.

3. Data sources and administrative influences
The group discussed the data sources that are available to complete a hospital mortality analysis. In many countries, both hospital death records and civil registrations death records were available. Some considerations are:
- Hospital death records may contain admission, discharge or underlying cause of death diagnoses. Underlying cause of death should be used for this type of analysis. Admission diagnoses are not useful for this type of analysis, and the utility of discharge diagnoses requires further investigation. For injuries, hospitals often do not record the underlying cause of death based on the external cause of the injury (V-Y codes), instead they record the consequences of the external causes (S-T codes).
- Discharge diagnoses can be influenced by insurance systems. Diagnoses that are not covered by insurance schemes were less likely to be used in some cases.
- In countries where care is administered in private hospitals, hospital death records may not be easily available.
- In some countries, the person who assigns an ICD cause based on a death certificate is trained to do so for one system (e.g., the civil registration system) but not in the other (e.g., hospital death records).
4. **Cause list**

The group discussed how the short-list of causes used during the workshop could be tailored for different countries' analyses. One should consider isolating diseases with high mortality when constructing the short-list. A reasonable next step would be to create regional cause lists, which could be derived from the proposed ICD short list developed for verbal autopsy tools. For example, the African participants would list malaria as a separate cause, which was not reasonable for many other countries.

5. **Sample size**

A number of countries had access to only numbers of hospital deaths (under 50 000 death records), resulting in unstable estimates using the hospital mortality method. The African group, Asian group, and American group 2 each suggested that regional pooling of data could be used to increase sample sizes.

6. **Stata software**

Several participants raised concerns about the availability of Stata software to allow them to continue to use the method (and analyse their data in other ways). Although a few participants already have Stata, the vast majority do not use it. A possible solution would be to develop an application which could be accessed from a web site to allow users to apply the method, but that would not allow much flexibility for the users. A more practical strategy needs to be considered in order to disseminate this method more widely.

**Proposed next steps**

WHO, IHME, HMN and many country participants expressed interest in continued collaboration. Specifically, six next steps were identified by the group:

1. Participants from countries with high use of garbage codes were charged with implementing systems to reduce the use of garbage codes.
2. New and innovative ways to understand and reduce use of ill-defined cause-of-death codes are needed. An important first step is to understand the process of assigning a specific cause of death given a set of information.
3. IHME and WHO will continue to develop algorithms to redistribute deaths assigned to garbage codes, and will share these methods with participants upon completion.
4. For those countries for which the hospital mortality method can currently be used, IHME, the WHO and the country participant should work together to apply and evaluate the method.
5. Many participants wished to learn Stata to facilitate the analysis of their data.
6. HMN will continue to support country initiatives to improve the quality of data collection as well as efforts to apply methods for better estimates of causes of death.

**Acknowledgments**

We gratefully acknowledge funding for this workshop provided by the Health Metrics Network, the Institute for Health Metrics and Evaluation, the Japanese Ministry of Health, Labour and Welfare, and the World Health Organization.
## Appendix 1. Workshop Agenda

### Thursday, 15 May 2008

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenter(s)</th>
</tr>
</thead>
</table>
| 9:00 – 9:20 | Opening                                      | Ties Boerma (WHO)  
Sally Stansfield (HMN)  
Chris Murray (IHME)   |
| 9:20 – 9:30 | Workshop overview                            | Kenji Shibuya (WHO)                   |
| 9:30 – 10:30 | Introduction to a new method for hospital data analysis  
- Theory, method, application and validation  
- Current status and future directions  
- Discussion, Q & A | Chris Murray (IHME) |
| 10:30 – 11:00 | COFFEE BREAK                                |                                       |
| 11:00 – 12:30 | Step-by-step approach to hospital data analysis  
- Overview: analytical approaches, data sources, data quality measures, analysis plan  
- Examples  
- Q & A | Dennis Feehan (IHME)  
Rafael Lozano (IHME)  
Jeanette Kurian (IHME) |
| 12:30 – 14:00 | LUNCH                                       |                                       |
| 14:00 – 15:30 | Country data analysis I  
- Prepare data and begin analysis | WHO and IHME staff                   |
| 15:30 – 16:00 | COFFEE BREAK                                |                                       |
| 16:00 – 18:00 | Country data analysis II  
- Continuation of analysis  
- Calculate one set of results for your country | WHO and IHME staff |
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 – 9:30</td>
<td><strong>Overview</strong></td>
<td>Kenji Shibuya (WHO) Jeanette Kurian (IHME)</td>
</tr>
<tr>
<td></td>
<td><strong>- Summary of Day 1</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>- Overview of Day 2</strong></td>
<td></td>
</tr>
<tr>
<td>9:30 – 10:30</td>
<td><strong>Country data analysis III</strong></td>
<td>WHO and IHME staff</td>
</tr>
<tr>
<td></td>
<td><strong>- Synthesize and graph results</strong></td>
<td></td>
</tr>
<tr>
<td>10:30 – 11:00</td>
<td><strong>COFFEE BREAK</strong></td>
<td></td>
</tr>
<tr>
<td>11:00 – 12:30</td>
<td><strong>Country data analysis IV</strong></td>
<td>WHO and IHME staff</td>
</tr>
<tr>
<td></td>
<td><strong>- Finalize and summarize preliminary findings</strong></td>
<td></td>
</tr>
<tr>
<td>12:30 – 14:00</td>
<td><strong>LUNCH</strong></td>
<td></td>
</tr>
<tr>
<td>14:00 – 15:45</td>
<td><strong>Country data analysis V</strong></td>
<td>WHO and IHME staff</td>
</tr>
<tr>
<td></td>
<td><strong>- Share the results and discuss with group members</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>- Identify key issues and gaps in data</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>- Prepare for group presentations (coffee and refreshments available)</strong></td>
<td></td>
</tr>
<tr>
<td>15:45 – 17:15</td>
<td><strong>Group presentation and future directions</strong></td>
<td>Chris Murray (IHME) Kenji Shibuya (WHO)</td>
</tr>
<tr>
<td></td>
<td><strong>- Preliminary results</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>- Discussion and feedback from participants</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>- Next steps (follow-up analysis, a global database of hospital records, and collaborative studies)</strong></td>
<td></td>
</tr>
<tr>
<td>17:15 - 17:30</td>
<td><strong>Closing</strong></td>
<td>Nosa Orobaton (HMN) Chris Murray (IHME) Carla AbouZahr (WHO)</td>
</tr>
</tbody>
</table>
Appendix 2. Introductory Presentation

### Outline
- Introduction
- Methods
- Validation
- Results
- Discussion

### Population Causes of Death: Key Health Information
- Reliable information on leading causes of death is a key input for health policy.
- Causes of death should guide both investment decisions as well as help track progress of priority health programs.
- MDG indicators such as maternal mortality, HIV, TB and malaria mortality are illustrations of the importance of cause of death data.

### Three Common Problems
1. Many deaths in low and middle-income countries are not recorded in vital registration systems.
2. Some deaths do not have sufficient diagnostic information available at the time of completing a death certificate to ascertain true cause.
3. Death certification leads to the coding of the underlying cause of death to a ‘garbage code’.

### Low Coverage of Vital Registration Systems
In many countries, vital registration systems capture deaths in urban communities or for richer households. The cause composition of deaths in incomplete systems is likely to be biased towards the causes of death that affect the better off.
- It is difficult to accurately determine how complete is a vital registration system.
- Range of demographic techniques including Synthetic Extinct Generations, General Growth Balance and others have been developed to assess completeness.

### Availability of vital registration data

### Information Available for Cause Certification
- Accurate completion of an death certificate following the principles of the ICD depends on the diagnostic information available to the individual completing the death certificate.
- Individuals who have not had contact with health services prior to death will have much less information available for certification.
- Extent of diagnostic testing, imaging and clinical history will all influence quality of certification.
- Deaths outside of hospital likely to be less accurate.
Proper Assignment of Underlying Cause

Quality of cause of death data depends not only on the information available to the certifier but on the training and skill of the certifier. Often deaths are assigned underlying causes that are ‘garbage codes’. For example, heart failure, general atherosclerosis, ill-defined etc.

Quality of Cause of Death Coding

Potential to Use Deaths in Hospital

In many countries with incomplete or low-quality vital registration data, deaths in hospital may provide a useful source of information. Deaths in hospital are not a representative sample of deaths in the community. Because of the natural history of each cause of death and the propensity of different individuals to seek healthcare, the causes in hospital will be different than in the community.

Mapping From Deaths in Hospital to the Community

Deaths in hospital in general have better information available for certification than deaths outside of hospital. If we can understand the probability of a death in the community occurring in the hospital as a function of cause, age, sex, and other variables, then we can map from deaths in hospital to deaths in the population. Using existing data on deaths in hospital is also low-cost as many countries are already collecting this information.

Outline

Introduction
Methods
Validation
Results
Discussion

Basis of the Method

• We use observed proportions of in-hospital death by age-sex-cause group to correct observed hospital CSMFs, yielding robust estimates of population CSMFs.
• To validate our method, we used vital registration data from Mexico for the years 1998-2005, from South Africa for 2002-2005 and from the United States for 1999-2002.
• We also explored the extent to which we can apply probabilities of in-hospital death from one population to estimate population CSMFs in another.

Definitions

The population cause-specific mortality fraction is simply the number of deaths from cause \( j \) divided by all deaths:

\[
\text{CSMF}_j = \frac{\sum \sum D_{aj}}{\sum \sum \sum D_{as}}
\]

All deaths due to cause \( j \)

All deaths

Definitions

Proportion of deaths in group \( a \), sex \( s \) from cause \( j \) that occur in hospital

\[
H_{adj} = \frac{D_{asj}}{P_{asj}}
\]

\( H_{adj} \) = number of deaths in hospital for age-group \( a \), sex \( s \) from cause \( j \)

\( D_{asj} \) = number of population deaths in age-group \( a \), sex \( s \) from cause \( j \)

\( P_{asj} \) = proportion of deaths in age-group \( a \), sex \( s \) from cause \( j \) that occur in hospital.
Definitions

We can estimate deaths from cause \( j \) in an age-sex group by dividing hospital deaths by the proportion of deaths that are expected to occur in hospital:

\[
CSMF = \frac{\sum_{a} \sum_{s} H_{asj} P_{asj}}{\sum_{a} \sum_{s} H_{asj}}
\]

If we are able to estimate the values of \( P_{asj} \) for a population, then in-hospital deaths can be easily corrected to yield population CSMFs.

Required Information

• Deaths in hospital by age and sex accurately assigned an underlying cause of death according to the International Classification of Diseases (ICD)
• An estimate of the proportion of in-hospital death by age, sex and cause group, \( P_{asj} \), obtained from a subset of that population or a similar population in another country.

a) Deaths in Hospital

• Nearly all middle-income and many low-income countries record in hospital deaths by cause
• In a number of them the cause attribution may be sufficiently high quality to obtain more detailed data that would allow tabulation by age, sex and cause.

b) The Challenge for Operationalizing this Method: Probabilities of In-Hospital Death

• Method accuracy depends on the accuracy of \( P_{asj} \) estimated for a subset of the population or estimated in some other community.
• This accuracy in turn depends on how stable \( P_{asj} \) are across communities with different socio-economic levels and over time.
• Obtaining a reasonable estimate of \( P_{asj} \) depends on complete or near complete vital registration (VR) data that accurately assign the underlying cause of death and whether the death occurred in hospital.

Outline

Introduction
Methods
Validation
Results
Discussion

The Mexico Study

• We first validated this approach using individual death records from Mexico 1998-2005
• Vital registration is estimated to be greater than 90% complete in Mexico and closer to 95% complete for adult.
• Mexico collects information on the location of death (in-hospital or not), so we can both predict population CSMFs and compare them to the observed CSMFs using vital registration data.
• Mexico’s states also represent a tremendous range of socioeconomic and health conditions.

Causes of Death

We based our analysis on 45 cause groups that are mutually exclusive and collectively exhaustive. To determine these, we started with the Global Burden of Disease cause list adjusted to the U.S. cause-of-death profile, which includes 109 causes.

Method Validity

Our primary measure of method validity is the average relative error (ARE) for the 45 CSMFs. This metric can be calculated for any population for which CSMFs are being predicted. Formally, it is defined as:

\[
ARE = \frac{\sum_{j} \left( \frac{CSMF_{j} - CSMF_{j}}{CSMF_{j}} \right) - 1}{45}
\]

This metric directly measures the deviation between estimated and true CSMFs.

We tested this approach in two ways:

1) Demonstrated that the method can provide good estimates of population CSMFs using a range of hypothetical coverage of national vital registration data.
• The values of \( P_{asj} \) for a country can be estimated using the available VR data in a country.
• We simulated partial VR coverage in Mexico by using \( P_{asj} \) estimates derived from the more socioeconomically advanced states. We ordered states on the basis of the literacy rate from the 2000 Census.
• We assumed that most VR data come from the more developed parts of the country, especially in nations with low levels of VR coverage. For each level of partial VR coverage, we computed new \( P_{asj} \) estimates and used this set of probabilities to correct Mexico’s hospital CSMFs to estimate population CSMFs.

2) Explored whether \( P_{asj} \) values measured in one population can be used to estimate population CSMFs using in-hospital deaths in another community.
2) Explored whether $P_{in}$ values measured in one population can be used to estimate population CSMFs using in-hospital deaths in another community.

- We used VR data for 1998-2005 for the Distrito Federal and the Estado de Mexico, which together form the main urban and periurban center in Mexico, to calculate $P_{in}$ values.
- We would expect that an urban area such as these two together would have higher access to hospital services than a poor rural area.
- We then applied these fractions of in-hospital deaths to the population can be used to estimate population CSMFs using in-hospital deaths in another community.

Average Percent Error for CSMF

We used VR data for 1998-2005 for the Distrito Federal and the Estado de Mexico, which together form the main urban and periurban center in Mexico, to calculate $P_{in}$ values. We would expect that an urban area such as these two together would have higher access to hospital services than a poor rural area. We then applied these fractions of in-hospital deaths to the population can be used to estimate population CSMFs using in-hospital deaths in another community.

Average Relative Error in Population CSMFs when Based on Hospital Deaths, by State versus the Proportion of All Deaths Occurring in-Hospital, Mexico 1998-2005

- These sub-groups serve to demonstrate how socio-economic status affects the overall probability of dying in hospital:
  - For HIV/AIDS, diabetes mellitus and cerebrovascular disease, the proportion dying in hospital at any age-group is lower in municipalities with lower socio-economic status as assessed by literacy rates.
  - For road traffic accidents, however, there is no marked difference by level of development in the proportion of in-hospital deaths, as might be expected.

- These four causes illustrate that the proportion of in-hospital deaths is a distinct function of age, cause, and level of community development:
  - This diverse pattern confirms that CSMFs based solely on in-hospital deaths are likely to be inaccurate.

- The previous figure shows average relative error for hospital CSMFs as a function of the percent of deaths in-hospital for each Mexican state.
  - As expected, the average percent error steadily rises as the proportion of deaths in-hospital falls.
  - In other words, in states with a smaller proportion of in-hospital deaths, the effects of selection bias on the hospital CSMFs are greatest.
Results

• The previous figure systematically explores the relationship between the amount of VR data used to calculate the PASJ in Mexico (from 9% to 100%) and the average relative error across 45 causes of death at the national level.
• Even if VR in Mexico covered only a small fraction of the country’s most developed states, our methods suggest that we would be able to measure CSMFs quite accurately if data on causes of death in hospital were available.

Average Relative Error

- In the state with the lowest fraction of deaths in hospital, Oaxaca, the ARE is 30% using our correction method.
- The ARE across the 45 CSMFs is even lower for the states of Guerrero and Chiapas.
- While these levels of error are much higher than we obtain at the national level, the results still demonstrate the possibility of estimating plausible CSMFs for a large set of causes even in settings where the PASJ cannot be measured directly, but must be borrowed from another population.

Other Applications

Where deaths in hospital are recorded and assigned causes according to the ICD, but vital registration data may not be available, it may be worthwhile to use PASJ values for a neighboring country.

For example:
- India for Pakistan and Bangladesh
- South Africa, Zimbabwe, or Mozambique for other Southern African countries

In Development: Logistic Regression Model

\[ \logit(hospital) = \beta_0 + \beta_1(age) + \beta_2(sex) + \beta_3(gdp) + \beta_4(cause) + \beta_5(prop_hosp) + \beta_6(age*cause) + \beta_7(prop_hosp*cause) \]

• The logit result is a predicted probability that the individual should have died in hospital, given his or her covariates
• Using the relationship \( H = O^{\text{PASJ}} \) at the individual level with index \( i \), each hospital death represents \( \frac{1}{P_{ij}} \) community deaths, and the CSMF is defined as:

\[
\text{CSMF}_{ij} = \sum_{k} \frac{\sum_{n} \text{estimated deaths due to cause } j}{\sum_{n} \text{total deaths due to cause } j} \]
Discussion

• When high quality ICD-coded data on deaths in hospital and high quality ICD-coded data from vital registration from a small subset of the population or a similar population are available, population CSMFs can be estimated with an acceptable level of error.
• The results are robust even when using less than 10% of VR data to estimate the proportion of in-hospital death for each age, sex, and cause group.

Implications for Assessing Quality of VR

• South Africa results illustrate that for some causes especially when assigned outside of hospital, the model suggests the number of deaths is too high or low.
• This could be a true pattern or possibly an indicator of low quality of cause certification outside of hospital especially for conditions such as cancers or other diseases requiring sophisticated diagnostics.
• The comparison of hospital and VR cause of death patterns compared to benchmarks may be a useful tool for identify potential quality problems.

Future Work: This Workshop

• Workshop is an opportunity to both further validate the method with full VR data as well as apply it in countries with substantial or no VR.
• Results obtained will be highly informative in both cases: cause estimates for areas previously without, and validation of method and assessment of data sources for areas with national VR.
• Presentations of results will also act as a forum for discussion of the challenges in collecting reliable cause-of-death data:
  - Completeness of VR data
  - Quality of coding (measaling, ill-defined, missing data, underlying cause)
  - Hospital data considerations:
    - Data originated in hospital, discharge, unverified cause of death
    - Sources of data sources (hospital databases or hospital death records in VR)
  - Non-ICD vs. ICD, large hospitals vs.
Appendix 3. Step-by-step Instructions

HOSPITAL METHOD WORKSHOP

ESTIMATING CAUSE-SPECIFIC MORTALITY FRACTIONS IN STATA: UNIT RECORD HOSPITAL DATA

METHOD CONCEPTS FOR REFERENCE

Our quantities of interest are:

\[ H_{asj} = \# \text{ in-hospital deaths in age group } a, \text{ sex } s, \text{ from cause } j \]
\[ D_{asj} = \# \text{ population deaths in age group } a, \text{ sex } s, \text{ from cause } j \]
\[ P_{asj} = \text{ proportion of deaths in age group } a, \text{ sex } s, \text{ from cause } j \text{ that occur in hospital} \]

These quantities have the relationship

\[ H_{asj} = D_{asj} \cdot P_{asj} \]

That is, for a given age group, sex, and cause of death, multiplying the number population deaths \( D_{asj} \) by the proportion of deaths that occur in hospital, \( P_{asj} \), should equal the number of hospital deaths \( H_{asj} \). It follows that if we have estimates of the number of hospital deaths and the proportion of deaths occurring in hospital, we can estimate the number of population deaths.

\( H_{asj} \) can be estimated from hospital data by summing deaths in age-sex-cause groups. \( P_{asj} \) requires in- and out-of-hospital mortality, which we can obtain from a vital registration system.

Note: research has shown that \( P_{asj} \)'s taken from vital registration data covering a different area than the hospital data can produce acceptable estimates of \( D_{asj} \), for the area covered by the hospital data.

Once we estimate \( D_{asj} \) as \( H_{asj} / P_{asj} \), we can estimate the cause-specific mortality fraction due to cause \( j \):

\[ \text{CSMF}_j = \frac{\# \text{ population deaths from cause } j}{\text{total } \# \text{ population deaths}} = \frac{\text{sum of } D_{asj} \text{ across ages and sexes}}{\text{sum of } D_{asj} \text{ across ages, sexes, causes}} \]

We may be interested in comparing these estimates to “hospital CSMFs,” the estimates you would get from using hospital data only:

\[ \text{Hospital CSMF}_j = \frac{\# \text{ in-hospital deaths from cause } j}{\text{total } \# \text{ in-hospital deaths}} \]

(For validation only): If complete vital registration data exists for the estimation area, then the measure of estimation error can be computed as an average relative error across all causes:

\[ \frac{1}{34} \sum_{j=1}^{34} \left| \frac{\hat{\text{CSMF}}_j}{\text{CSMF}_j} - 1 \right| \]

\[ \text{ARE} = \frac{\sum_{j=1}^{34} \left| \frac{\hat{\text{CSMF}}_j}{\text{CSMF}_j} - 1 \right|}{34} \]
**VARIABLE DEFINITIONS**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>year</td>
<td>Year of death of decedent</td>
</tr>
<tr>
<td>age</td>
<td>Age at time of death</td>
</tr>
<tr>
<td>agecat</td>
<td>Age category at time of death - see below</td>
</tr>
<tr>
<td>sex</td>
<td>Sex of decedent: 1=male, 2=female</td>
</tr>
<tr>
<td>icd</td>
<td>ICD-coded underlying cause of death of decedent or discharge diagnosis (ICD-9 or ICD-10)</td>
</tr>
<tr>
<td>hmlist</td>
<td>Corresponding short-list code; merged in using the ICD 10, 9, 8, or 6/7 map provided</td>
</tr>
<tr>
<td>hospital</td>
<td><strong>VR DATA ONLY</strong>: Indication of the place of death: 1=in-hospital, 0=out-of-hospital</td>
</tr>
<tr>
<td>residence</td>
<td><strong>OPTIONAL</strong> variable indicating subnational residence information of decedent</td>
</tr>
<tr>
<td>number</td>
<td><strong>AGGREGATED DATA ONLY</strong>: Number of deaths in the year-agecat-sex-icd(-residence) group</td>
</tr>
</tbody>
</table>
STARTING YOUR STATA SESSION

Throughout this guide,
- Commands to be typed into Stata’s command line will be in Courier font. Carefully observe quotation marks, commas, and parentheses.
- Pressing “Enter” on your keyboard will execute commands.
- Portions of commands that are italicized may require user-specific input (for example, “yourcountry”).

To begin your Stata session,

1. Open Stata and set the memory. Stata’s default memory allocation (10mb) is smaller than what you will need to read in your hospital and/or civil registration (CR) data, so you need to increase it before opening your dataset.

   set mem 500m

2. Set the working directory. This tells Stata where to look for data:

   cd “C:\Documents and Settings\All Users\Desktop\Workshop”

3. Open a log file. Your log file will record the output of the results window from now until you close it. You may find it to be a useful reference later. Put your initials and the date in the file name, without using any spaces.

   log using yourinitials_date_workshop

PART 1: Compute the number of hospital deaths per agecat-sex-cause group (H_{asj})

The first step is to prepare the hospital death data by aggregating deaths into agecat-sex-cause groups (where “agecat” refers to age category/group). Although CSMFs are reported for all ages and sexes combined, during the analysis we stratify by agecat and sex to reduce confounding, as age and sex both influence patterns of in-hospital mortality. In other words, computing $H_{asj}$ (# hospital deaths) and $P_{asj}$ (proportion of in-hospital deaths) is done for every agecat-sex-cause group separately because we expect those quantities to be quite different for different ages, sexes, and causes.

1. Open your hospital or civil registration data in Stata. Similar to how you open data in Excel, choose “File” → “Open” from the menu bar at the top, and find your data in the Workshop folder on the Desktop.

2. Now your variable list should list the variables in your data: year, age, agecat, sex, hmlist, and hospital IF you are using CR. To view these variables, you must open the Data Browser. Open the Data Browser and confirm that you have opened the correct dataset. This can be done most easily by typing

   browse
Note: once you are done looking at your data, you must close the Data Browser to reactivate the command line.

3. Depending on your data, you may want to limit the data according to a variable (note that your original data will not be affected by this step).

If you are using civil registration data, you must limit the data to hospital deaths only:

    keep if hospital == 1

4. Deaths with missing age or sex information cannot inform the agecat-sex-cause estimates, so they must be dropped from the data. Stata denotes missing values with a period, “.” (Note that dropping missing values is a short-term solution. If a significant percentage of your data has missing values, dropping these records has the potential to bias your results.)

    drop if agecat == .
    drop if sex == .
    drop if hmlist == .

Note: two separate lines of code requires that you press “Enter” after EACH line.

5. Recall that $H_{asj}$ is the number of deaths per agecat-sex-cause group. To calculate this quantity, generate a new variable “Hasj” that stores this value for each agecat-sex-cause group:

    bysort hmlist sex agecat: generate Hasj = _N

If you are using a shortlist of causes that is NOT “hmlist:” make sure to substitute the variable name of your shortlist every time you see “hmlist” as part of a command.

Note that we use age categories for the analysis; using individual ages would most likely result in too few deaths per $H_{asj}$ group.

6. The new “Hasj” variable has been created in a way that preserves the unit-record data. We no longer need the unit-record information, however. Condense the dataset to contain only the necessary aggregate information, namely the list of agecat-sex-cause groups and their corresponding number of hospital deaths:

    collapse (max) Hasj, by(hmlist sex agecat)

7. Browse the data to look at “Hasj.” Confirm that it generally varies for different agecat-sex-hmlist combinations. You may also want to confirm that it varies in ways you expect—for example, for hmlist 7 (birth asphyxia/trauma), you should see some number of deaths in agecat 0, but no deaths in the higher age categories.

    browse hmlist sex agecat Hasj

8. Sort your data for use later,
sort hmlist sex agecat

9. Save your data using the command line. Put your country’s name in the filename.

   save "yourcountry_Hasj"

You have now have a dataset of Hasj values that can be used in computing Hasj/Pasj = Dasj, the estimated number of population deaths per agecat-sex-cause group. We will put this dataset aside for now and move on to computing Pasj, the probability of dying in hospital per agecat-sex-cause group.

If you are calculating Pasj’s with your own civil registration data: proceed to Part 2.

Otherwise: Open the Pasj dataset you were assigned using File→Open, and proceed to Part 3. (However, you may want to read through Part 2 to understand the computation of Pasj’s.)

**PART 2: Compute the probability of in-hospital death for agecat-sex-cause groups (Pasj)**

Note: See Part 4 for details on doing multiple analyses with additional Pasj data sources.

1. Open your civil registration data in Stata. Similar to how you open data in Excel, choose “File” → “Open” from the menu bar at the top, and find your data.

2. **SKIP this step IF you are not limiting your CR to a particular region of the country.** The accuracy of the Pasj estimate depends on how well the CR captures both in- and out-of-hospital deaths. If your CR system is estimated to have low coverage in some areas, a concern is that the deaths not captured are primarily out-of-hospital deaths, since those are more logistically challenging to register than deaths in hospital. If you believe that certain regions have higher coverage than others, it may be better to restrict the data to those regions:

   keep if residence == regionnameorcode

3. Again, deaths with missing age or sex information cannot inform the agecat-sex-cause estimates, so they must be dropped from the data. Stata denotes missing values with a “.”

   drop if agecat == .
   drop if sex == .
   drop if hmlist == .

4. Recall that Pasj is estimated as

   Pasj = the proportion of deaths in an agecat-sex-cause group that occur in hospital.
   = (# in-hospital deaths observed in an agecat-sex-cause group)/(total deaths in that group)
Since the “hospital” variable is coded as 0 or 1, this proportion is just the mean of the “hospital” variable and we can generate it as such:

```stata
bysort hmlist sex agecat: egen Pasj = mean(hospital)
```

5. If extremely few deaths occur in hospital for a particular agecat-sex-cause group, $P_{asj}$ will be very small and will probably incorrectly inflate estimates of deaths from that cause (remember that the estimation process requires dividing by $P_{asj}$). For agecat-sex-cause groups with two or fewer deaths, set $P_{asj}$ to zero to avoid this problem:

```stata
duplicates tag hmlist sex agecat, gen(tag)
replace Pasj = 0 if tag < 2
```

6. Similar to step 6 of Part 1, condense the data,

```stata
collapse (max) Pasj, by(hmlist sex agecat)
```

7. Sort the $P_{asj}$ data

```stata
sort hmlist sex agecat
```

8. Save your data using the command line. Put your country’s name in the filename.

```stata
save "yourcountry_Pasj"
```

Now that the two quantities required for estimation, $H_{asj}$ and $P_{asj}$, have now been computed for matching agecat-sex-cause groups. Thus we can move on to estimating CSMFs.

**PART 3: Estimate cause-specific mortality fractions**

Since age, sex, and cause all influence whether a death occurs in hospital or not, $P_{asj}$ is unique for each agecat-sex-cause group and only describes the probability of dying in hospital for that one agecat-sex-cause group. Thus, in order to calculate $D_{asj} = H_{asj} / P_{asj}$, the values of $H_{asj}$ and $P_{asj}$ must be first be matched by agecat-sex-cause groups.

**Open your $P_{asj}$ dataset if it is not already open using File→Open.**

1. Before matching $P_{asj}$ and $H_{asj}$, take a look at the $P_{asj}$’s you have and confirm that age, sex, and cause do influence $P_{asj}$. You can do this by making a scatterplot of $P_{asj}$ vs hmlist:

```stata
scatter Pasj hmlist, xlabel(#34) xlabel(.)
```

The Y axis here is the probability of dying in hospital, and the X axis is the hmlist cause-of-death list. You should see a range of data points for each hmlist cause—this represents the range of $P_{asj}$’s from the different age categories and sexes. Using the hmlist definitions at the end of this document, observe whether different causes have different ranges of $P_{asj}$’s. For example, how do the $P_{asj}$’s for hmlist 6, maternal conditions, compare to the $P_{asj}$’s for hmlist 30, road traffic accidents?
Now try a scatter plot of $P_{asj}$ against agecat:

```
scatter Pasj agecat, xlabel(#34) ylabel(.
```

Now the X axis represents age categories. Is there an observable trend of probabilities over age? In the oldest age groups, it is likely that you will observe an overall downward trend in the $P_{asj}$'s due to many elderly dying at home or in nursing homes.

2. Sort the data once more:

```
sort hmlist sex agecat
```

3. The $P_{asj}$'s for each age-sex-cause group need to be matched with the corresponding $H_{asj}$ for that age-sex-cause group. This can be done by “merging” the two datasets:

```
merge hmlist sex agecat using "yourcountry_Hasj"
```

4. The merge created a “_merge” variable that indicates the success of the matching process. When _merge is 3, all observations in both datasets were matched. Check the success of the merge:

```
tab _merge
```

**NOTE: SKIP steps 5-6 IF _merge was 3 for all records.**

5. If _m is 1 for some records: this means that deaths in an agecat-sex-cause group were observed in the CR but not in the hospital data. Set $H_{asj}$ to 0 for these records:

```
replace Hasj = 0 if _merge == 1
```

6. If _m is 2 for some records:
   a) This means that means that deaths in an agecat-sex-cause group were observed in the hospital data but not in the CR. This is unusual but may happen. You may want to browse these records to understand them better:

   ```
browse if _merge == 2
```

   b) Since no deaths from these agecat-sex-cause groups were observed in the CR, our best guess is that all deaths from this group occurred in hospital. Set $P_{asj}$ to 1 for these records:

   ```
replace Pasj = 1 if _merge == 2
```

7. Recall that $D_{asj}$ is estimated as $H_{asj}/P_{asj}$. Generate the variable $D_{asj}$:

   ```
generate Dasj = Hasj/Pasj
```

8. If $P_{asj} = 0$, the division by zero will generate a missing value “.”. Again, it is unusual for this to happen—all deaths occurring out-of-hospital for the agecat-sex-cause group—but if it does, we cannot use $P_{asj}$ to help us estimate $D_{asj}$. The most
information we have is from the number of hospital deaths, so set $D_{asj} = H_{asj}$ in these cases:

\[ \text{replace } D_{asj} = H_{asj} \text{ if } D_{asj} == . \]

**Note:** this step highlights the importance of having enough deaths in the civil registration data to capture reliable estimates of $P_{asj}$. The more deaths in the data, the more likely it is that $P_{asj}$ will reflect the true probability. This is the motivation for pooling data across years—if we can assume that $P_{asj}$ does not vary sharply from year to year, then we can pool adjacent years to get more deaths and thus more reliable estimates of $P_{asj}$.

9. The $D_{asj}$’s are the estimates of the total number of deaths that occurred in each age-sex-cause group. Since

$$\text{CSMF}_j = \frac{\text{mortality from cause } j}{\text{total mortality}},$$

summing the $D_{asj}$’s over agecat and sex will yield the numerator.

\[ \text{bysort hmlist: egen numerator = sum(Dasj)} \]

10. The denominator is the sum of all the $D_{asj}$’s, or the estimated total number of deaths. Obtain this value by summing all the $D_{asj}$’s:

\[ \text{egen denominator = sum(Dasj)} \]

11. Now you can generate the CSMF estimates for each cause:

\[ \text{generate CSMF = numerator/denominator} \]

12. Condense the data,

\[ \text{collapse (max) CSMF, by(hmlist)} \]

13. Sort the data,

\[ \text{sort hmlist} \]

14. Save your data using File ➔ Save As or using the command line (include your country name as well as the source of the $P_{asj}$’s in the filename):

\[ \text{save “yourcountry est_CSMFs - Pasjsource Pasj”} \]

15. For use later in Excel, save your data as a comma-separated value file (note the following command is one line of text):

\[ \text{outsheet hmlist CSMF using “yourcountry est_CSMFs - Pasjsource Pasj.csv”, comma} \]

16. You can explore the estimates for each hmlist cause using a bar graph (note that there will be time later to explore these in more detail):

\[ \text{tw bar CSMF hmlist, xmlabel(#34) xlabel(.)} \]

24
This bar graph plots the CSMFs by cause. Refer to the cause list to observe which causes contribute the highest and lowest fractions. (If you want to save this graph, in the graph window, choose File→Save As.)

17. **Time permitting**, go on to the additional analyses. When you are finished with your Stata session for the day, close your log file:

```
log close
```

**PART 4: Repeat analysis with different data sources**

There are several potential data sources for $H_{asj}$ and $P_{asj}$:

- $H_{asj}$
  - Hospital databases from the country
  - Civil registration from the country
- $P_{asj}$
  - Civil registration from a region of the country
  - Civil registration from another country
  - Logistic regression model developed from pooling several countries’ CR data

You have completed the analysis using one combination of data sources. Repeating the analysis with different combinations will produce additional CSMF estimates that will be valuable for comparative purposes. Synthesizing and discussing the results will help us understand which data sources work best, in general and for your country in particular.

Fortunately, you do not have to retype all your commands into Stata to repeat the analysis. Stata allows the execution of pre-written sets of commands through files with the extension “.do”. The IHME/WHO staff has prepared a file that you can use. What you will need to do is edit the file to reflect the data you want to use.

1. Open the program Notepad (Start → Programs → Accessories → Notepad)

2. Open the file “workshop CSMFs.do” from Notepad. You can find it in the “Workshop” folder on the Desktop.

3. You should see the following at the top of the file:

```
local Hasjcountry = “countryname”
local Pasjcountry = “countryname”
local Hasj_source = “set filepath to the Hasj dataset created in Part 1”
local Pasj_source = “set filepath to a Pasj dataset in the Workshop folder on the Desktop”
```

4. Edit the above text in quotes with the appropriate country names and filepaths of the datasets. **For example**, if you are estimating for Mexico using CR from the United States,
local Hasjcountry = “Mexico”
local Pasjcountry = “US”
local Hasj_source = “C:\Documents and Settings\All
Users\Desktop\Workshop\Mexico_Hasj”
local Pasj_source = “C:\Documents and Settings\All
Users\Desktop\Workshop\US_Pasj”
local output = “C:\Documents and Settings\All Users\Desktop\Workshop”

5. Save “workshop CSMFs.do” using File → Save.

6. To execute the commands in “workshop CSMFs.do”, go back to Stata and choose File → Do from the menu, and open the “workshop CSMFs.do” file.

7. You can monitor the progress of the commands in the results window. The CSMF results will be saved in files names similarly to the ones you saved in steps 14-15 of Part 3, except with the “Pasjsource” portion reflecting the new Pasj source country.

**PART 5: Estimate observed cause-specific mortality fractions from CR:**

**Note:** Parts 5 and 6 are for those participants who brought CR data.

1. Start with your CR data as you did in Part 2.

2. Recall that a CSMF_j = (mortality from cause j)/(total mortality). The denominator is simple: it is the total number of deaths recorded in your hospital data. In the case of unit-record data, this is the same as the number of observations in the data. Count the number of observations in the data:

   count

3. You should see the results of the “count” command in the results window. Store this value in a variable called “denominator”:

   generate denominator = _N

   (You can confirm that this is the correct number by typing tab denominator).

4. To calculate the mortality from each cause, we can generate a new variable that stores the number of observations per “hmlist” cause:

   bysort hmlist: generate numerator = _N

5. Condense the data:

   collapse (max) numerator, by(hmlist denominator)

6. Browse the data to understand the new variable “numerator.”

   browse hmlist numerator
7. Now we can generate CSMF\textsubscript{j} by dividing the numerator by the denominator,

\begin{equation*}
\text{generate obs}_\text{CSMF} = \text{numerator/denominator}
\end{equation*}

8. Sort the data,

\begin{equation*}
\text{sort hmlist}
\end{equation*}

9. Save your data using the command line:

\begin{equation*}
\text{save "yourcountry}_\text{obs}_\text{CSMFs"}
\end{equation*}

10. For use later in Excel, save your data as a comma-separated value file (note that the following command is meant to be all one line):

\begin{equation*}
\text{outsheet hmlist obs}_\text{CSMF using "yourcountry}_\text{obs}_\text{CSMFs.csv", comma}
\end{equation*}

11. There will be time later to look at these CSMFs more closely. For now, you can informally explore them using a bar graph:

\begin{equation*}
\text{tw bar obs}_\text{CSMF hmlist, xlabel(#34) ylabel(.)}
\end{equation*}

(If you want to save this graph, in the graph window, choose File→Save As.)

**PART 6: Compute average relative error of CSMF estimates**

**Note:** Parts 5 and 6 are for those participants who brought CR data.

For a region with CR data of high completeness, the best estimates of CSMFs for the region will come directly from the CR data. It is thus a valuable exercise to estimate CSMFs using the method and compare the results to the CR estimates. This will contribute to our understanding of how well the method performs in countries like yours and/or with the data sources you used.

1. Open your results from step 5 in Stata using File→Open or the command line:

\begin{equation*}
\text{use "yourcountry}_\text{obs}_\text{CSMFs", clear}
\end{equation*}

2. Sort the data

\begin{equation*}
\text{sort hmlist}
\end{equation*}

3. Merge the observed CSMFs with a set of method-estimated CSMFs.

\begin{equation*}
\text{merge hmlist using "yourcountry}_\text{est}_\text{CSMFs - Pasjsource Pasj"}
\end{equation*}

4. Check that your merge was successful, _merge = 3:
5. The “relative error” for each cause can be thought of as the proportion of the “true” value by which the estimated value deviates. In this case, we consider the true value to be an observed CSMF from the CR (stored in your variable “obs_CSMF” from Part 5), and the estimated value the CSMF from the method (stored in your variable “CSMF” from Part 4). Then the relative error is

\[
\text{Relative Error for cause } j = \left( \frac{CSMF_i}{obs\_CSMF_i} \right) - 1
\]

This quantity will be positive if the method overestimates, and negative if the method underestimates. Generate a variable of relative errors:

```
generate RE = CSMF/obs_CSMF - 1
```

6. Use a bar graph to look at the relative errors to see which causes have been overestimated, which have been underestimated, and which seem to be outliers. Consult Appendix A for descriptions of hmlist causes.

```
tw bar RE hmlist, xlabel(#34) ylabel(.)
```

7. We are also interested in one average error estimate (ARE) across all causes as an overall measure of the method’s average accuracy. This can be computed easily from the relative error by taking the absolute value for each, and then taking the average of the absolute relative errors (summing them all and then dividing by the number of causes, 34):

\[
\text{ARE} = \frac{\sum_{j=1}^{34} \left| \left( \frac{CSMF_i}{CSMF_j} \right) - 1 \right|}{34}
\]

Generate a variable with the absolute values of the relative errors:

```
generate absRE = abs(RE)
```

8. Use Stata’s “summarize” function to find the average or mean of these absolute errors. The “mean” value you see is the average relative error (ARE).

```
summarize absRE
```

How does the method do?

9. Save your data,

```
save "yourcountry RE - Pasjsource Pasj"
```

10. Also save your data in .csv format for use in Excel (note the following is one line of text):

```
outsheet hmlist RE using "yourcountry RE - Pasjsource Pasj.csv", comma
```
<table>
<thead>
<tr>
<th>hmlist</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>2</td>
<td>HIV/AIDS</td>
</tr>
<tr>
<td>3</td>
<td>Diarrhoeal diseases</td>
</tr>
<tr>
<td>4</td>
<td>Other infectious and parasitic diseases</td>
</tr>
<tr>
<td>5</td>
<td>Respiratory infections</td>
</tr>
<tr>
<td>6</td>
<td>Maternal conditions</td>
</tr>
<tr>
<td>7</td>
<td>Birth asphyxia and birth trauma</td>
</tr>
<tr>
<td>8</td>
<td>Other perinatal conditions</td>
</tr>
<tr>
<td>9</td>
<td>Nutritional deficiencies</td>
</tr>
<tr>
<td>10</td>
<td>Malignant neoplasms, specified</td>
</tr>
<tr>
<td>11</td>
<td>Other malignant neoplasms</td>
</tr>
<tr>
<td>12</td>
<td>Benign neoplasms</td>
</tr>
<tr>
<td>13</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>14</td>
<td>Endocrine disorders</td>
</tr>
<tr>
<td>15</td>
<td>Neuropsychiatric conditions</td>
</tr>
<tr>
<td>16</td>
<td>Rheumatic heart disease</td>
</tr>
<tr>
<td>17</td>
<td>Hypertensive heart disease and inflammatory heart disease</td>
</tr>
<tr>
<td>18</td>
<td>Ischaemic heart disease</td>
</tr>
<tr>
<td>19</td>
<td>Cerebrovascular disease</td>
</tr>
<tr>
<td>20</td>
<td>Other cardiovascular diseases</td>
</tr>
<tr>
<td>21</td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>22</td>
<td>Asthma</td>
</tr>
<tr>
<td>23</td>
<td>Other respiratory diseases</td>
</tr>
<tr>
<td>24</td>
<td>Peptic ulcer disease, cirrhosis of the liver, and appendicitis</td>
</tr>
<tr>
<td>25</td>
<td>Other digestive diseases</td>
</tr>
<tr>
<td>26</td>
<td>Genitourinary diseases</td>
</tr>
<tr>
<td>27</td>
<td>Musculoskeletal diseases</td>
</tr>
<tr>
<td>28</td>
<td>Skin diseases, sense organ diseases, and oral conditions</td>
</tr>
<tr>
<td>29</td>
<td>Congenital anomalies</td>
</tr>
<tr>
<td>30</td>
<td>Road traffic accidents, poisonings, falls, fires, drownings</td>
</tr>
<tr>
<td>31</td>
<td>Other unintentional injuries</td>
</tr>
<tr>
<td>32</td>
<td>Intentional injuries</td>
</tr>
<tr>
<td>33</td>
<td>Ill-defined diseases</td>
</tr>
<tr>
<td>34</td>
<td>Ill-defined injuries/accidents</td>
</tr>
</tbody>
</table>
Appendix 4. Group Presentations

African Group

WHO HOSPITAL DATA WORKSHOP
MAY 2008

DATABASE ANALYZED:

AFRO GROUP
ETHIOPIA
GHANA
KENYA
MOZAMBIQUE
TANZANIA

Kenya

MOZAMBIQUE

USE OF PROPORTION GIVES DIFFERENT RESULTS

Kenya

Issues

Weakness of vital registration system
Garbage code
Assigning of wrong codes
Difficulty in attribution of the right ICD diagnosis
(Mortality – Morbidity)
Region specific ICD cause group mortality list
Isolate predominant diseases e.g. malaria
Data quality
Discharge diagnosis vs cause of death
Cause of deaths e.g. cancers not covered by health insurance
Death certificate vs patient files

Recommendations

Strengthen country-specific mortality information system
Hospital EMRs
Vital registration
Data analysis software
Avail/State and assure use
Or other options
Training on:
Stata
ICD Ix-coding/mapping for HCWs-physicians, coders

DATABASE ANALYZED:

ETHIOPIA: H E1, Pasj SA – Iran – Mex – US
GHANA: H Ga, Pasj SA – Iran – Max – US
KENYA: H K, Pasj SA – Iran – Max – US
MOZAMBIQUE: HCR Moz, Pasj S.A. – CR Mex
TANZANIA: H Max, Pasj SA – Iran – Max – US

MOZAMBIQUE

USE OF PROPORTION GIVES DIFFERENT RESULTS

TANZANIA:

MOZAMBIQUE:

KENYA:

GHANA:

DATABASE ANALYZED:

ETHIOPIA: H E1, Pasj SA – Iran – Mex – US
GHANA: H Ga, Pasj SA – Iran – Max – US
KENYA: H K, Pasj SA – Iran – Max – US
MOZAMBIQUE: HCR Moz, Pasj S.A. – CR Mex
TANZANIA: H Max, Pasj SA – Iran – Max – US

African Group
Next Steps

1. To elaborate a Region specific ICD cause group mortality list
2. To develop training on ICD-10 and mortality coding system (Hospital and CR)
3. To develop regional mechanics for analysis of data and linking with regional policy
4. To develop a regional minimum standard for mortality information system (age group, sex, race etc.)

American Group 1

CSMF's Nicaragua - Pasj Mexico - EEUU - logit

Bahia
Background - Barbados
- Total Pop – 270,000 (2001)
- Hospital Mortality Data
  - Total Entries - 6,518
  - Average Annual Death – 2,300
- Civil Registration Data - Exist but not used for this analysis
- Deaths Certified – 100%

Background - Guyana
- Total Pop. – 751,400 (2002)
- Hospital Mortality Data – 2006
  - Total from Main Hospital - 1,732
- Civil Registration – 2001 – 2007
  - Total Entries - 33,079
  - Average Annually deaths – 5,000
  - Sex – 2
  - Age – 387
  - Place of death – 204
- Deaths Certified – 72%

Background - Suriname
- Total Pop – 500,000 (2005)
- Certified Mortality Data
  - Years – 2002 - 2006
  - Total Entries – 14,992
  - Average Annual Death – 3,300
- Hospital Deaths Certified – 70%

Finding - Barbados

Findings - Guyana
Finding - Suriname

Suriname main causes of death

- HIV/AIDS
- Malignant neoplasms
- Diabetes mellitus
- Hypertensive heart disease
- Inflammatory heart disease
- Ischaemic heart disease
- Cerebrovascular disease
- Road traffic accidents, poisonings, falls, fires, drownings
- Intentional injuries
- Other cardiovascular diseases
- Ill-defined diseases

Guyana Garbage Code

- #20 (4.1 – 5.5%)
  - 40% Heart Failure
  - 12% Unspecified Heart Failure
  - 5% Atherosclerosis
- #33 (0.8 – 2.5%)
  - 37% Senility
  - 22% Other Unspecified
  - 7% Malaise/Fatigue

Comments

- Under/Late reporting
  - From the interior (Guy, Sur)
  - Hospitals (Sur)
- Missing Information on Death Certificate (Guy, Sur)
- Inability to identify underlying cause of death (All)
- All Use Death Certificate
- Data sets are too small for analysis

Future Collaboration

- Use of Multi year data in individual countries
- Increase collaboration with Civil Registry (assembling in ICD 10 coding)
- Pooling of Data for Caribbean Sub Region to complete regression analysis
- Improve systems for compilation of mortality data
- Ask/Implement WHO methods/recommendations for Mortality analysis

Asian Group

Workshop Hospital data
Vevey 16 May 2008
Summary Presentation

- Malaysia
- Mongolia
- Myanmar
- Philippines
- Thailand
- Sri Lanka
Facilitator: Ms Mie INOUE

Data Source

- Civil registration: Philippines, Thailand, Sri Lanka
- Hospital statistics: Myanmar, Mongolia, Malaysia
- Data series vary between 1-5 years
Leading CSMF

- Philippines – Diabetes (13)
- Sri Lanka, Thailand, Myanmar, Malaysia – ill defined conditions (33)
- Mongolia – Specified Malignant (10)

Second Leading CSMF

- Philippines, Mongolia – CVDs (19)
- Sri Lanka – Neuropsychiatry (15)
- Thailand – Specified Malignant (10)
- Myanmar, Malaysia – Other infections and parasitics infections (4)

CSMF changes using different Pasj

(Example of Malaysia)

<table>
<thead>
<tr>
<th>Country</th>
<th>CSMF</th>
<th>hmlist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>0.125054</td>
<td>33 4 10</td>
</tr>
<tr>
<td>USA</td>
<td>0.125584</td>
<td>6 10</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.1312576</td>
<td>2 1 1</td>
</tr>
</tbody>
</table>

Garbage code 33 analysis observed vs model

- Thailand: >35% for observed, ~<10% with models
- Sri Lanka: 22 observed, ~12% with models
- Philippines, Mongolia: <5% observed and model

Garbage code 33 analysis variation with models

CSMF by agecat (example of Philippines)

Recommendations

- Improve quality of raw data collection
- Use regional Pasj
- WHO/HMN/IHME make a STATA programme available at subsidize cost
- WHO/HMN/IHME provides a programme codes of STATA and other statistical package
- Have a network of Hospital mortality data method for regular review
Eastern Mediterranean Group

Dataset
- **Algeria**: Hospital data (34% of total deaths in 2006)
- **Egypt**: Hospital data and Civil Registration data (95% coverage in 2007)
- **Oman**: Hospital data and Civil Registration data (90% coverage in 2005-2007)
- **Saudi Arabia**: Hospital data (sample of 5,000 deaths in the dataset)

Comparing different methods for Algeria

Comparing different methods for Egypt

Vital Registration with blank codes removed from the dataset

Hospital dataset

Percentage of deaths by age group from Vital Registration with blank codes removed from the dataset for different causes
European Group

1. Countries: Belarus, Georgia, Kazakhstan, Poland, Turkey

2. Data available: Relatively full datasets were available for Kazakhstan and Poland. Partially complete data for the capital of Georgia (Tbilisi) – in the vital registration the place of death is not mentioned. For Belarus only hospital data for the capital city (Minsk) are available for analysis.

3. General findings:
   - Coding of cases are different in the hospital statistics and the VR. An assumption is made that hospital data contain clinical diagnosis for the last episode or diagnosis at discharge rather than an underlying cause of death.

4. Kazakhstan: \( H_{asj} \) and \( P_{asj} \) were generated using data for Kazakhstan
   - The cause of death distribution is similar in hospital data and VR.
   - Garbage codes were high: I50 (heart failure) makes up to 25% of the "other cardiovascular causes" group (group 20).

5. Poland: \( H_{asj} \) and \( P_{asj} \) were generated using data for Poland
   - The cause of death distribution is different in hospital data and VR, especially for ill-defined deaths (group 33).
   - Garbage codes: The fraction of deaths coded to group 20 is 50% larger in hospital data than in VR. In the VR ICD10 code I70 (atherosclerosis) makes up to 41% of group 20.

6. Georgia: No place of death is mentioned in the VR. Thus, \( P_{asj} \) from the US and Mexico were used. Results were similar with both \( P_{asj} \).
   - Garbage codes: I50 makes up approximately 33% of group 20; R57 (shock, not elsewhere classified) makes up 80% of group 33.

6. Belarus. No data from VR is available. Thus, \( P_{asj} \) from the US and Mexico were used. Results were similar with both \( P_{asj} \).
   - Garbage codes: There were almost no garbage codes (less than 4%).
Appendix 5. List of Participants

WORKSHOP ON HOSPITAL MORTALITY DATA ANALYSIS

VEVEY, SWITZERLAND, 15 - 16 May 2008
18 June 2008

LIST OF PARTICIPANTS

Temporary Advisers

Dr Soraya Belamri
Maître assistante en épidémiologie
Institut National de santé publique
4 chemin El Bakr
El Biar, Alger
ALGERIA
Telephone No. : 00213 773 669 614
Email : sbelamri@hotmail.com

Lic Carlos Guevel
Estadistico
Dirección de Estadísticas e Información de Salud
Ministerio de Salud de la Nación
Avenida 9 de julio 1925, piso 6to
oficina 604 C1073ABA Capital Federal
ARGENTINA
Telephone No. : 0054 11 4379 9024
Fax No. : 0054 11 4381 2015
Email : cguevel@deis.gov.ar

Dra Elida Marconi
Directora
Dirección de Estadísticas e Información de Salud (DEIS)
Av. 9 de Julio 1925, Piso 6
C1073ABA - Capital Federal
ARGENTINA
Telephone No. : 0054 11 4379 9024
Fax No. : 0054 11 4379 9024
Email : emarconi@deis.gov.ar

Mr Danny Gill
Director of Policy, Planning & Research
Ministry of Health
Jemotts Lane
Saint Michael
BARBADOS
Telephone No. : 001 246 426 5080
Fax No. : 001 246 426 5570
Email : directorpru@gmail.com
Dr A Grakovich
Director
Belarus Republican Scientific and Practical Centre for Health Care Technologies, Informatics, Management and Economy
ul. P. Brovki 7a
Minsk
BELARUS

Mr Antony Stevens
Departamento de Analisis de Situación de la Salud
Esplanada dos Ministéíos - Bloco G-1 piso-Sala 150 Brasilia
BRAZIL

Dr Jesus Rodriguez
Profesor Asociado
Centro de Proyectos para el Desarrollo-CENDEX - Universidad Javeriana Carrera 7 numero 40-90, Edificio Emilio Aranbo 5 piso Ap. aéreo 74791 Bogota
COLOMBIA

Msc Rosa Vargas
Jefe de la Unidad de Estadistica Ministerio de Salud Detras Más por Menos, Guadalupe, San José
COSTA RICA

Lic Estuardo Albán
Jefe de Estadisticas Sociales y de Salud Instituto Nacional de Estadisticas y Centros (INEC) Juan Larrea N15-36 y José Riofrio Casilla Postal 135C, Quito
ECUADOR

Dr Sohier Saad Botrous
General Manager Health Information Center Ministry of Health and Population 3 Magles El Shaab Street Kasr El Aini St. Cairo 11467
EGYPT
Lic Marlene Barrientos
Jefe de la Unidad de Información en Salud
Ministero de Salud Pública y Asistencia Social
Calle Arce No. 827
San Salvador
EL SALVADOR

Telephone No. : 00503 2205 7199
Fax No. : 00503 2222 3324
Email : mbarrientos@mspas.gob.sv

Mr Gadissa Lemecha
Head HMIS
Ministry of Health
Department of Planning and Program Development
Addis Ababa
ETHIOPIA

Telephone No. : 00251911693380
Email : gadissal@yahoo.com

Dr Marina Shakhnazarova
Head
Division of Data Analysis and Presentation
National Center for Disease Control and Public Health
9 M Asatiani Street
0177 Tbilisi
GEORGIA

Telephone No. : 0099595956101
Email : mshakh@caucasus.net

Mr Daniel Darko
Head
Centre for Health Information Management
Ghana Health Service
P.O. Box GP 2848
Accra
GHANA

Telephone No. : 0023321668152
Email : kddarko@yahoo.com

Dr Shamdeo Persaud
Chief Medical officer
Ministry of Health
1 Brickdam
Georgetown
GUYANA

Telephone No. : 00592 226 1224
Fax No. : 00592 225 0113
Email : cmoguyana@gmail.com

Dr Gasima Bermagambetova
Head
Health Statistics and Information Analysis Department
Ministry of Health of Kazakhstan
66 Moskovskaya str.
473000 Astana
KAZAKHSTAN

Telephone No. : 0077172 743169
Fax No. : 0077172 74 31 51
Email : g.bermagambetova@mz.gov.kz
Dr Sergon Kibet
Head
Division of HMIS
Afya House, Cathedral Road
P.O. Box 30016, Nairobi
KENYA

Telephone No.: 00254 020 353 2622
Email: kibetsergon@yahoo.com

Dr Jameela Binti Zainuddin
Senior Principal Assistant Director
Malaysia National Health Account (MNHA) Unit
Planning and Development Division
Ministry of Health Malaysia
Level 6, Block E6 Complex E
Federal Govt. Administrative Centre
MALAYSIA

Telephone No.: 00603 8883 2094
Email: jzmohealth@yahoo.com

Ms Khad Narantuya
Head
Department of Statistics and Information National Center for Health Development (NCHD)
P.O.Box 187, Enkhtaivan Street - 13B
Ulaanbaatar 210648
MONGOLIA

Telephone No.: 00976881 18099
Email: Khadnaraa@yahoo.com

Dr Alessandro Campione
Assessor
Departamento Informação para a Saúde Vigilancia Epidemiológica
Ministério da Saúde
Ave Eduardo Mondlane 1081
Maputo
MOZAMBIQUE

Telephone No.: 00258 82 811 4130
Email: alessandro.campione@gmail.com

Dr Ercilia de Almeida
Head of HIS
Ministério da Saúde, Mozambique
Av Eduardo Mondlane 1008
CP 264, Maputo
MOZAMBIQUE

Telephone No.: 000258 213098 73
Email: ealmeida@misau.gov.mz

Dr Thet Thet Mu
Deputy Director HMIS
Department of Health Planning
Nay Pyi Taw
Ministry of Health
MYANMAR

Email: thetthetmu@gmail.com
Dr Luis Carballo Palma
Resp. Oficina de Estadísticas
División de Sistemas de Información
DGPD
Ministerio de Salud
Complejo Nacional de Salud Concepción
Palacios, Aptdo Postal 107, Managua
NICARAGUA

Dr Salah Nasser Al-Muzahmi
Director
Health Information and Statistics
Ministry of Health
P.O. Box 393
Postal cod 100, Muscat
OMAN

Dr Rosalinda Arandia
Medical Center Chief
Quirino Memorial Medical Center
P.Tuazon Street, Project 4
Quezon City
PHILIPPINES

Ms Lourdes J Hufana
Director
Civil Registration Department
National Statistics Office
3rd Floor, Vibal Building
Times Street Corner EDSA
West Triangle, Quezon City 1100
PHILIPPINES

Dr Pawel Gorynski
Head of Department of Medical Statistics
National Institute of Hygiene
Chocimska 24
00-791 Warsaw
POLAND

Mr Nabil Othman
Supervisor of Information Center
Ministry of Health
P.O. Box 11176
Riyadh
SAUDI ARABIA
Dr WMTB Wijekoon  
Director, Planning  
Ministry of Healthcare and Nutrition  
Dean's Road  
Colombo 8  
SRI LANKA  
Telephone No.: 0094 112 674 683  
Email: wmtbw@health.gov.lk

Dr Widya Punwasi  
Coordinator Vital Statistics of the  
Epidemiology Unit of the Bureau of  
Public Health  
Ministry of Health  
Rode Kruislaan 22  
Paramaribo  
SURINAME  
Telephone No.: 00597 499703  
Fax No.: 00597491452  
Email: widyapunwasi@yahoo.com

Mr Claud John Kumaija  
Head of HMIS  
Directorate of Policy and Planning  
Ministry of Health and Social Welfare  
P.O. Box 9083  
Dar es Salaam  
UNITED REP. TANZANIA  
Email: claudjohnk@yahoo.com

Dr Narong Kasitipradith  
Deputy Director  
Bureau of Planning and Strategy  
Ministry of Public Health  
Royal Thai Government  
Tivanond Road  
Nonthaburi 11000  
THAILAND  
Email: narong@health.moph.go.th

Dr Hakki Gürsöz  
Deputy Director  
Ministry of Health of Turkey  
School of Public Health  
Sihhiye  
Ankara  
TURKEY  
Telephone No.: 00903123091224  
Email: hakki.gursoz@hm.saglik.gov.tr

Dr Handan Kalaycioglu  
Department of Strategy Development  
Ministry of Health  
Sihhiye  
Ankara  
TURKEY  
Telephone No.: 0090312485 5303  
Email: h.kalaycioglu@hotmail.com
Institute for Health Metrics and Evaluation

Mr Dennis Feehan
Researcher
Institute for Health Metrics and Evaluation
University of Washington
2301 5th Ave. Suite 600
Seattle, WA 98121
USA
Email : feehan@u.washington.edu

Ms Jeannette Kurian
Post Bachelor Fellow
Institute for Health Metrics and Evaluation
University of Washington
2301 5th Ave, Suite 600
Seattle, WA 98121
USA
Email : kurian@u.washington.edu

Dr Rafael Lozano
Visiting Professor
Institute for Health Metrics and Evaluation
University of Washington
2301 5th Ave, Suite 600
Seattle, WA 98121
USA
Telephone No. : 001206 897 2836
Email : rlozano@u.washington.edu

Dr Christopher JL Murray
Director
Institute for Health Metrics and Evaluation.
Department of Global Health
University of Washington
Box 356340
Seattle, WA 98195-8166
USA
Email : cjlm@u.washington.edu

World Health Organization

Headquarters

Mrs Carla ABOU-ZAHR
Coordinator, CHI
Telephone No. : 13367/14678
Email : abouzahr@who.int

Dr Ties BOERMA
Director, MHI
Telephone No. : 11481
Email : boermat@who.int

Ms Jessica HO
Technical Officer, SEV
Telephone No. : 12290
Email : hoj@who.int

Mr Kacem IAYCH
Intern: SEV
Email : iaych@who.int
Ms Mie INOUE  
Statistician, SEV  
Telephone No. : 12309/12855  
Email : inouem@who.int

Dr Robert JAKOB  
Medical Officer, CAT  
Telephone No. : 15877  
Email : jakobr@who.int

Ms Veronique JOSEPH  
Technical Officer, SEV  
Telephone No. : 14379  
Email : josephv@who.int

Mrs Doris MA FAT  
Statistician, SEV  
Telephone No. : 12841  
Email : mafatd@who.int

Ms Wahyu Retno MAHANANI  
Technical Officer, CHI  
Telephone No. : 11017  
Email : mahananiw@who.int

Dr Nosakhare Gregory OROBATON  
Manager, HMN  
Telephone No. : 12267  
Email : orobatonn@who.int

Mrs Susan PICCOLO  
Secretary, SEV  
Telephone No. : 12855  
Email : piccolos@who.int

Dr Kenji SHIBUYA  
Coordinator, SEV  
Telephone No. : 12370  
Email : shibuyak@who.int

Dr Sally Katherine STANSFIELD  
Executive Secretary  
HMN  
Telephone No. : 13592  
Email : stansfields@who.int

Dr Gretchen STEVENS  
Technical Officer, SEV  
Telephone No. : 11031  
Email : stevensg@who.int

Temporary Adviser - WHO

Dr Remigijus Prochorskas  
Asigalio str. 35-40  
Kaunas, LT-49148  
LITHUANIA  
Telephone No. : 00370 37721534  
Email : rpr@kaunas.init.lt

Other Regional Offices

Dr Edoh W. SOUMBEY-ALLEY  
Statistician  
AFRO  
Email : soumbeye@afro.who.int
Dr Samuel MIKHAIL
DHS/EIP
EMRO

Dr Sahar PARVEZ
EMRO

Dr Alejandro GIUSTI
PAHO

Dr Fatima MARINHO
Chief, Health Analysis and Statistics
PAHO

John SILVI
Statistician, Health Analysis and Statistics
PAHO

Dr Gunasena Sunil SENANAYAKE
Regional Adviser
SEARO

Email: mikhails@emro.who.int

Email: parvezs@emro.who.int

Telephone No.: 56 2 437 4627
Email: giustia@chi.ops-oms.org

Email: marinhof@paho.org

Telephone No.: 001 202 974 3141
Email: silvijhn@paho.org

Email: senanayakes@searo.who.int
Appendix 6. Host Organization Overviews

Health Metrics Network (HMN)

Health Metrics Network (HMN) is a global partnership that facilitates better health information at country, regional and global levels. Partners include developing countries, multilateral and bilateral agencies, foundations, other global health partnerships and technical experts. Most importantly, HMN seeks to bring together health and statistical constituencies in order to build capacity and expertise and enhance the availability, quality, dissemination and use of data for decision-making.

For more information visit [www.who.int/healthmetrics/en](http://www.who.int/healthmetrics/en)

Institute for Health Metrics and Evaluation (IHME)

The Institute for Health Metrics and Evaluation at the University of Washington works to monitor global health conditions and health systems, as well as to evaluate interventions, initiatives, and reforms and as such provide high quality and timely information on health so that policymakers, researchers, donors, practitioners, local decision-makers, and others can better allocate limited resources to achieve optimal results.

For more information visit [www.healthmetricsandevaluation.org](http://www.healthmetricsandevaluation.org)

World Health Organization (WHO)

WHO is the directing and coordinating authority for health within the United Nations system. It is responsible for providing leadership on global health matters, shaping the health research agenda, setting norms and standards, articulating evidence-based policy options, providing technical support to countries and monitoring and assessing health trends.

For more information visit [www.who.int/healthinfo/en/](http://www.who.int/healthinfo/en/)