Regional Capacity Building Workshop on Undertaking Inequality Assessments in Civil Registration

2-5 August 2022 (Virtual)

Renee Sorchik
Consultant, Civil Registration and Vital Statistics Expert
Day 2: Wednesday, 3 August 2022, 11:00-14:00 (BKK Time)

Assessing data for usability
Today we will cover:

1. Considerations for using data from household surveys
2. Complexities of measuring disaggregated (birth and) death registration completeness
3. Do your definitions align sufficiently enough between your different data sources to use the data for analysis?
4. How to assess the quality of the data to determine if it’s usable for analysis?
Lecture 7: Considerations for disaggregation

Renee Sorchik
Consultant, Civil Registration and Vital Statistics Expert
Survey data considerations

• In some cases, surveys such as the MICS or DHS for birth registration completeness, or other household surveys for death registration completeness may already provide the evidence needed for an IA

• If results are published, you can review the data and consider the groups being left behind, and the recommendations for improving registration within these disadvantaged groups

• Even if the published report is only representative at the national level, sometimes a dataset is available which can allow you to disaggregate further
Survey data considerations

- Reports and datasets from such surveys are useful to understand variations in registration completeness, but may need to be taken with a note of caution, especially if underscored in the methodology.

- In these cases, it’s important to read the methodology and see if there are any factors for considerations for disaggregated data.
  - For example, were there any regions that were not part of the sampling frame? In which case they would not be represented in the data.
  - What is the sample size of households per region?
  - What was the response rate?
  - Did any areas or groups of people (certain ethnicities, single mothers, women-headed households etc.) have an unusually low response rate?
  - Was the sample married women only? What about mothers under age 15?
  - Who may have been left out of the survey?
Sometimes there is no existing data and we have to perform analyses on our own

- Registered deaths (for a specific sub-group) from the Registrar General’s Office as the numerator, divided by estimated deaths (from the same sub-group) from a variety of different data sources (discussed more in later slides).

Deaths registered by Registrar General during a specified time period
Estimated number of deaths from another data source for the same time period
Methods of measurement

- Direct calculation with a deaths “gold standard” – a source that is considered to be “true”

\[
\text{Completeness of death registration (\%)} = \frac{\text{Number of registered deaths}}{\text{Actual number of deaths}} \times 100
\]

- Direct calculation using questions on a census or survey of registration completeness, administrative data sources

- Indirect demographic methods to estimate deaths (Brass Growth Balance, Bennett–Horiuchi, Preston-Coale)

- Capture re-capture methods
Goal 1D: % deaths registered

Completeness of death registration (%) = \( \frac{\text{Number of registered deaths in a year}}{\text{Actual number of deaths in a year}} \times 100 \)

• **Numerator: from civil registration data**
  • Number of deaths registered that calendar year

• **Denominator (# of deaths that year) should come from the “best” source:**
  • Recorded deaths from MOH if complete
  • Census data
  • Estimates derived from census data
  • Estimates derived from surveys or sample registration systems
  • Estimates derived from indirect demographic methods
  • Global Burden of Disease Estimates
  • If no other data, use crude death rate from Census or UN Statistical yearbook
But remember, more detail is needed for inequality assessments...

Deaths registered in CAPITALVILLE to children age 1 year or less by Registrar General between Jan 1, 2020 to Dec. 31, 2021

Deaths in MoH system in CAPITALVILLE for children age 1 year or less between Jan 1, 2020 to Dec. 31, 2020

Are these defined the same?
Complexities of measuring (birth and) death registration completeness
Death registration completeness for 2020

Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General

Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020

But registered when?????
Death registration completeness for 2020

But registered when????? OPTION 1: Same year as registration

Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General

Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020
Death registration completeness for 2020

But registered when????? OPTION 1: Same year as registration

 Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General

 Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020
Death registration completeness for 2020

But registered when????? OPTION 1: Same year as registration

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy to measure</td>
<td>“Inequitable” towards later year events</td>
</tr>
<tr>
<td>Timely, do not need to wait (a year) to measure</td>
<td>Unlikely to include allowable legal time for registration for December events</td>
</tr>
</tbody>
</table>

Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General

Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020
Death registration completeness for 2020

But registered when????? OPTION 2: Registered at any point in time (all events in current database)

_Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General_

_Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020_
Death registration completeness for 2020

But registered when?????? OPTION 2: Registered at any point in time (all events in current database)

2010 would have low registration completeness using OPTION 1. But completeness would go up with time even though all deaths occurred in 2010.

Caution: Make sure to select deaths that occurred in the year of interest as opposed to those that were registered that year. E.g. deaths registered in 2020 may have occurred in 2010
Death registration completeness for 2020

But registered when????? OPTION 2: Registered at any point in time (all events in current database)

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy to measure</td>
<td>Measurement changes over time for same year, doesn’t tell you how well system is improving</td>
</tr>
<tr>
<td>Timely, can use what is in database at current time</td>
<td>Hard to make comparisons internationally or within same country to other years</td>
</tr>
</tbody>
</table>

*Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General*

*Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020*
OPTION 3: Registered within 1 year* of the time of occurrence

Registered between Jan 1, 2020 to Jan 1, 2021

Deaths occurring
Jan. 1, 2020

Jan. 1, 2021

Deaths occurring in 2020

Deaths occurring Dec. 31, 2020

Dec. 31, 2021

*Or could use specified legal registration time frame. A specified delay is needed for comparison purposes
OPTION 3: Registered within 1 year of the time of occurrence

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement does not change over time for same year</td>
<td>More difficult to measure, need calculations and more variables</td>
</tr>
<tr>
<td>Can make comparisons internationally or within same country to other years (how is completeness improving over time?)</td>
<td>Not as timely, must wait at least a year from last date of event to calculate</td>
</tr>
</tbody>
</table>
Death registration completeness for 2020

Deaths occurring between Jan 1, 2020 to Dec. 31, 2020 that are registered by Registrar General

Deaths in MoH system occurring between Jan 1, 2020 to Dec. 31, 2020

What about the denominator?

Assumption: There is not as much delay in the recording of deaths. If this is not the case in your country, similar considerations would need to be made.
Birth registration completeness

• Similar considerations will need to be made for birth registration completeness.

• Generally speaking, this affects the numerator (registered events) more than the denominator (estimated events).

• Estimated births through MOH databases such as birth notifications, vaccine registries etc. are likely to have records made at the time of service and delay is less likely.

• Estimates from surveys, census, or UNWPP would already have delay considerations factored in.
Lecture 8: Assessing inequalities in registration completeness – Do your definitions align?

Renee Sorchik
Consultant, Civil Registration and Vital Statistics Expert
### Potential variables to investigate for death registration inequalities

#### For the decedent

- Date of birth
- Date of death
- Date of death registration
- Date of death notification
- Place of death (home vs facility vs community)
- Cause of death
- Age
- Geography (ideally as usual residence)
- Sex
- Ethnicity
- Occupation
- Education
- Wealth quintile
- Religious affiliation
- Migration or citizenship status
- Living status (alone?)

#### For the mother and father of deceased children

- Date of birth
- Age (at time of child’s birth and death)
- Geography (ideally as usual residence)
- Ethnicity
- Wealth quintile
- Education
- Occupation
- Religious affiliation
- Migration or citizenship status
Do the metadata of these align across sources?

If not, could new variables be created or is it possible to use a proxy?

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deaths registered by CRO</th>
<th>MOH recorded deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male/Female</td>
<td>Male, Female, Other</td>
</tr>
<tr>
<td>Age</td>
<td>Single years of age</td>
<td>Age groups, ends at 80+</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Two categories</td>
<td>Four categories</td>
</tr>
<tr>
<td>Residential address</td>
<td>Uses standard administrative boundaries, but captured as free text</td>
<td>Not available, but contains boundaries of health facilities, not aligned with admin boundaries</td>
</tr>
</tbody>
</table>
Do the metadata of these align across sources?

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deaths registered by CRO</th>
<th>MOH recorded deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male/Female</td>
<td>Male, Female, Other</td>
</tr>
<tr>
<td>Age</td>
<td>Single years of age</td>
<td>Age groups, ends at 80+</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Two categories</td>
<td>Four categories</td>
</tr>
<tr>
<td>Residential address</td>
<td>Uses standard administrative boundaries, but captured as free text</td>
<td>Not available, but contains boundaries of health facilities, not aligned with admin boundaries</td>
</tr>
</tbody>
</table>

- Male/Female align across sources, but CRO does not collect data for other sexes
- Check the % for ‘other’ – how high is it? How many are blank?
- Likely no way to recode or use a proxy, will need to focus assessment on M/F
Do the metadata of these align across sources?

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deaths registered by CRO</th>
<th>MOH recorded deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male/Female</td>
<td>Male, Female, Other</td>
</tr>
<tr>
<td>Age</td>
<td>Single years of age</td>
<td>Age groups, ends at 80+</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Two categories</td>
<td>Four categories</td>
</tr>
<tr>
<td>Residential address</td>
<td>Uses standard administrative boundaries, but captured as free text</td>
<td>Not available, but contains boundaries of health facilities, not aligned with admin boundaries</td>
</tr>
</tbody>
</table>

- Age: Will need to recode single years of age in 5-year age groups capped at 80+
Stillbirths and data quality

• MOH data often include stillbirths.

• While stillbirths are important to report on their own, to align with international standards, we do not include stillbirths in death registration completeness calculations.

• You will need to filter out these records in the MOH dataset

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Stillbirth</th>
<th>Infant death</th>
<th>Grand Total</th>
<th>% stillbirth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous</td>
<td>88</td>
<td>253</td>
<td>341</td>
<td>26%</td>
</tr>
<tr>
<td>Marzian</td>
<td>42</td>
<td>65</td>
<td>107</td>
<td>39%</td>
</tr>
<tr>
<td>Grand Total</td>
<td>130</td>
<td>318</td>
<td>448</td>
<td>29%</td>
</tr>
</tbody>
</table>
### Do the metadata of these align across sources?

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deaths registered by CRO</th>
<th>MOH recorded deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male/Female</td>
<td>Male, Female, Other</td>
</tr>
<tr>
<td>Age</td>
<td>Single years of age</td>
<td>Age groups, ends at 80+</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Two categories</td>
<td>Four categories</td>
</tr>
<tr>
<td>Residential address</td>
<td>Uses standard administrative boundaries, but captured as free text</td>
<td>Not available, but contains boundaries of health facilities, not aligned with admin boundaries</td>
</tr>
</tbody>
</table>

- How do each of the data sources define each ethnicity category?
- Do they agree on definitions? Are they close enough that there would be little discrepancy?
Using our Poplandia example

• Let’s say there are 2 major ethnicities, one minor one, and the rest are others.
  - Indigenous Poplandians
  - Marzian Poplandians
  - Mercurian Poplandians
  - All other ethnicities

• But our 2 data sources do not collect the same number of ethnicities
  - CRO collect 2 categories and MOH collects 4

**Civil registry ethnicity categories**
- All other ethnicities
- Indigenous Poplandians

**MOH ethnicity categories**
- All other ethnicities
- Indigenous Poplandians
- Mercurians
- Marzians
Solution

- Assuming the definitions of Indigenous population align or have little discrepancy between sources, recode the other 3 categories from MOH into an all “other” category.
- These two data sources can be used to analyze inequalities in registration of indigenous populations

*Policy recommendation: Work with the CRO to increase the number of ethnicity categories for the data it collects so you can analyze more ethnicities in the future.
Do the metadata of these align across sources?

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deaths registered by CRO</th>
<th>MOH recorded deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male/Female</td>
<td>Male, Female, Other</td>
</tr>
<tr>
<td>Age</td>
<td>Single years of age</td>
<td>Age groups, ends at 80+</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Two categories</td>
<td>Four categories</td>
</tr>
<tr>
<td>Residential address</td>
<td>Uses standard administrative boundaries, but captured as free text</td>
<td>Not available, but contains boundaries of health facilities, not aligned with admin boundaries</td>
</tr>
</tbody>
</table>

- Geography poses several challenges as the varying definitions of the data do not align
Problems with geography

CRO Database

<table>
<thead>
<tr>
<th>DateOfBirth</th>
<th>Sex</th>
<th>Residential Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/9/10</td>
<td>M</td>
<td>111 Elm st. Eastern, Poplandia</td>
</tr>
<tr>
<td>2/23/99</td>
<td>F</td>
<td>123 Main street #300, Eastern Poplandia</td>
</tr>
<tr>
<td>12/14/06</td>
<td>M</td>
<td>600 Centerville drive, Capitalville, Poplandia</td>
</tr>
<tr>
<td>6/28/76</td>
<td>M</td>
<td>144 Oak street, Western Poplandia</td>
</tr>
<tr>
<td>1/31/04</td>
<td>M</td>
<td>Pine health center</td>
</tr>
</tbody>
</table>

CRO captures blank open field that has to be recoded for use of decedent’s province/division/district.

MOH Database

<table>
<thead>
<tr>
<th>DateOfBirth</th>
<th>Sex</th>
<th>Facility</th>
<th>Health Divis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/9/10</td>
<td>M</td>
<td>Elm clinic</td>
<td>Eastern</td>
</tr>
<tr>
<td>2/23/99</td>
<td>F</td>
<td>Oak Clinic</td>
<td>Western</td>
</tr>
<tr>
<td>12/14/06</td>
<td>M</td>
<td>Capitalville H Western</td>
<td></td>
</tr>
<tr>
<td>6/28/76</td>
<td>M</td>
<td>Oak Clinic</td>
<td>Western</td>
</tr>
<tr>
<td>1/31/04</td>
<td>M</td>
<td>Pine health c</td>
<td>Southern</td>
</tr>
</tbody>
</table>

Not “apples to apples”

MOH does not have decedent’s address, instead has location of health facility that issued the death notice. Uses district and Division and is coded to health division boundaries which do not align with administrative boundaries.
Recode into district/division/Province,

• Might not be possible depending on technical capacity and quality of data field

<table>
<thead>
<tr>
<th>DateOfBirth</th>
<th>Sex</th>
<th>Residential Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/9/10 M</td>
<td>M</td>
<td>111 Elm st. Eastern, Poplandia</td>
</tr>
<tr>
<td>2/23/99 F</td>
<td>F</td>
<td>123 Main street #300, Eastern Poplandia</td>
</tr>
<tr>
<td>12/14/06 M</td>
<td>M</td>
<td>600 Centerville drive, Capitalville, Poplandia</td>
</tr>
<tr>
<td>6/28/76 M</td>
<td>M</td>
<td>144 Oak street, Western Poplandia</td>
</tr>
<tr>
<td>1/31/04 M</td>
<td>M</td>
<td>Pine health center</td>
</tr>
</tbody>
</table>

If MOH has GPS coordinates or a database that can map health boundaries to administrative boundaries, recode records to align with district/division/province administrative boundaries.

• May not be possible if GPS points are only for clinics and don’t include surrounding areas or if no other means to map the two boundaries to one another exist
• Investigate whether proxies can be used, particularly for urban districts or areas
• Your investigations and partnerships with data providers will determine what can and cannot be done and will help inform the data to be used for the inequality assessment.

• Be sure to make note of what is not being collected or which sources are not aligned by definition. This is important information to include in a report and for future interventions.
Exercise 3: Assessing usability of data by analyzing definitions
Lecture 9: You now know which variables align, how to assess the quality of the data?

Renee Sorchik
Consultant, Civil Registration and Vital Statistics Expert
### How to assess data quality of variables

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DeathDate</td>
<td>RegistrationDate</td>
<td>BirthDate</td>
<td>Age</td>
</tr>
<tr>
<td>2</td>
<td>12/23/12</td>
<td>1800-01-01 12:00:00</td>
<td>9/25/62</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>11/1/12</td>
<td>3/15/03</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>4</td>
<td>2/8/19</td>
<td>3/1/22</td>
<td>5/25/65</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>11/11/17</td>
<td>2/28/22</td>
<td>9/28/49</td>
<td>68</td>
</tr>
<tr>
<td>6</td>
<td>12/7/20</td>
<td>2/25/22</td>
<td>12/30/76</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>5/2/16</td>
<td>2/25/22</td>
<td>7/26/38</td>
<td>77</td>
</tr>
<tr>
<td>8</td>
<td>9/23/18</td>
<td>2/23/22</td>
<td>10/24/50</td>
<td>67</td>
</tr>
</tbody>
</table>

**Red Circles:**
- Red circle on 3/15/2103: Indicates an error in the data.
- Red circles around dates: Highlighting dates that need further investigation.

**Highlighted Data:**
- Column A (DeathDate) and Column C (RegistrationDate) are highlighted to indicate areas of concern.

**Analysis:**
- Review dates for logical consistency and accuracy.
- Investigate NULL values for possible errors or missing data.
- Ensure dates are in proper format and reflect real-world data.

**Recommendation:**
- Perform data validation tests for each variable.
- Use data cleaning tools to identify and correct errors.
- Develop a data quality assessment plan to ensure ongoing integrity.

---

[Image of spreadsheet with highlighted cells]

**Legend:**
- **Red Circle:** Error indication.
- **Highlighted:** Data points for review.

**Further Analysis:**
- Conduct a statistical analysis of data distribution.
- Evaluate missing data for patterns that may indicate data quality issues.

**Conclusion:**
- Data quality is crucial for effective data-driven decisions.
- Continuous monitoring and improvement are necessary to maintain high data integrity.
Understanding data quality

• If there are 100,000 records in the registered death data set, you cannot fix them all in a timely manner.

• Underlines the importance of:
  • Entering the data correctly and completely from day one so they are usable for policy.
  • Having a regular data cleaning policy or “plan”

• Important to understand the extent of errors (how many records are “unusable”? Fixable?
  • 100 records out of 100,000 has little effect. 10,000 out of 100,000 starts to impact your analyses. 30,000 out of 100,000 might render them unusable.
Understanding data quality

• How many records with errors are “acceptable”?  
• This depends on the variable and type of data.  
• Generally speaking, 10% or less of records with errors will not void your analysis  
• If you are looking at sex with just M/F, 10% of missing records is unlikely to obscure sex differentials in registration.  
• However, if you are analyzing an indigenous hill tribe that constitutes 5% or less of the population, if even half of the missing 10% of records corresponds to this ethnicity, your analysis will not be accurate.
Questions to ask

• How many duplicate records are there? Is there a way to find and remove them?
• How many missing or erroneous records are there? How many can be fixed?
• What percent of the general population does my sub-population comprise?
• Based on what I know, or what the data collector may know, what is the likelihood that this sub-population corresponds to my missing records?
• If all the missing records corresponded to my population, how would this change the outcome of analysis? My conclusion or policy recommendations?
  • What if half or even 10% of the missing records were my population of interest?
  • If possible, you can include margins or errors based on the number of missing or erroneous records
• Do other sources (quantitative or qualitative) support the findings even with imperfect data? (more on this later)
• Remember, it’s OK to say the data was inconclusive. A recommendation for further and better data collection is a step in the right direction!
Where possible, you can derive or edit missing variables

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeathDate</td>
<td>RegistrationDate</td>
<td>BirthDate</td>
<td>Age</td>
</tr>
<tr>
<td>12/23/12</td>
<td>1/800-01-01 12:00:00</td>
<td>9/25/62</td>
<td>50</td>
</tr>
<tr>
<td>11/1/12</td>
<td>3/15/03</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>2/8/19</td>
<td>3/1/22</td>
<td>5/25/65</td>
<td>53</td>
</tr>
<tr>
<td>11/11/17</td>
<td>2/28/22</td>
<td>9/28/49</td>
<td>68</td>
</tr>
<tr>
<td>12/7/20</td>
<td>2/25/22</td>
<td>12/30/76</td>
<td>43</td>
</tr>
<tr>
<td>5/2/16</td>
<td>2/25/22</td>
<td>7/26/76</td>
<td>77</td>
</tr>
<tr>
<td>8/23/18</td>
<td>2/23/22</td>
<td>10/24/50</td>
<td>67</td>
</tr>
</tbody>
</table>
Report both adjusted and unadjusted rates

• If you do adjust your data, it’s important to mention the adjustment technique in your methodology, and how it might affect your conclusions.

• It’s important to report both the original registered counts of disaggregated deaths (by age, sex, ethnicity etc.) as well as adjusted numbers

• Be transparent about how numbers were adjusted
  • Did you use percent distribution from vital statistics?
  • Did you adjust by completeness by age group? By sex?
  • Was another imputation method used?
Important to look at the consistency of your data over time to check data quality

• How has number of (registered) deaths fluctuated over time?
• Are there patterns in the data we would expect to see?
• Important to compare your registered death data to your estimated death data over time and look for similar patterns
• If the ‘big picture’ is consistent (at the national level), you can feel more comfortable with disaggregating your data
How has number of (registered) deaths fluctuated over time?

Registered deaths in Poplandia (including delayed registration)

Average per year: ~663,000
Are there patterns in the data we would expect to see?

Registered deaths in Poplandia (including delayed registration)
How do we explain this?

- Are yearly deaths going down (due to a changing population structure or improving health situation)?
- Or is registration delayed?
- What about the impact of COVID???
- We expect to see ~660,000 deaths a year (average of 10 year period)
What is happening over time?

Timeliness of registration improved after 2010-2011 but TBD if it’s still improving in 2017 onwards.
What about when we disaggregate?

Is the drop post 2016 for Indigenous persons real? Or did something happen with our data?
Snapshot of 2018: Comparing registered deaths to “all” deaths

- Two sources share similar patterns. Is that a sign of good data quality???
- Late registration is low compared to timely registration, but will it increase post-COVID??
- We want to know, who are the missing ~150,000 deaths who are not registered???
What about birth data quality?

• The points covered above will largely cover most analyses of your death data

• They would also apply to looking at recorded births such as birth notifications and neonatal vaccine administration

• If you are using UNWPP data for your denominator, the lecture series by Tom Moultrie covers data quality considerations in detail

• However, there are some intricacies to consider when trying to look back at births retrospectively (using MOE data)
Estimating births through school enrollment data

• Ministries of Education (MOE) often has school enrollment data which you can use to estimate the number of births that occurred (5-6) years ago.

• While this gives estimates that are 5+ years old, this data is useful because it is often available in a more disaggregated form than other types of data.

• However, if migration of young children is substantial and you do not have good estimates for migration data, this methodology may not be appropriate for your country.

• Because it’s retrospective, you will need to factor in children lost to infant and child mortality in the interim years.

• The methodology of reverse survival is outside the time and scope of this workshop, but the reliability of infant and child mortality data along with the migration context and capacity for analysis should be considered when using school enrollment data.
Considerations for school enrollment data

• What % of children are in school in your country? Can you adjust for those not in school by sex, ethnicity, location etc.?

• For children that are in school, how complete is the school enrollment data set?
  • Does it only cover urban areas? Certain provinces? Government schools?

• How many children attend kindergarten? Is Kindergarten compulsory? Do many families choose private kindergarten?

• Is it better to analyze those in grade 1?
Considerations for school enrollment data

Let’s say we look at year one data. We expect most student to be age 6-7, so born in 2014 or 2015.

However, there are a few born later and some whose birth dates are much earlier.

Are these errors? Or is something else going on?
Considerations for school enrollment data

- We are not interested in level of academic achievement, thus it would be better to query the MOE database for all students born in 2014, 2015, 2016 etc. regardless of their grade level.

- However, if there is a variable for disability status, and the CRO also can be disaggregated by CRO or the data can be linked via unique ID, this is important information for measuring the completeness of registration for ALL births.
Considerations for school enrollment data

• You may not have data to flag students with disabilities etc. In this case, an in-depth discussion with MOE would be needed to inform about quality of data.

• We also need to examine at year-to-year variations in enrollment to assess data quality, similar to what we discussed above for deaths.

• We want to ensure coverage is good and factor in the percent of kids believed to be in school.

• And as discussed above, make sure our definitions align.
Exercise 4: Assessing quality of data sources and determining usability