



**United Nations**

Department of  
Economic and  
Social Affairs

# *Population concepts used in the Demographic Yearbook System: Improvement of vital statistics metadata*

(P. Gerland, Population Division, DESA, United Nations)

United Nations Expert Group Meeting on the United Nations Demographic Yearbook System



Session III: Tuesday 10 Nov. 2020



**United  
Nations**

Department of  
Economic and  
Social Affairs

# DYB Questionnaire on Vital Statistics

Metadata questionnaire,  
DYB database and analytical use for..

- [UN World Population Prospects](#)  
(235 countries/areas for 1950 -2020 annual time series of population estimates and vital rates by age and sex)
- [SDG global monitoring of adolescent fertility rates](#)
- [UN IGME child mortality estimation group](#)
- [UN MMEIG group on maternal mortality](#)

# UNSD DYB: vital statistics

43 tabulations requested to NSOs:

## Fertility (13)

- **Total live births by sex, and by age of mother** and urban/rural residence

## General mortality (9)

- **Total deaths by sex, and by age and sex + life tables** (abridged or complete age)

## Infant deaths (3)

- Total infant deaths by sex of infant and urban/rural residence of mother
- **Infant deaths by age and sex (since 1970)**

## Foetal deaths (3)

- *Late foetal deaths by sex of foetus and urban/rural residence of woman*
- *Foetal deaths by gestational age*

+ Abortions (3), Marriage (6), Divorce (5)

# UNSD DYB: vital events by age and sex

- DYB series start in 1948
- Most available series only by abridged age groups, by single age only since 2007 for subset of countries
- Urban series only available for selected countries x years and abridged age groups: e.g., mortality data
  - Africa: Egypt (1971-)
  - Asia: Kazakhstan (1989-), Kyrgyzstan (1987-), Maldives (1984-)
  - LAC: Chile (1983-), Colombia (1995-), Cuba (1979-), Ecuador (1979-), El Salvador (1972-), Panama (1966-)
  - Europe: Bulgaria (1966-), Poland (1965-), Romania (1966-)
- Note: the urban/rural definitions for each country needs to be validated for each census or time period they apply.

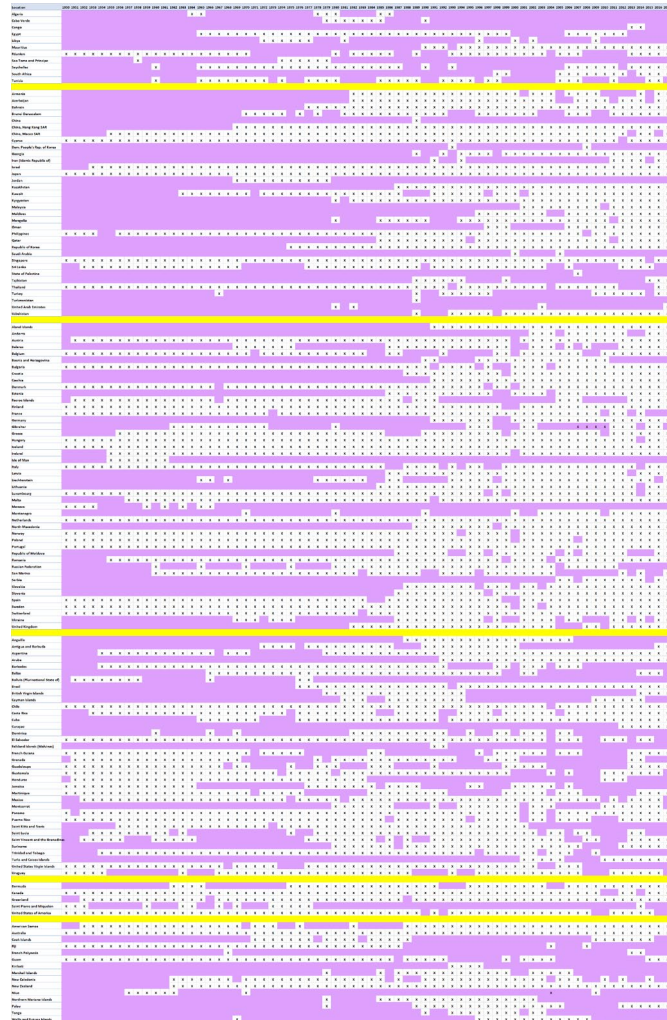
# Geographic and temporal coverage

- Many countries/areas have various gaps in time series since the 1950s, and age classifications lack harmonized reporting standards as used by WHO mortality database\*
- Time series for countries/areas typically only reflect the current territorial situation at time of reporting to DYB
- Due to various historical territorial changes (e.g., countries independence or merging), **time series for these new locations are incomplete for earlier periods** for these respective areas (e.g., former USSR, Yugoslavia, etc.)
- Availability of data, especially disaggregated by age and sex, is often lacking for various territorial dependencies and smaller populations.



# 1950-2017 data gaps in VR reporting to DYB\* for locations with $\geq 90\%$ completeness in 2019

Live births  
by age of  
mother  
for 147  
locations



Africa

Asia

Europe

Latin America

North America

Oceania

Deaths by  
age and sex  
(abridged)  
for 121  
locations



Africa

Asia

Europe

Latin America

North America

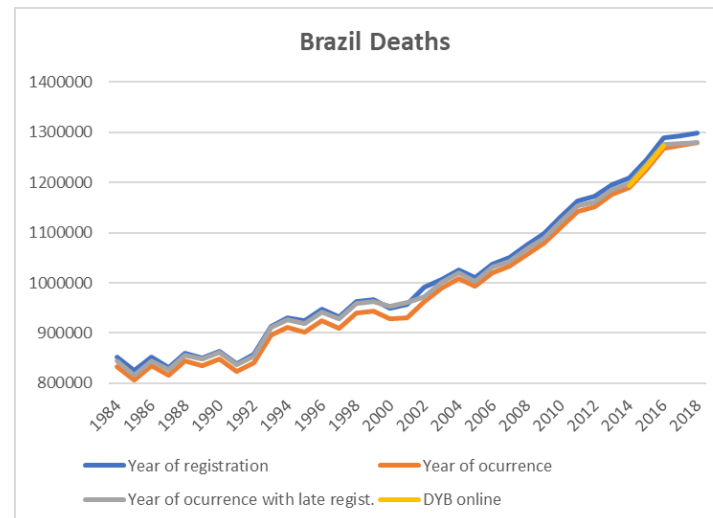
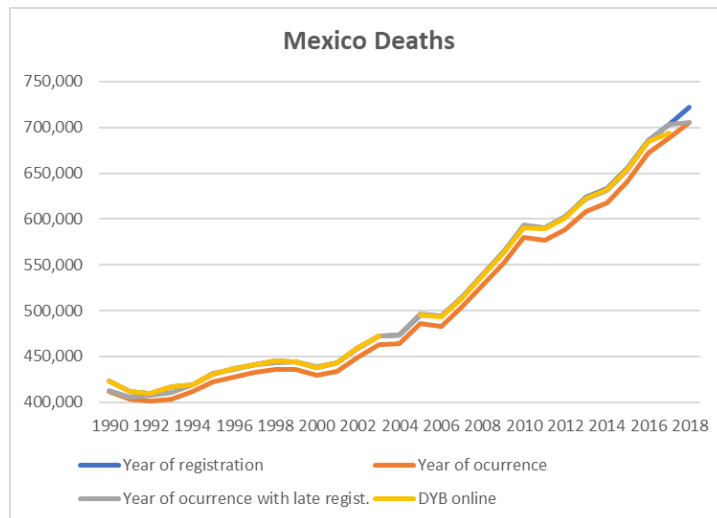
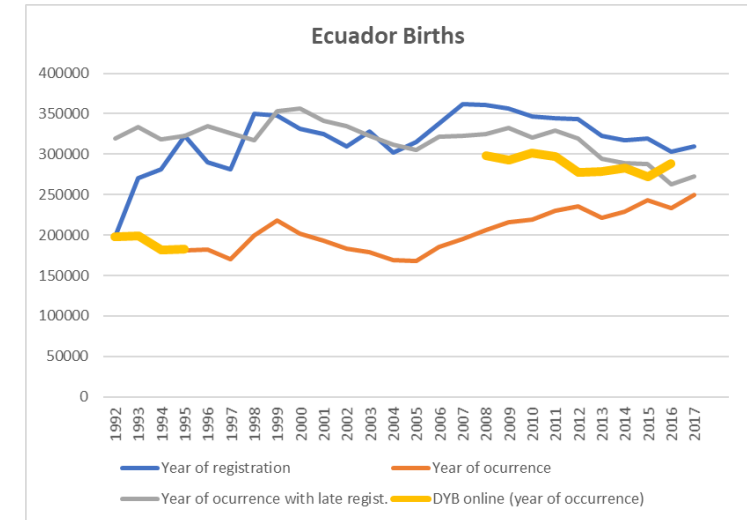
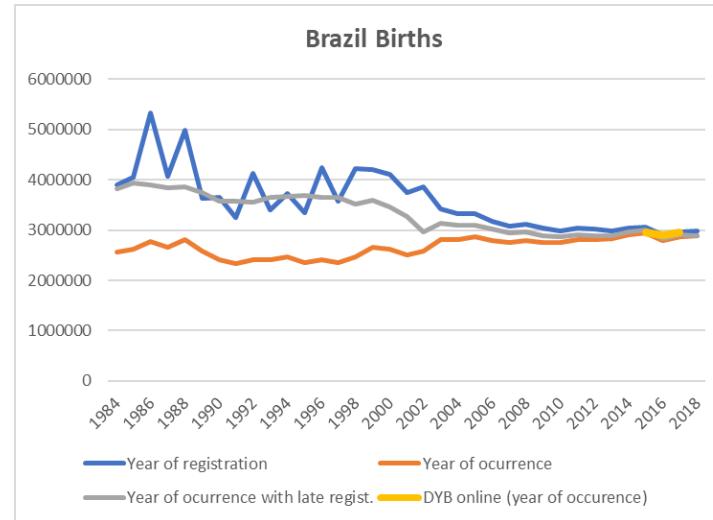
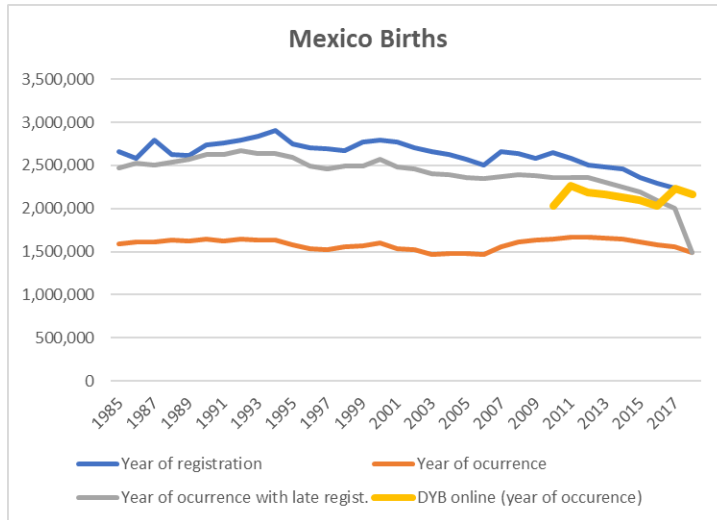
Oceania

# Quality of VS obtained from CR (1)

I. Basis of tabulation	Live births	Deaths	Infant deaths	Late foetal deaths	Marriages	Divorces
Present basis of tabulation:						
- date of occurrence						
- date of registration						
How long has the present basis of tabulation been used?						

- The Year of occurrence/registration is coded with the DYB data allowing more easily various analytical use.
- In a majority of instances, most countries/areas are reporting by year of occurrence (as per P&R recommendations), but a subset of countries have only been reporting by year of registration.
- Current concepts do not reflect whether the vital events are for the *de-facto*, *de-jure* or usual resident population only. In particular, whether data are only for nationals, or include also foreigners (e.g., special situation of GCCs or various territorial enclaves/islands and benefit to collect and use both numerators and denominators for native and foreign-born population). Idem for the inclusion/exclusion of special population groups like refugees, asylum-seekers and stateless persons.
- How territorial coverage issues get treated and reflected in the time series have major implications for analyzing change over time. Use of footnotes can be helpful, but key issues/limitations should be codified through a standard typology for easier analytical purposes (e.g., geographic coverage, population coverage, methodological issues).

# DYB coding (year of occurrence vs. registration), and late registrations, especially for live births



Currently, late registration do not explicitly appear in DYB, only indirectly and if revised series are submitted for more recent reporting years



# Quality of VS obtained from CR (2)

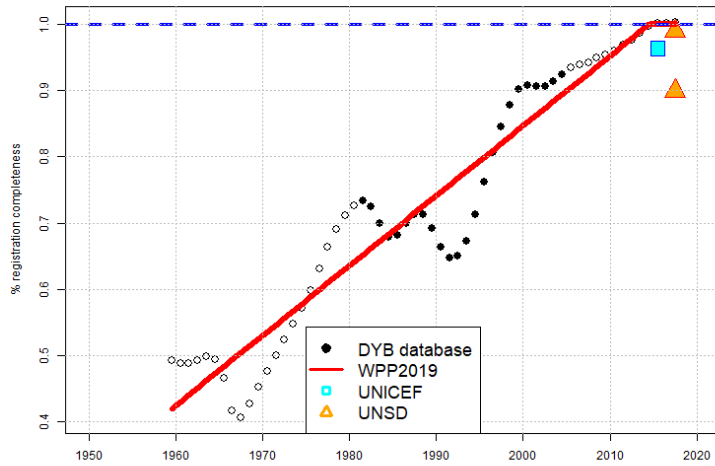
II. Estimated completeness of registration	Live births	Deaths	Infant deaths	Late foetal deaths	Marriages	Divorces
100 per cent						
90 - 99 per cent						
80 - 89 per cent						
70 - 79 per cent						
60 - 69 per cent						
50 - 59 per cent						
Under 50 per cent						
Please specify:						
(a) Year(s) to which completeness estimate refers						
(b) Basis of completeness estimate						
-Demographic analysis						
-Dual record check						
-Questions in population census						
-Questions in sample surveys						
-Other (specify)						
-No evaluation						
Please include any reports describing completeness of registration and methods used in arriving at estimated completeness:						

# Estimated completeness of registration

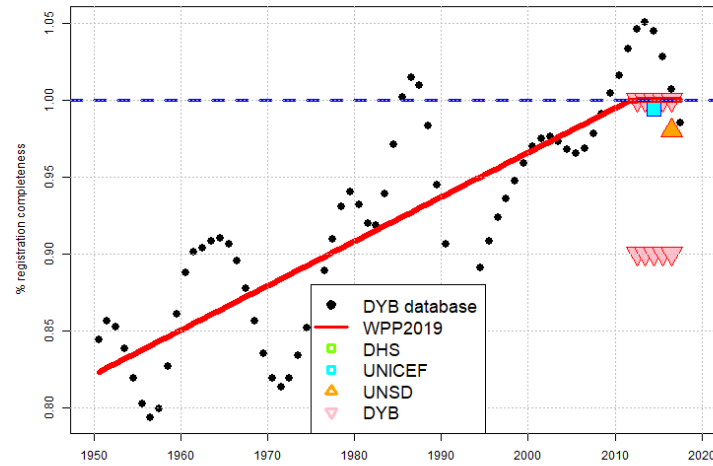
- Recent revision of the set of categories for monitoring of SDG indicator 17.19.2b\* (used to be under 50%, 50-74%, 75-89%, 90% or more)
- Collect and report the percentage rather than only a category to measure progress over time
- Pending challenges:
  - Frequent lack of comprehensive metadata documentation for the various set of vital events, especially in respect to the method(s) used to determine the completeness estimate ;
  - Metadata are summarized in the DYB publications through the reliability encoding ( $\geq 90\%$ ), and the basis for the completeness estimates eventually included in textual notes in the hardcopy / PDF report – but not available in machine readable codified format for analytical purposes ;
  - Information is only presented for the most recent year in each edition of the DYB. Lack comprehensive national time series of completeness estimates based on consistent methodology, type of data sources and assumptions are currently available.

# Trend in estimated completeness of registration

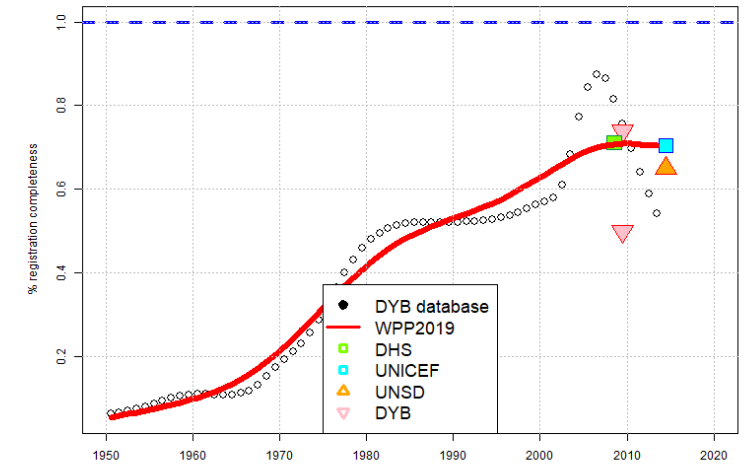
Brazil : Completeness of Live Births Registration



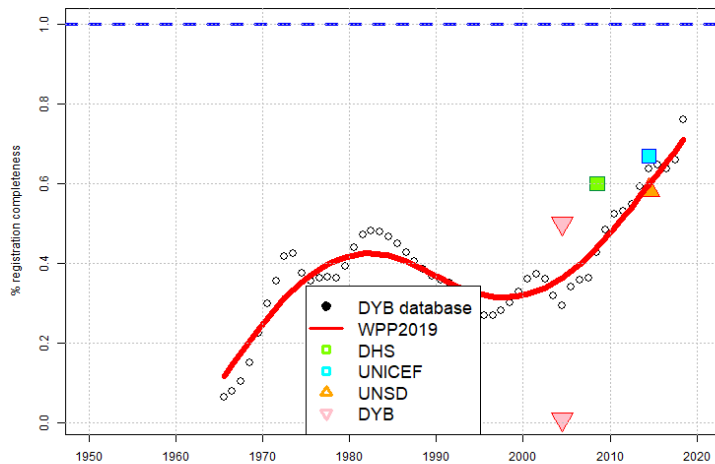
Egypt : Completeness of Live Births Registration



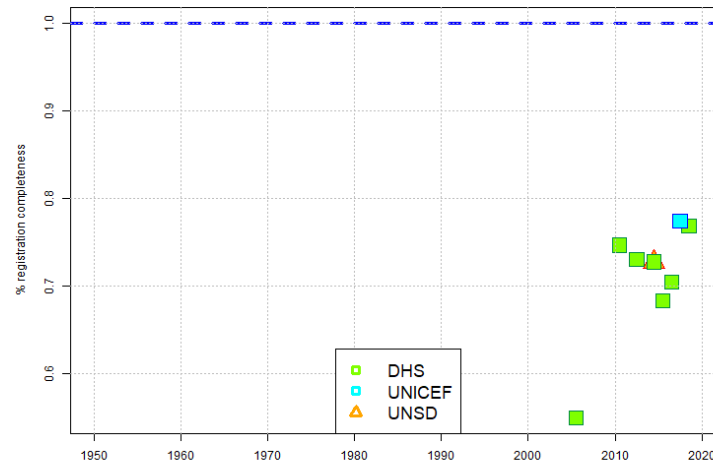
Ghana : Completeness of Live Births Registration



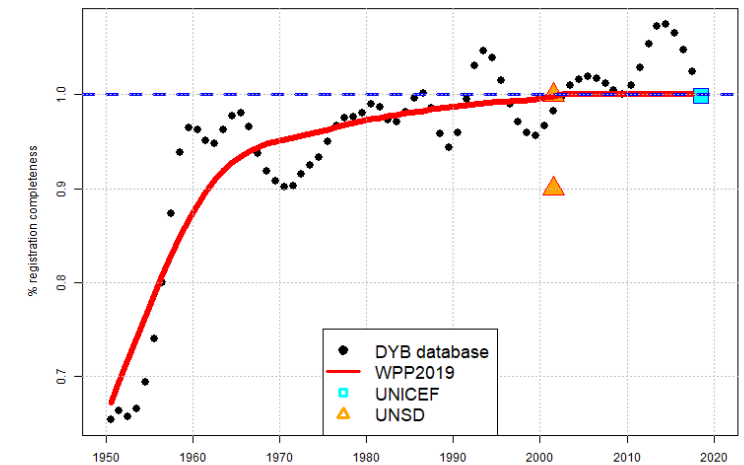
Kenya : Completeness of Live Births Registration



Senegal : Completeness of Live Births Registration



Tunisia : Completeness of Live Births Registration



# Quality of VS obtained from other sources

Basis of vital statistics estimates is	Live births	Deaths	Infant deaths	Late foetal deaths	Marriages	Divorces
- Population censuses (date)						
- Sample surveys						
- Population registers						
- Dual record systems						
- Other (specify)						
Please include any reports describing the methods used for estimates of vital statistics based on other sources than civil registration:						

By default, and historically, DYB has focused almost only on vital events from civil registration, and vital rates that have been published in the DYB reports are not stored in the DYB database or disseminated for analytical purposes through UNdata. End-users must (re-)compute their own rates.

The DYB collection, and coding of vital events and vital rates from censuses or some surveys has only started in the last 1-2 decades, and the coverage has been partial. But the DYB has not been collecting systematically vital statistics from surveys, and for countries with sample registration systems (e.g., India, Bangladesh) or annual population change surveys (e.g., China, Viet Nam) the reporting of vital rates through the DYB is lacking.

# Additional metadata for selected tables

Tables 14a, 14b, 15a and 15b - Life tables	
Do the life tables refer to de Facto population or de Jure population?	
Was any method used to smoothen the life table? Which one?	
Was any specific method used to close the life table at older ages (e.g., Gompertz, Makeham, etc.)? Which one?	
If any model life table or relational model was used to derive the life table (e.g., Coale-Demeny West, UN South Asian pattern), what model was used?	
Please mark with an "X" the applicable options	x
What source of data was used to compute the life tables?	
a) Unadjusted vital registration deaths	
b) Adjusted vital registration deaths	
c) Information on deaths from census	
d) Life expectancy at birth	
e) Under-five mortality	
f) Infant mortality	

- Unclear what data are used for numerator and denominator to compute mortality rates.
- To date this type of metadata information is not encoded in the data model of the DYB.
- When specific life tables are queried from the DYB database, there are no flags or technical documentation linked to the respective data records to know about the various limitations or adjustments made.
- With an increasing number of deaths occurring at older ages, the assessment, quality and reliability of mortality rates at older ages become ever more critical.



# From count data to vital rates

- Rates and ratios based on (1) same source of data (e.g., census, survey or registers) or (2) different data sources (e.g., vital registration and census population or estimates):
- If based on different data sources, potential issues with numerator vs. denominators: problem of matching statistical concepts and measurements, potential biases with differential reporting/selection between num./denom. For example:
  - Published data only available by year of registration instead of **year of occurrence**
  - Population estimates used as denominator: intercensal vs. post-censal estimates/projections, and **revised intercensal series after new census** + near-extinct generation population estimates at very old ages based on death registration (see Human Mortality Database method protocol).
  - For subnational, including urban: published data only available by place of occurrence/registration instead of **place of (usual) residence**
  - Additional disaggregation issues for various SDG characteristics (e.g., mortality by occupation in Australia or by ethnicity in New Zealand: different ways in which occupation/ethnicity is recorded on death certificates and on census forms).

# Vital registration as data source

## Benefits:

- Universal coverage (in principle)
- Continuity
- Can provide data for small areas

## Limitations :

- Potential problems with geographical coverage, completeness of reporting and data quality
- Year of registration vs. Year of occurrence and revisions for Late registration
- Place of registration vs. Place of residence vs. usual residents / native / foreigners
- Common lack of geographic disaggregation or only total counts, often lack of details by age & sex – especially for smaller locations/populations, and territorial dependencies;
- Common lack of additional characteristics for further disaggregation by socio-demographic, economic or cultural characteristics (currently only urban/rural collected by DYB for some tables) – but potential issues of matching concepts for numerators/denominators to compute vital rates.

# Recommendations about VS data

1. Greater interest/focus on more disaggregated age distribution (by single year of age) and use of annual time series + more harmonized age classifications (>1100 age groups since 1948 in DYB tables)
2. Report to UNSD DYB vital events by single age distribution for earlier years (i.e., 1950-2007) upon data availability
3. Collect and disseminate in database format vital rates, including from SRS and large (annual) nationally representative household sample surveys (especially used for SDG reporting on fertility and mortality for about 115 countries/areas)
4. Collect more comprehensively tabulations from recent births and household deaths collected in censuses (including maternal deaths)
5. Collect vital events (and population) for native and foreign -born population as per relevant national circumstances
6. Fill-in data gaps in time series upon national data availability
7. Idem for the computation of vital rates and life tables, report the revision year of the estimates and projections used for the population denominator. Use as much as possible **revised intercensal population estimates after each new census**.
8. Resubmit to UNSD DYB new series of revised life tables using updated intercensal population estimates as denominator to compute vital rates.

# Recommendations about concepts

1. Expand set of statistical concepts considered for the population covered by vital statistics: **de-facto, de-jure or usual resident population only**
2. Clarify the inclusion/exclusion status of special population groups like refugees, asylum-seekers and stateless persons
3. Codify DYB footnotes using a **standard typology for easier analytical purposes** (e.g., geographic coverage, population coverage, methodological issues)
4. Update the DYB coding and data submission process to deal with **late registrations** , especially for live births

# Recommendations about concepts

Tabulation on late foetal deaths: update the DYB VS questionnaire to the new ICD 11 definition for the collection of stillbirth data as per recommendations of the United Nations Inter -agency Group for Child Mortality Estimation (UN-IGME)\*, and endorsed by the WHO Mother and Newborn Information for Tracking Outcomes and Results (MONITOR) technical advisory group\*

- UN-IGME suggests to use the term “stillbirths” instead of “ foetal deaths” and to use stillbirths at 28 weeks or more of gestational age for international comparison and to collect stillbirths by gestational age.
- UN-IGME recommends using a stillbirth definition that uses the gestational age as the single criteria. Gestational age is preferred to birthweight and/or length criteria as it is a better predictor of maturity and viability, and is the most commonly available criteria across data sources.



# Recommendations about metadata

1. Metadata should be made publicly available in **machine readable format for analytical use**
2. Time series of such information should be made available jointly with the respective data tabulations
3. Additional follow -up may be required to clarify or fill-in missing metadata
4. **Method(s) and data used to determine the respective completeness estimate must be documented**
5. For completeness reporting, if based on analytical method and the use of population estimates and projections , **report the revision year of the estimates and projections** .
6. Collect and report the **percentage of completeness** rather than only a category to measure progress over time
7. Consolidate time series of completeness estimates based on consistent methodology, type of data sources and assumptions

# Recommendations about life tables

1. Encode some of the metadata information for life tables in the DYB data model
2. Improve the documentation and encoding of the **data used for numerator and denominator to compute mortality rates**
3. Add flags or codified footnotes to the respective data records to know about the various limitations or adjustments made
4. Improve data validation on the basic consistency of the various life table elements and relationships between indicators, and over age and sex