United Nations | Department of Economic and Social Affairs

Insights from the World Population Prospects: Estimation methodologies

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

Regional Webinar on
"National and International Population Estimates in the Arab Region: Challenges and Analytical Strategies"

26th February 2025, Beirut and New York

Outline

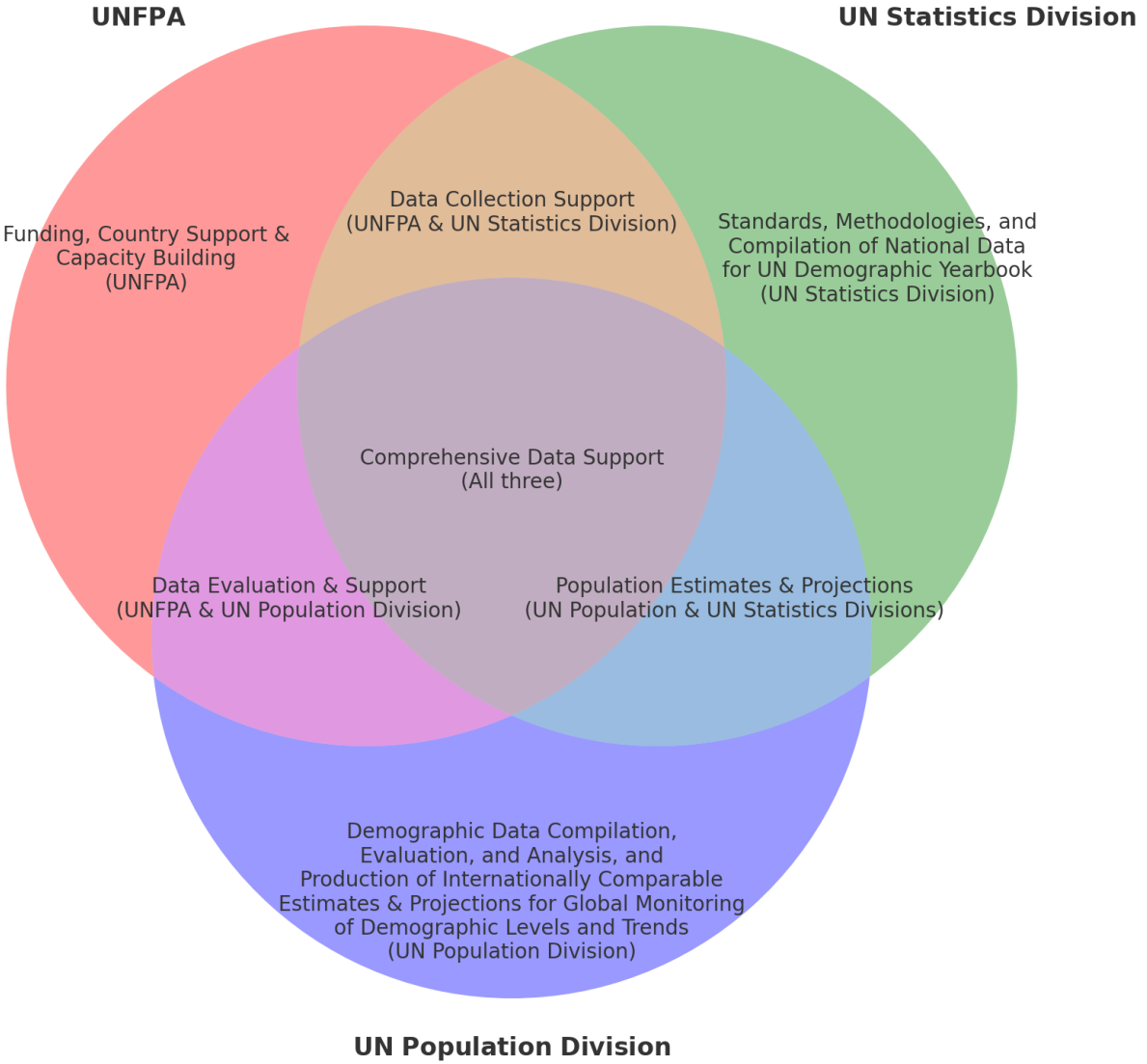
1. Typology of population data systems
2. World Population Prospects: methodology and data for population and demographic estimates and projections
3. Country data sources and availability for population estimates
4. WPP analytical workflow, estimation strategy and statistical modelling for components of population change (fertility, mortality, and international net migration)
5. Method protocol for the evaluation of census population data by age and sex
6. Global perspective on challenges and opportunities for internationally comparable estimates
7. Annexes

United Nations | Department of Economic and Social Affairs

Part 1

Typology of population data systems

Population data in the United Nations



Why population data systems matter?

- Crucial for informed decision-making and policy formulation.
- Population data informs resource allocation, service delivery, and policy development.
- Accurate and timely data is essential for monitoring progress towards the Sustainable Development Goals (SDGs).
- Modern data systems need to be resilient to adapt to evolving data demands and potential disruptions.

Fundamental law of demographic dynamics

- **Population change depends on components of demographic change and a balancing equation for demographic accounting**
- This is represented by the following relationship:

- Population estimates (and projections) are derived by combining population, natural change (births minus deaths) and net migration (immigrants minus emigrants) either at the aggregate or individual level.

Introduction about population estimates

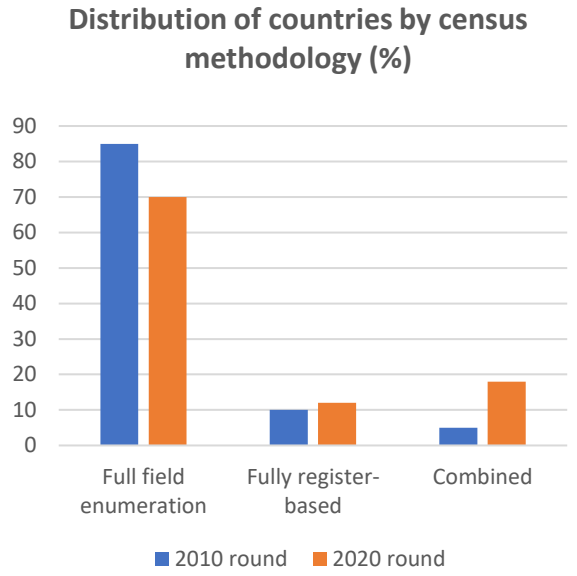
How do we know the current number, characteristics and distribution of people in each country or area?

… it depends on the **type of population data system** available in a given country …

Typology of Population Data Systems

Three main groups based on their primary data source:

1. Population counts (stocks)*
    1. **Full field-based Census systems**: relying on periodic complete field enumeration of the population.
    2. **Register-based systems**: relying on continuously updated population registers and records of vital events (births, deaths, migration) and (a) other administrative data linked at the individual level, or (b) existing sample surveys or integrated administrative sources and existing sample surveys
    3. **Combined systems**: combining data collected from full field enumeration or one or more surveys with administrative data sources or statistical registers.



2. Vital statistics (flows) for intercensal and post-censal periods
    1. Vital registration of births and deaths (full or partial completeness)
    2. Fertility and mortality statistics from censuses and surveys
    3. International and/or internal migration statistics from administrative data

(*) Source: UNSD surveys on 2010 round PHCs (126 responses) and 2020 round PHCs (158 responses)

Overview of Country Data Sources for Population Estimates

Data Sources:

- Population censuses: most countries conduct censuses, but coverage, frequency and quality can differ.
- Civil registration systems (CRS) and vital statistics: birth and death registration systems vary in completeness (from <50% to >90% globally) and accuracy.
- Migration data: often the weakest component
- Population registers: some countries have well-established registers, while others are developing them – but hard to know precisely due to lack of metadata.
- Administrative data (often only for selected subsets of population): Increasingly used, but access and quality can be challenging, and many integration challenges when unique national IDs are unavailable.
- Sample surveys (when other sources are unavailable): provide valuable data, but may have limitations in coverage and frequency.

Global Perspective:

- Differences in data availability and quality across regions.
- Censuses are conducted less frequently in lower-income countries due to costs

Intercensal and Post-Censal Population Estimates

Intercensal Estimates:

- Use data from the two most recent censuses combined with birth, death, and migration.
- Helps to maintain up-to-date population estimates between censuses, and to ensure the internal consistency between demographic components of change (especially migrations) and populations
- Best practice: after each new census, a new set of intercensal estimates should always be recomputed, disseminated and used to recomputed intercensal vital rates and life tables (European Commission 2003, Statistics Canada 2016, Spoorenberg 2020, United Nations 2024)

Post-Censal Estimates:

- Adjustments are made after a census to correct inaccuracies and update for population dynamics
- Post-censal estimates use the demographic balancing equation to project population figures
- Timeliness and frequency of updates: produced shortly after last census or updated annually or more frequently (monthly or near-real time daily updates) to incorporate extra / new data on vital statistics and migrations
- Timeliness and type of input data used for each demographic components
- Projection method: cohort-component (aggregate level) or microsimulation (individual level: Austria, Canada, Mexico based on UNECE Database on Population Projections Metadata)

United Nations | Department of Economic and Social Affairs

Part 2

World Population Prospects: methodology and data for population and demographic estimates and projections

The need for consistent population estimates
(for global monitoring of levels and trends by the United Nations)

World Population Prospects (WPP) 2024 – biennial update

- Comprehensive, standardized demographic dataset for all countries/areas Internally consistent prospective annual **population reconstruction** from 1950 to 2023 using all available data (i.e. population by age and sex, mortality, fertility and migration) using the cohort component method
- Population estimates are used to check consistency of WPP estimates (…and WPP allow to detect consistency problems in population estimates)

>> Important to obtain **accurate/consistent** national estimates of population size and composition by age and sex

  - WPP used as “denominator” in several SDG indicators

For further details, see United Nations (2024)

Aims of the UN World Population Prospects (WPP)

- What is the WPP?
  - Official UN estimates and projections of populations and key demographic indicators (fertility, mortality, migration), all by age and sex.
  - Covers 237 countries and areas (and more than 300 aggregates)
  - Annual estimates from 1950 to 2023, projections until 2100.
  - Fully updated every two years
- Significance
  - Provides a standardized and consistent set of demographic data for all countries spanning 150 years.
  - Used to inform policy decisions, assess future needs, and monitor progress towards Sustainable Development Goals (SDGs), incl. about ¼ of the SDG indicators.

Consistent population estimates

Several methods exist to produce annual population estimates

- Basic requirements:
  - An initial/base population count (recent, evaluated and if necessary adjusted, see P&R for Population and Housing Censuses, Revision 4);
  - A method of time adjustment to update the initial population count

Gold, silver and bronze standard methods

1. Continuously updated population registers
2. Regularly updated combined population systems
3. Cohort-component method applied to census data

→ Population data consistent by age, sex and along cohorts

Population balance (demographic accounting)

$$\mathbf{P_{t+n}} = \mathbf{P_t} + B_{t,t+n} - D_{t,t+n} + I_{t,t+n} - E_{t,t+n}$$

$$P_{t+n} - P_t = B_{t,t+n} - D_{t,t+n} + I_{t,t+n} - E_{t,t+n}$$

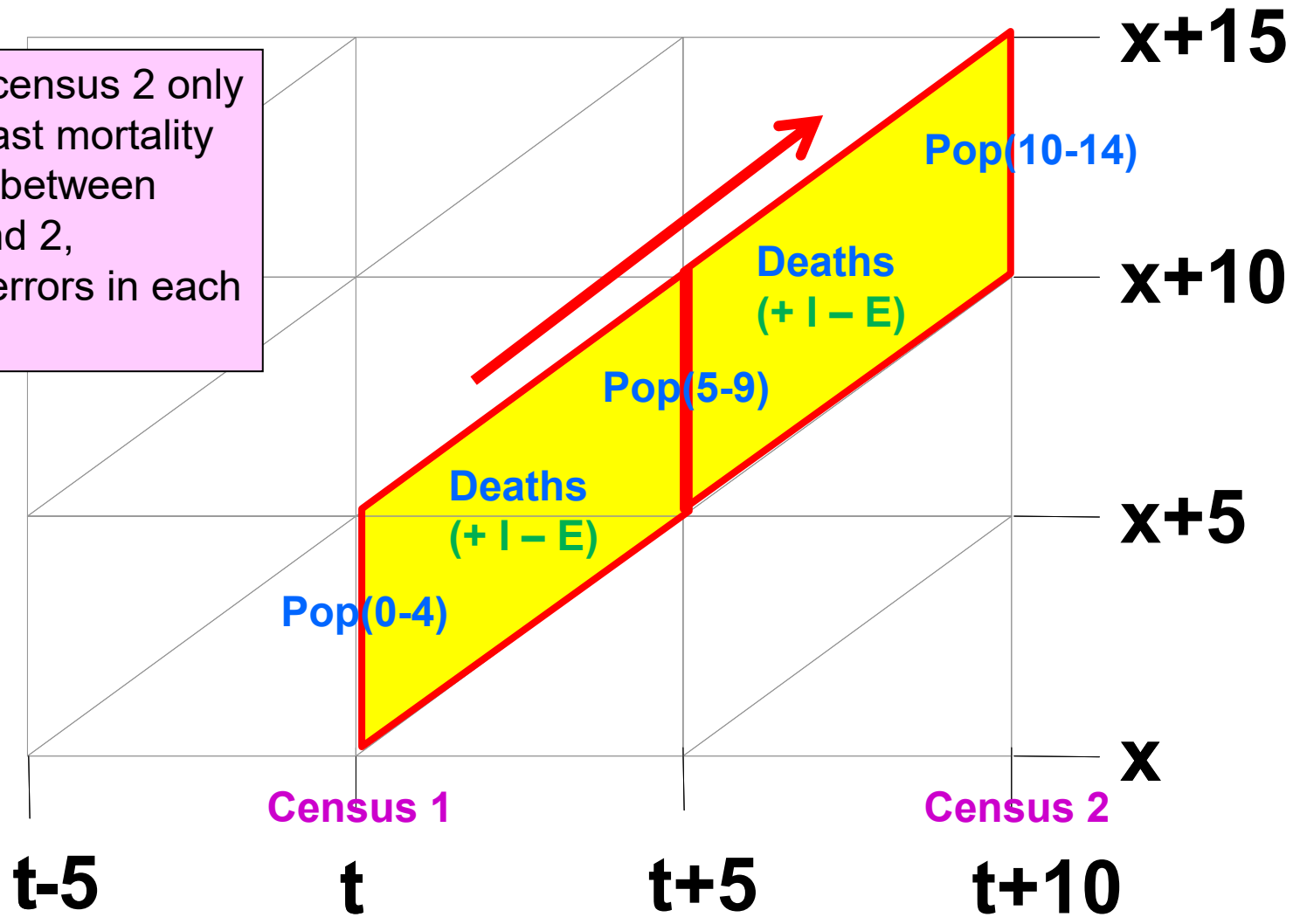
$\mathbf{P_t}$ is the population at time t

$\mathbf{B_{t,t+n}}$ and $\mathbf{D_{t,t+n}}$ are number of births and deaths occurring between t and t + n.

$I_{t,t+n}$ and $E_{t,t+n}$ are number of immigrants and emigrants from the country during the same period

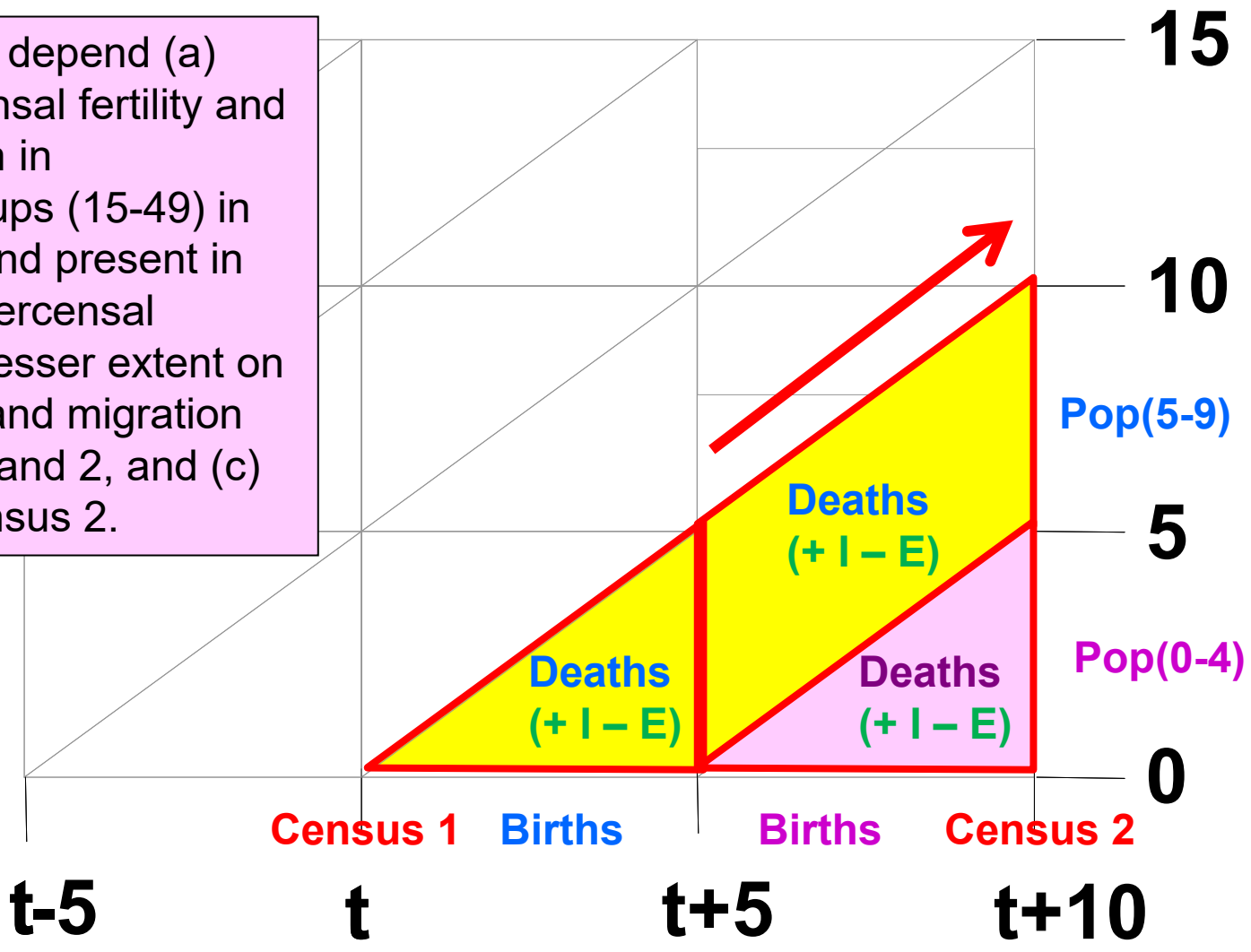
Cohort component

Population in census 2 only depends on past mortality and migration between censuses 1 and 2, and potential errors in each census…



Cohort component

Youngest age groups depend (a) mostly on the intercensal fertility and the number of women in reproductive age groups (15-49) in census 1 (surviving and present in the country during intercensal period), and (b) to a lesser extent on infant/child mortality and migration between censuses 1 and 2, and (c) potential errors in census 2.



Official population estimates: Current practice

Base population

- In majority of the regions, full (or partial) population census, or population register serve as base population for population estimates

Method of time adjustment

- Situation varies by region
- Practice of not bringing up to date the initial population still prevalent in few countries
- Not all methods produce consistent population estimates by age, sex and cohort

Adjustment to base population

- In many countries, no adjustment (or only age smoothing) is made to census data
- Under-enumeration, especially of young children, is not systematically corrected

>> Improper method of time adjustment and/or unadjusted base population produce national population estimates that are inconsistent and different from international (WPP) estimates

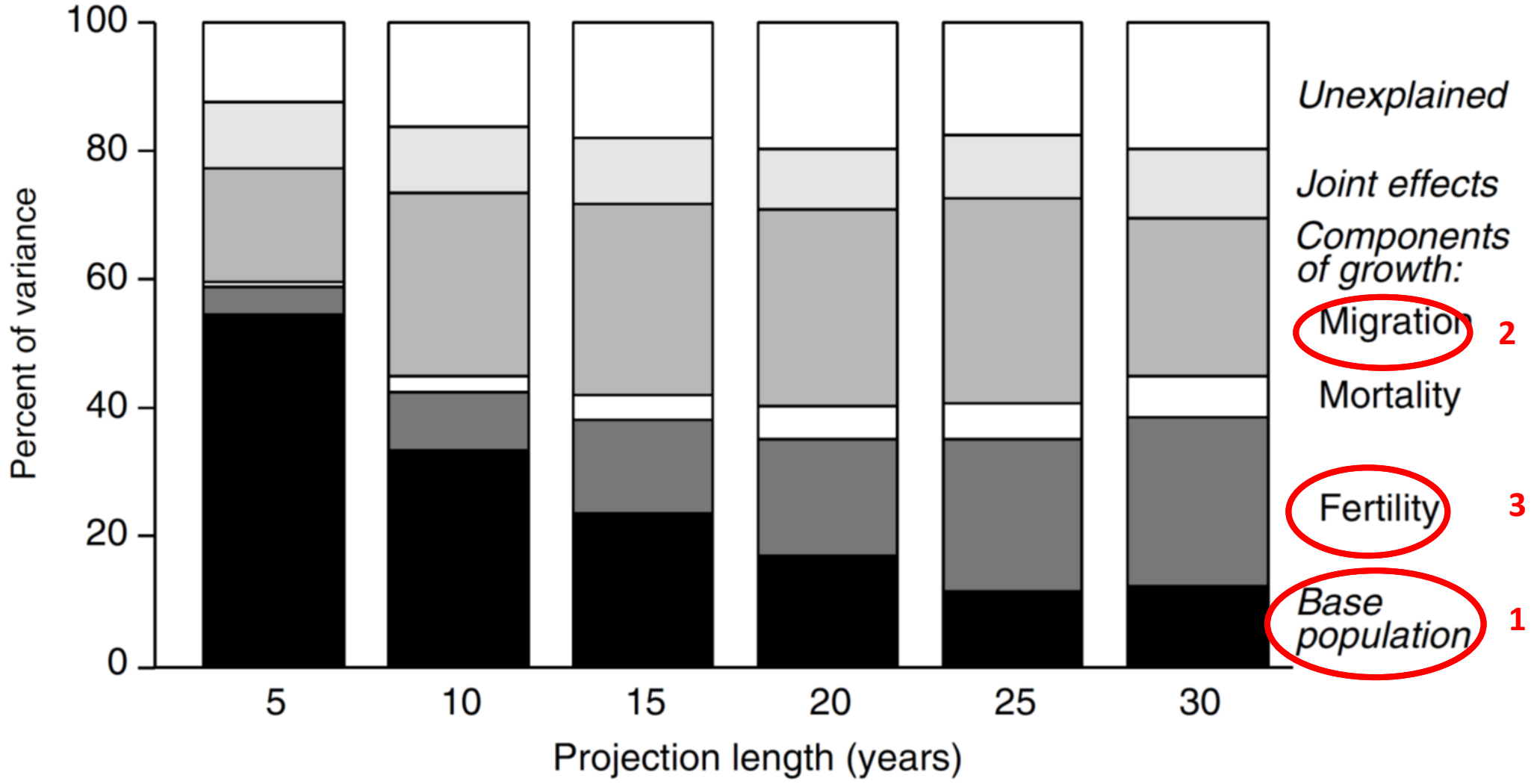
For further details, see Spoorenberg (2020)

Population projection accuracy: factors influencing forecast errors

- Gap/lag between latest available data and base year used for projection
- [Accuracy and reliability of the baseline population]
- Degree of disaggregation (sex, age, and other characteristics)
- Length of projection horizon (i.e., next 5-10 years vs. 1-3 future generations)
- Assumptions (fertility; mortality; migration)
  [and accuracy and reliability of past empirical estimates]
- [changing] demographic trends, reversal of trends, crises and new emerging issues (e.g., AIDS)

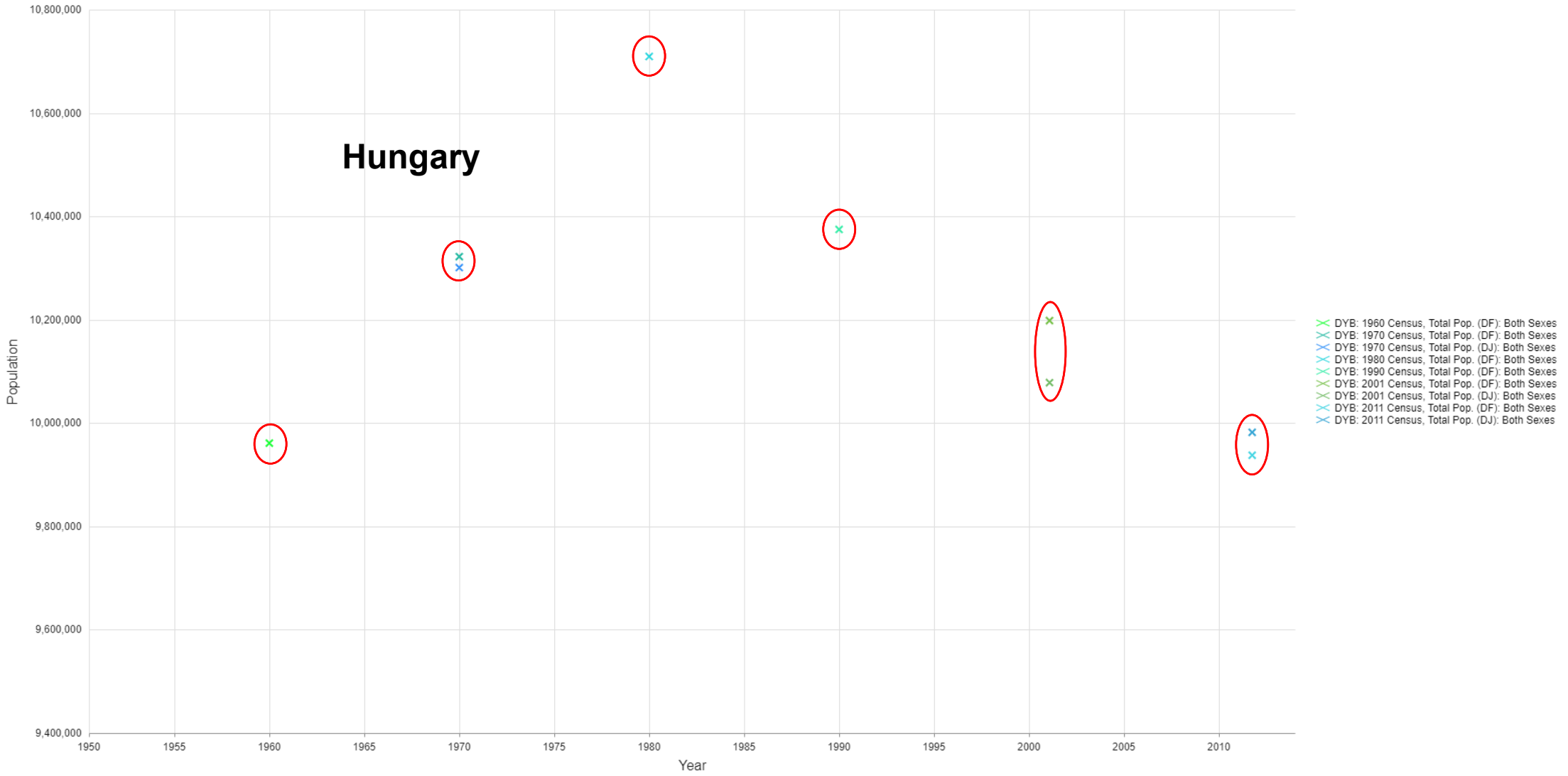
See **Keilman, N. (2001)** Data quality and accuracy of United Nations population projections, 1950-95. *Population Studies* 55: 149-164.

Sources of error vary in importance with the length of the projection horizon: 1950-1995.

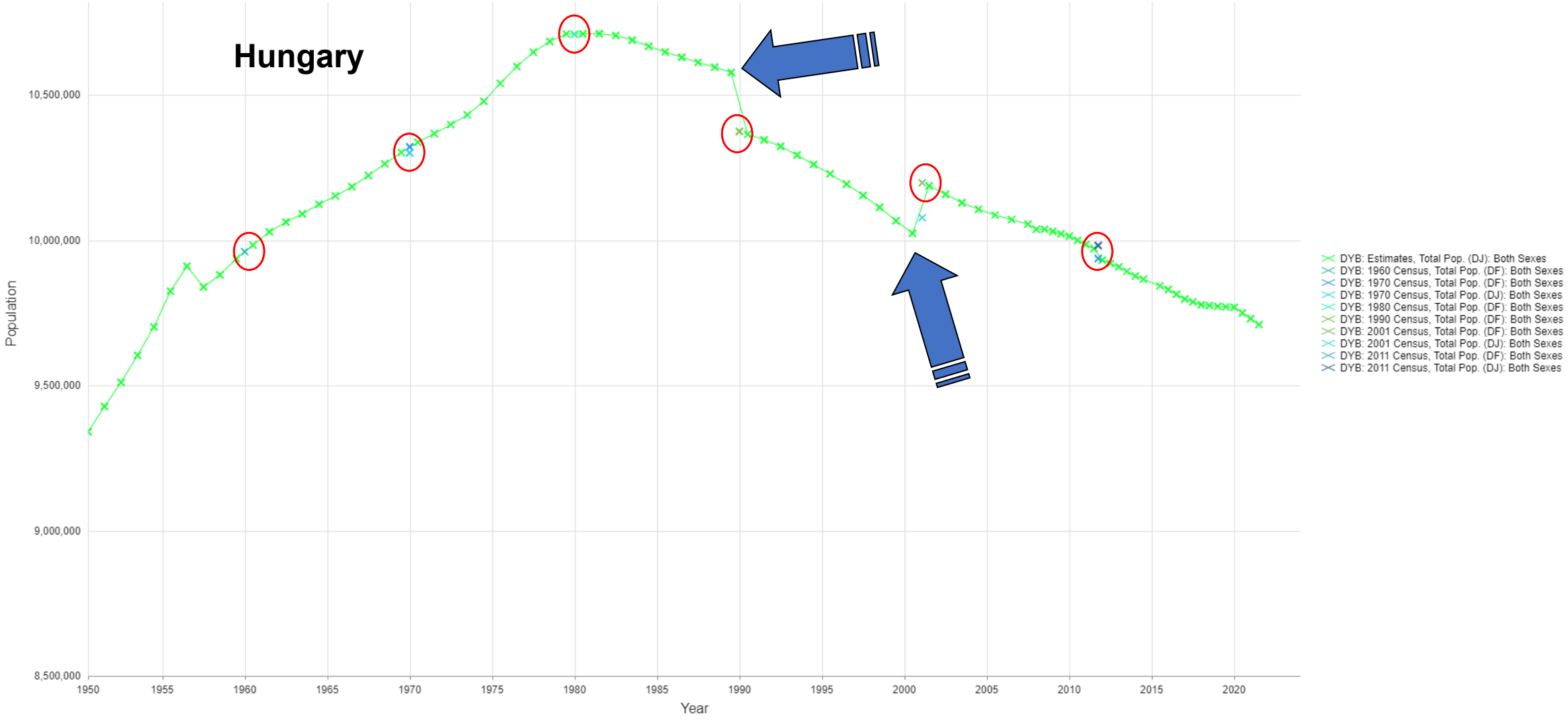


Source: Figure 2-6 (p.49) from National Research Council, & Committee on Population (2000). *Beyond six billion: Forecasting the world's population*. National Academies Press.

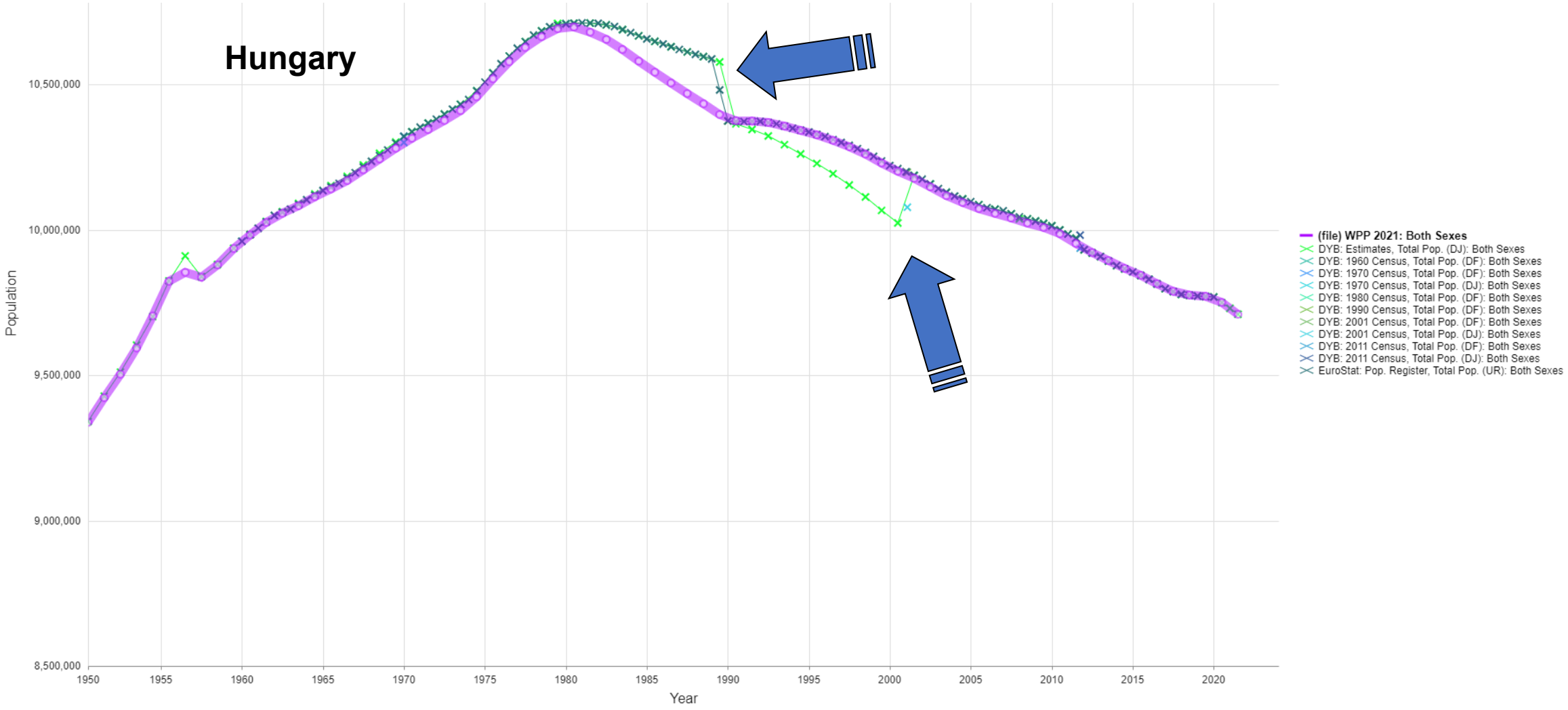
Official estimates and censuses



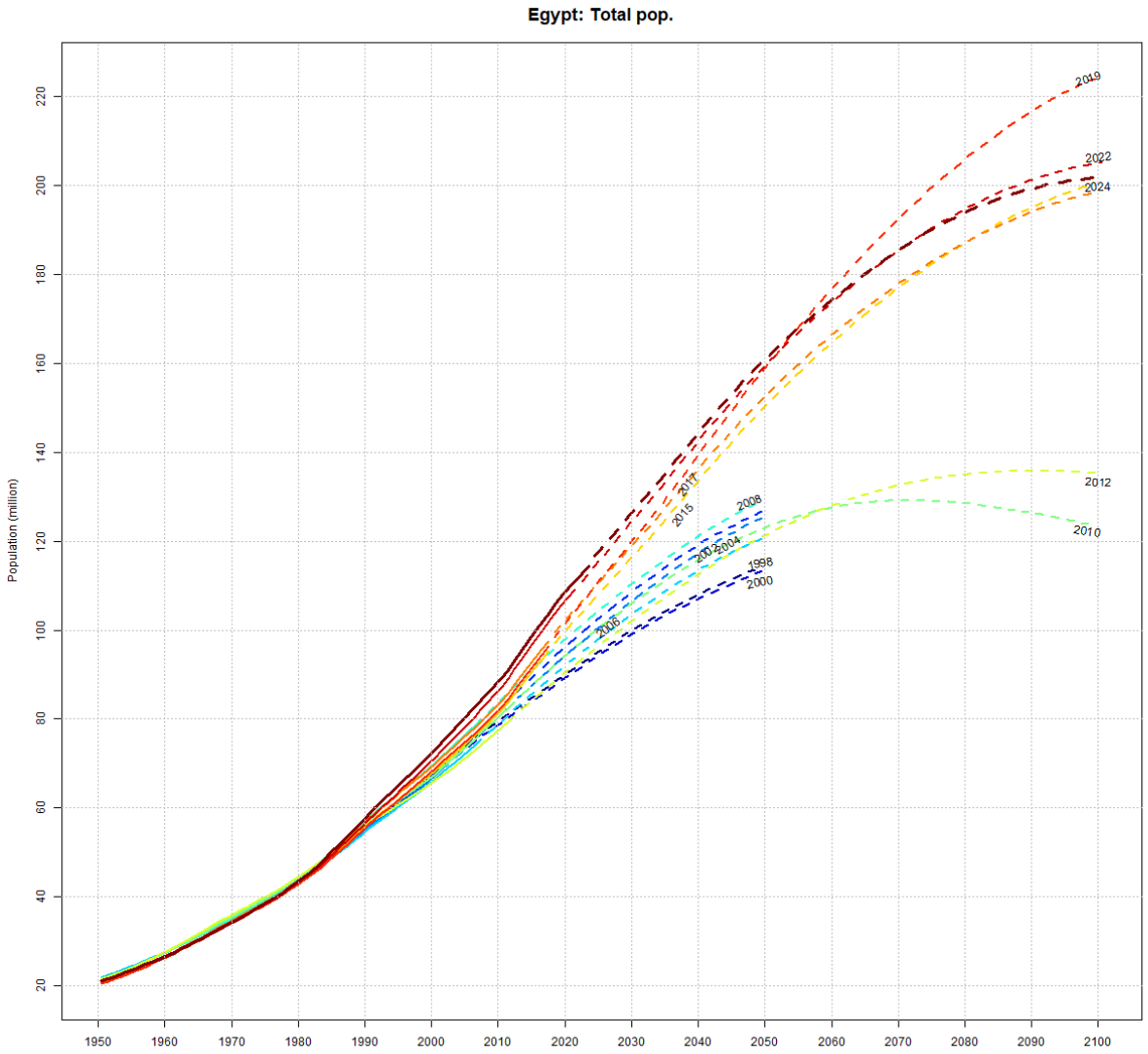
Official estimates and censuses

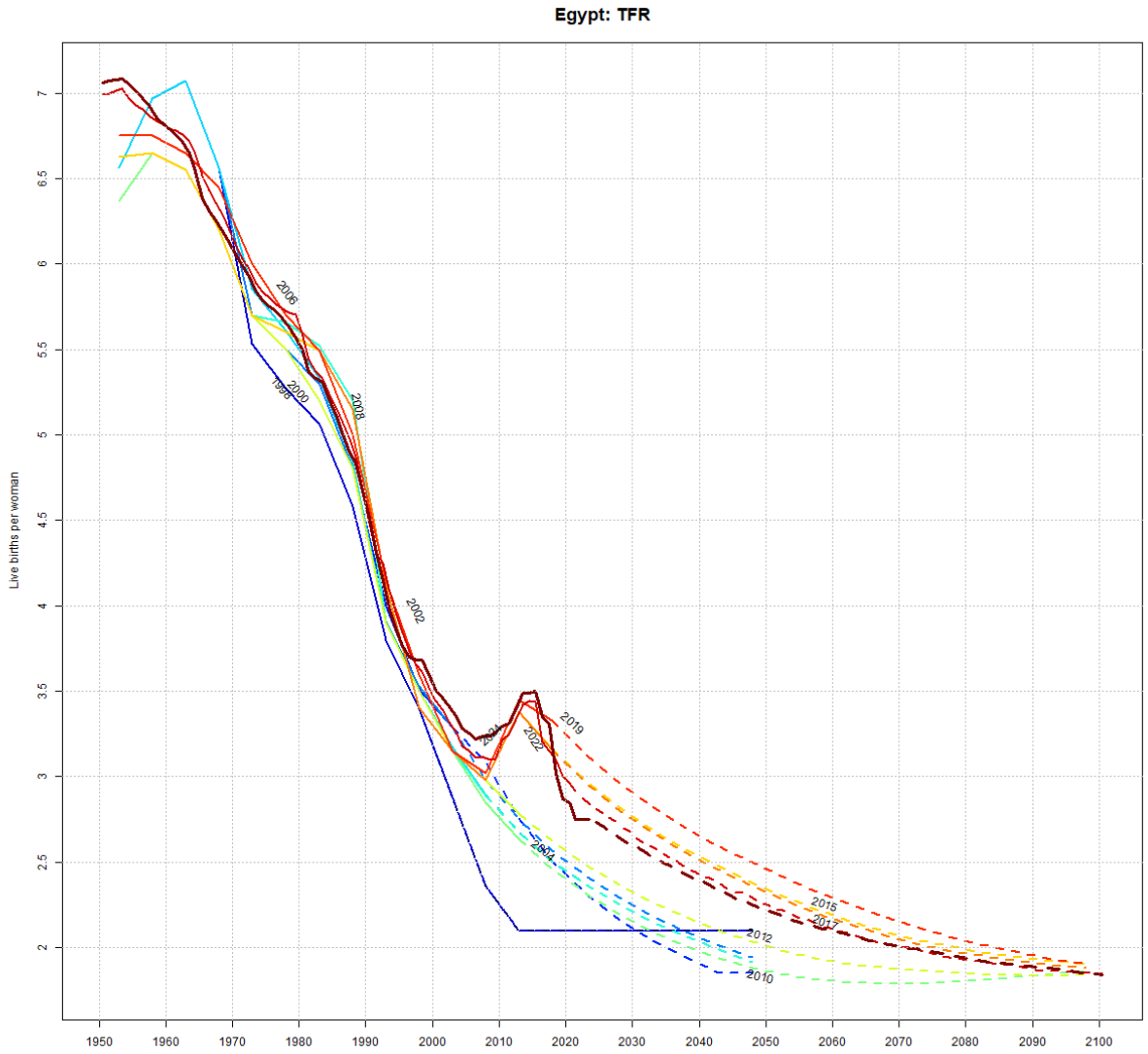


Official estimates and censuses

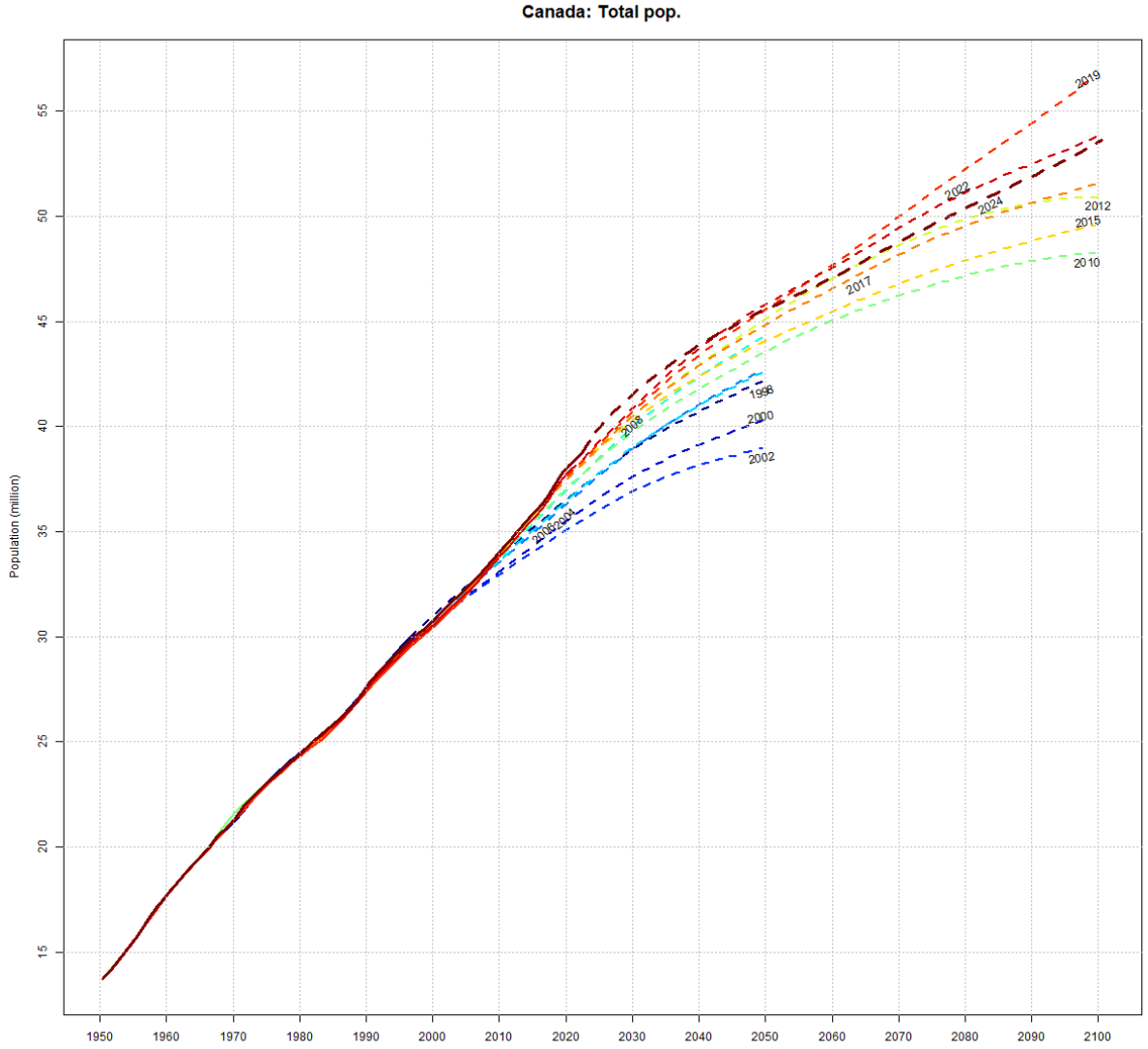


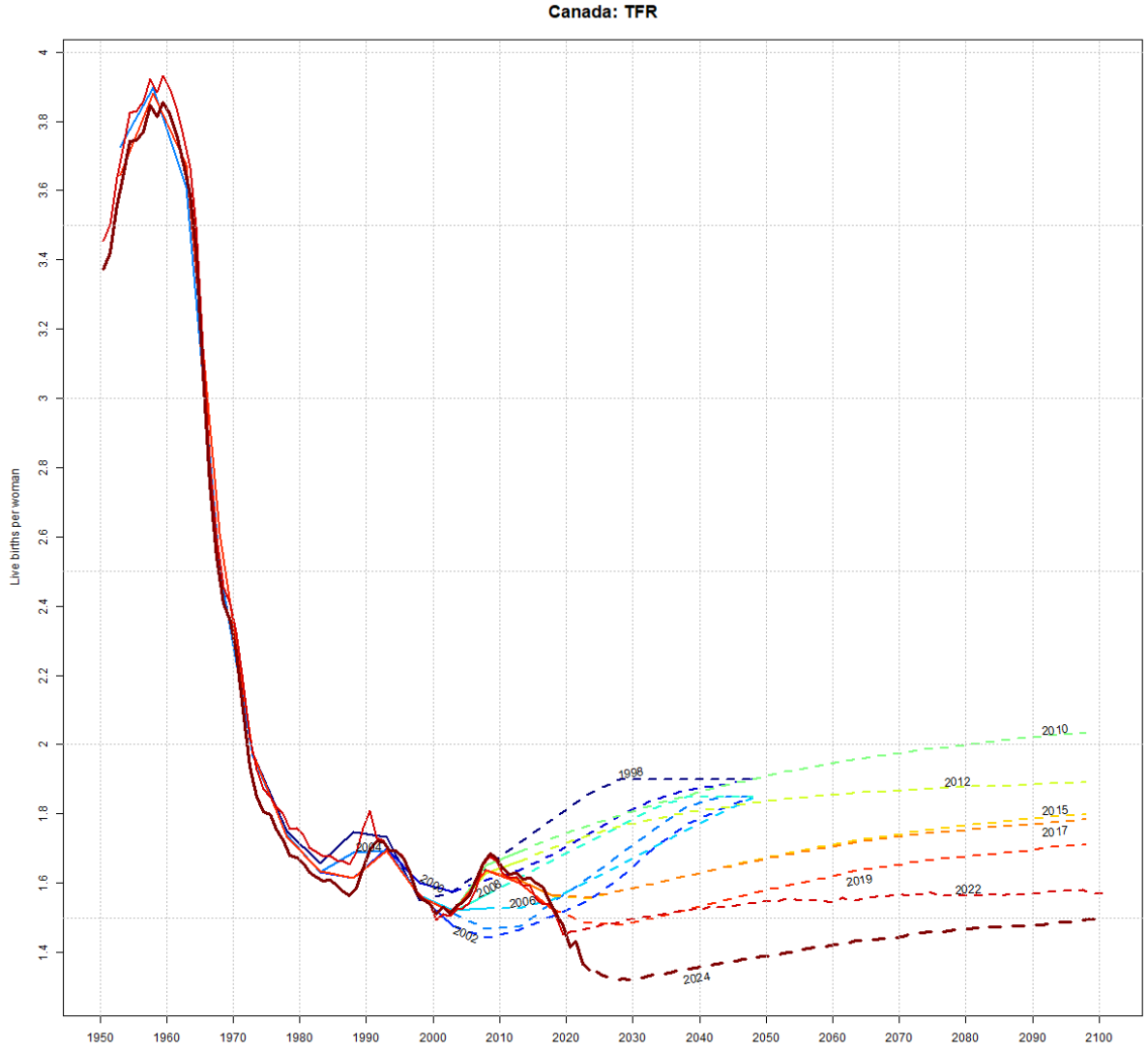
Egypt: WPP 1998-2024 revisions



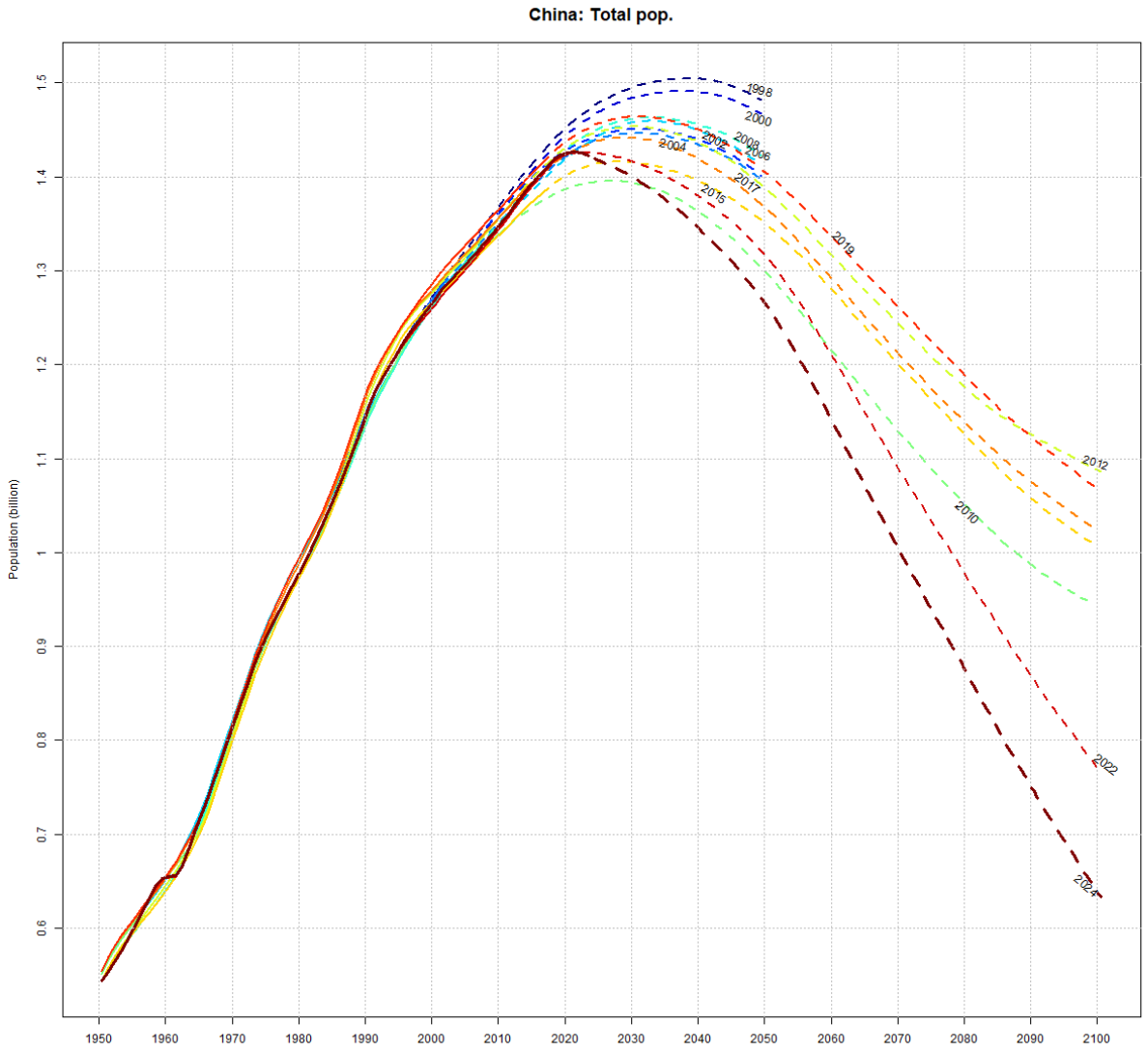


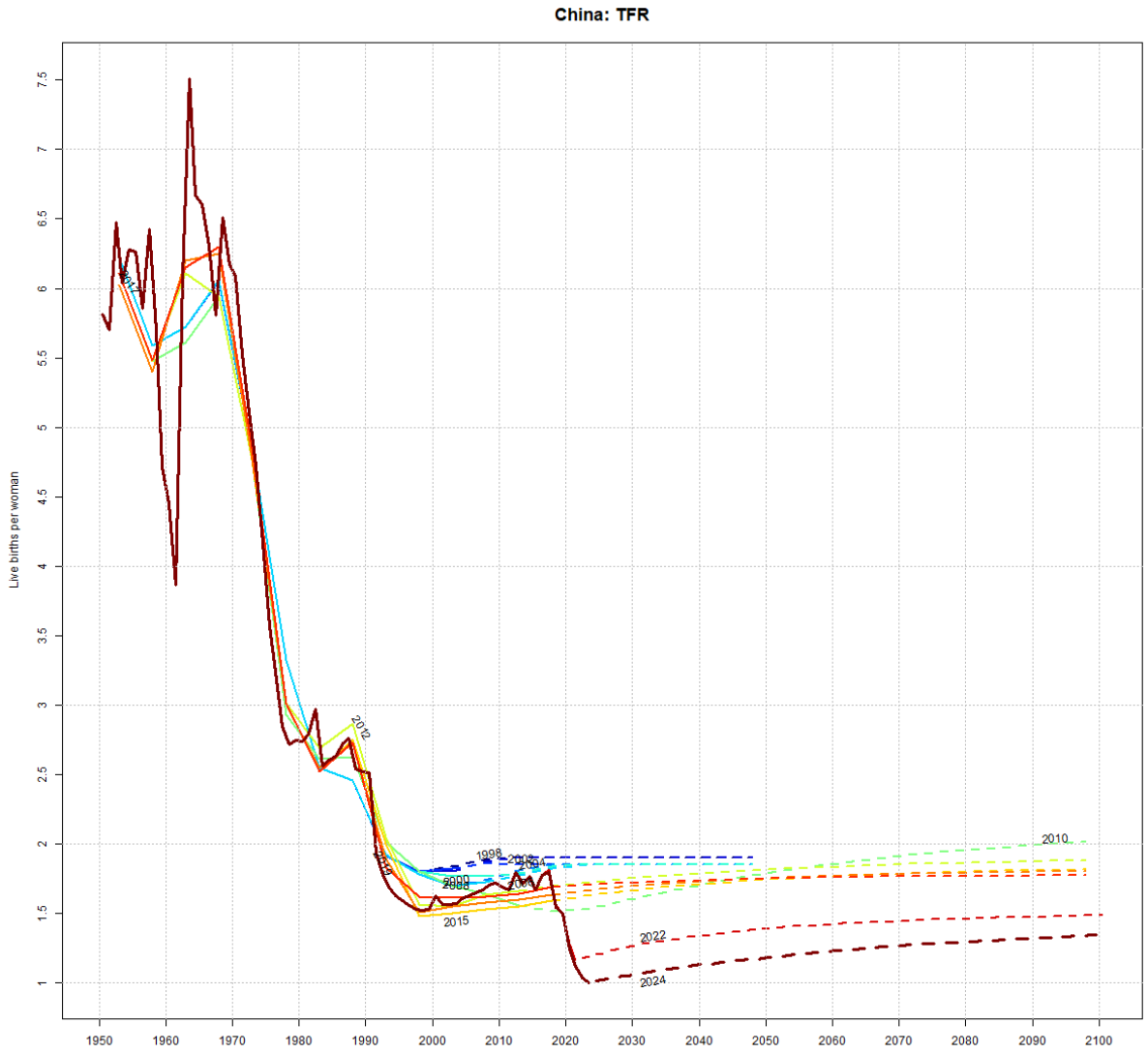
Canada: WPP 1998-2024 revisions



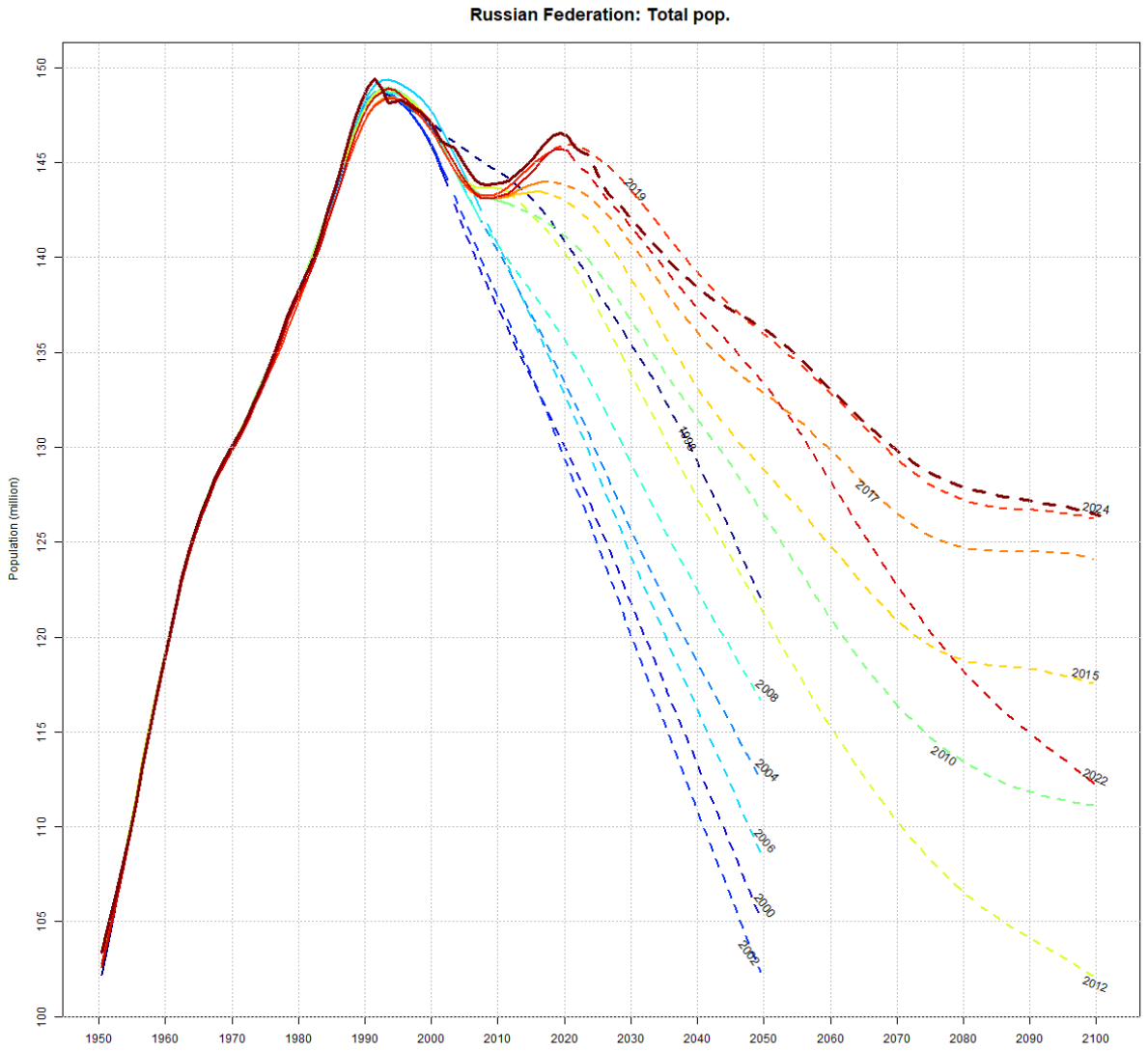


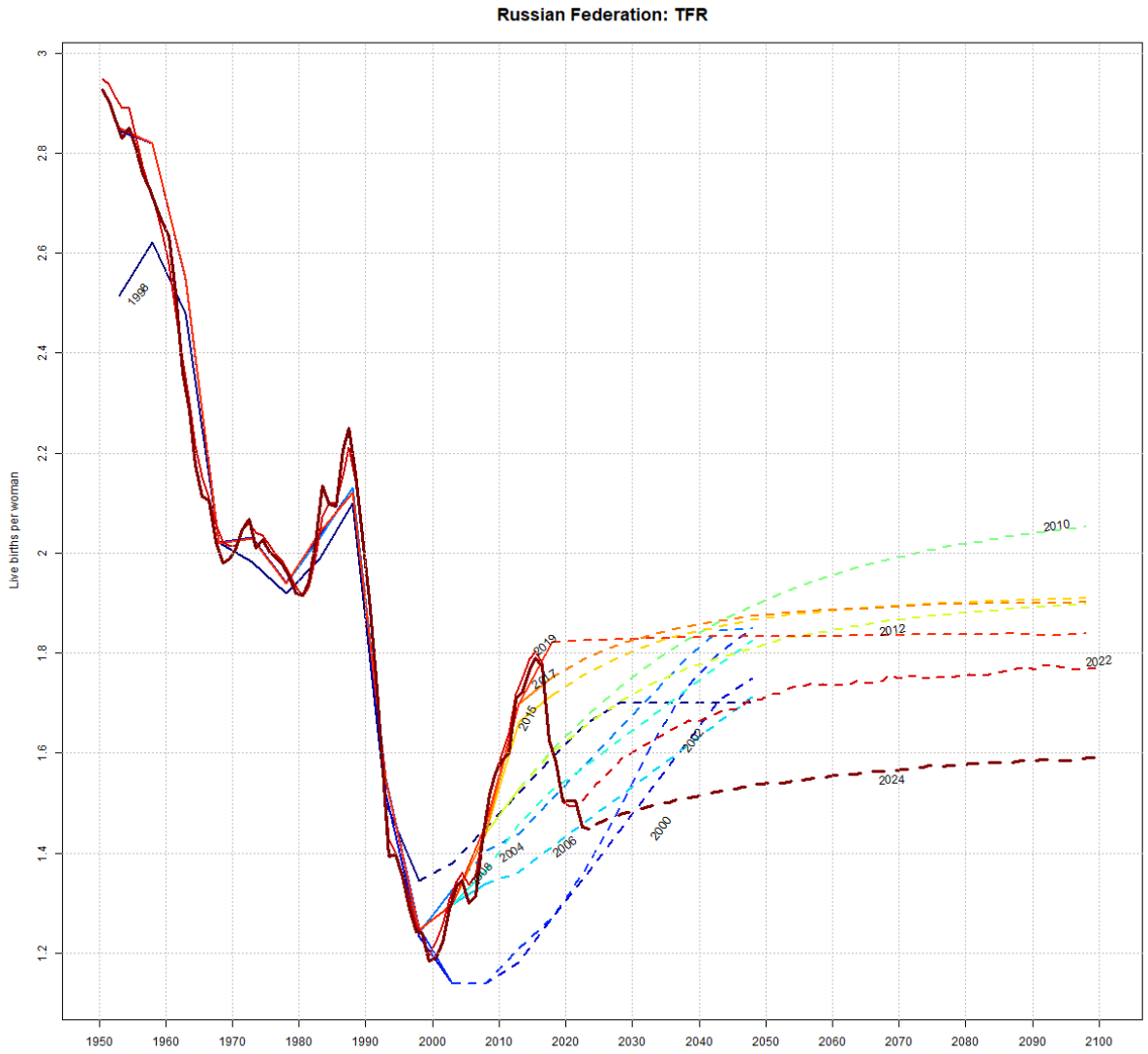
China: WPP 1998-2024 revisions





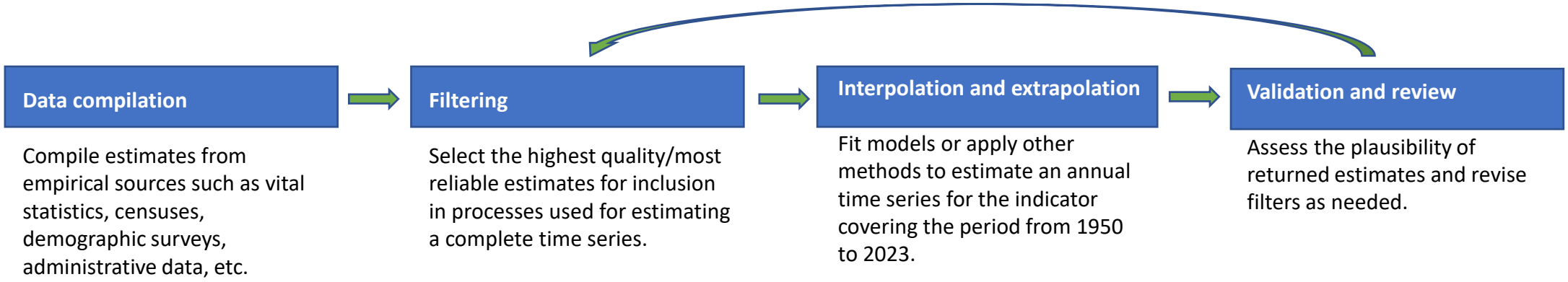
Russian Federation: WPP 1998-2024 revisions





Measurement and Methodologies in WPP

- **Probabilistic Modeling**: WPP uses Bayesian hierarchical models to project fertility, mortality and migration, incorporating uncertainty intervals.
- **Cohort Component Method**: Population estimates and projections rely on the cohort-component method, which estimates population by combining fertility, mortality, and migration data by age and sex over time.
- **Data Sources**: WPP draws from censuses, surveys, and vital registration systems across 237 countries to build population estimates.



See *WPP 2024 methodological report*, *and* structured metadata for further details

United Nations | Department of Economic and Social Affairs

Part 3

Country data sources and availability for population estimates

Data Availability for Population Estimates (WPP 2024)

- National statistical sources (either taken as-is or adjusted after in-depth evaluation) for 1950-2023:
  - **2,025 censuses** and 320 post-enumerations surveys
  - **2,970 surveys** (440 since 2015)
  - **vital registration systems** from **169 countries or areas**

Most recent census data:

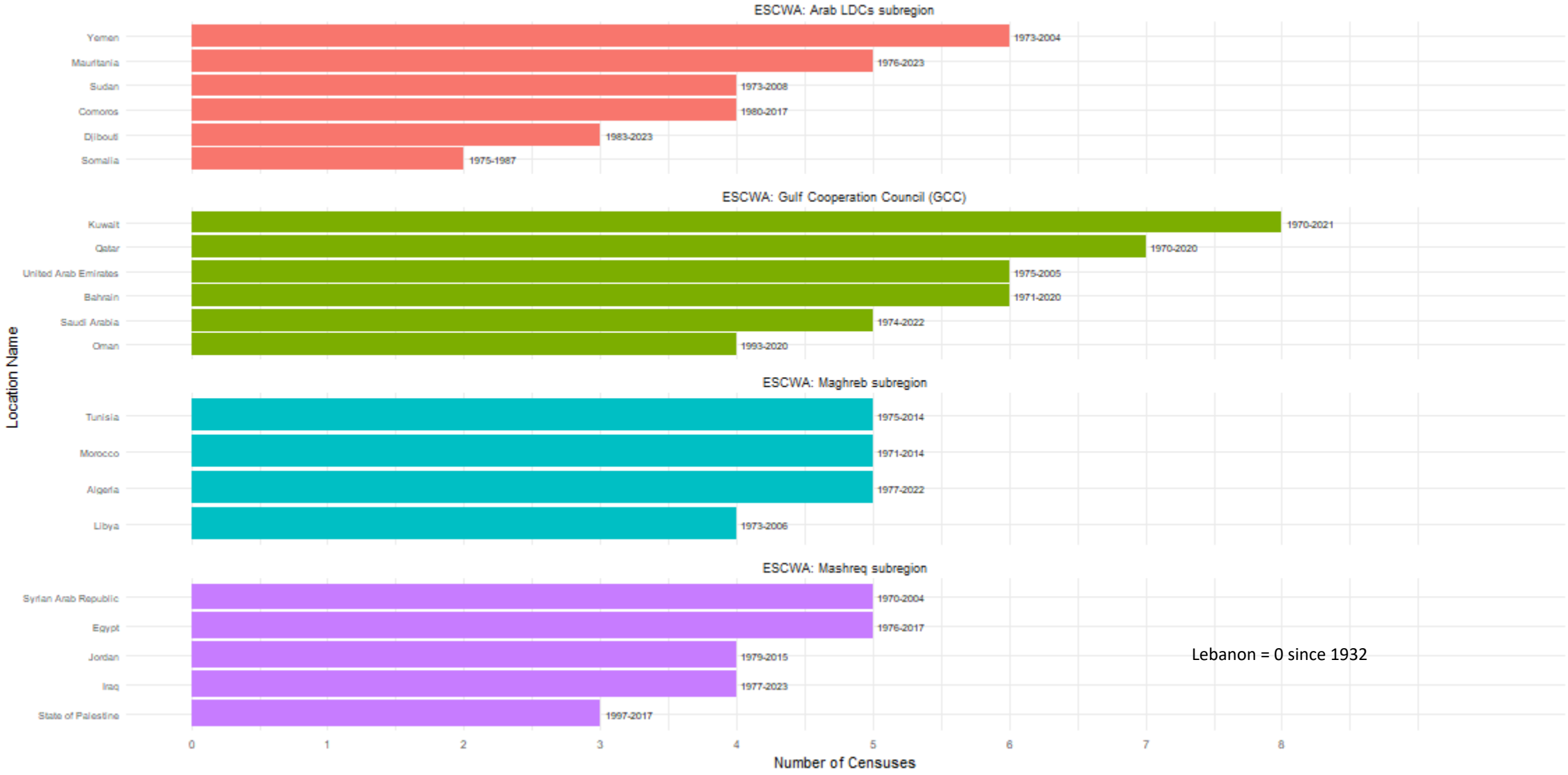
- 48% of countries or areas (n=114) have data since 2019 (and 54 per cent of the world population)
- 18% (n=43) from 2014-2018, 24% (n=57) from 2009-2013, 10% (n=23) before 2009

Latest available birth statistics (and their completeness) vary significantly by region:

- 83% of countries or areas (n=196) have fertility data since 2019, 14% (n=32) from 2014-18, 3% (n=8) before 2012
- Only 73% of the countries, territories and areas register at least 90% of births occurred (UNSD).

See *WPP 2024 methodological report*, *and* structured metadata for further details

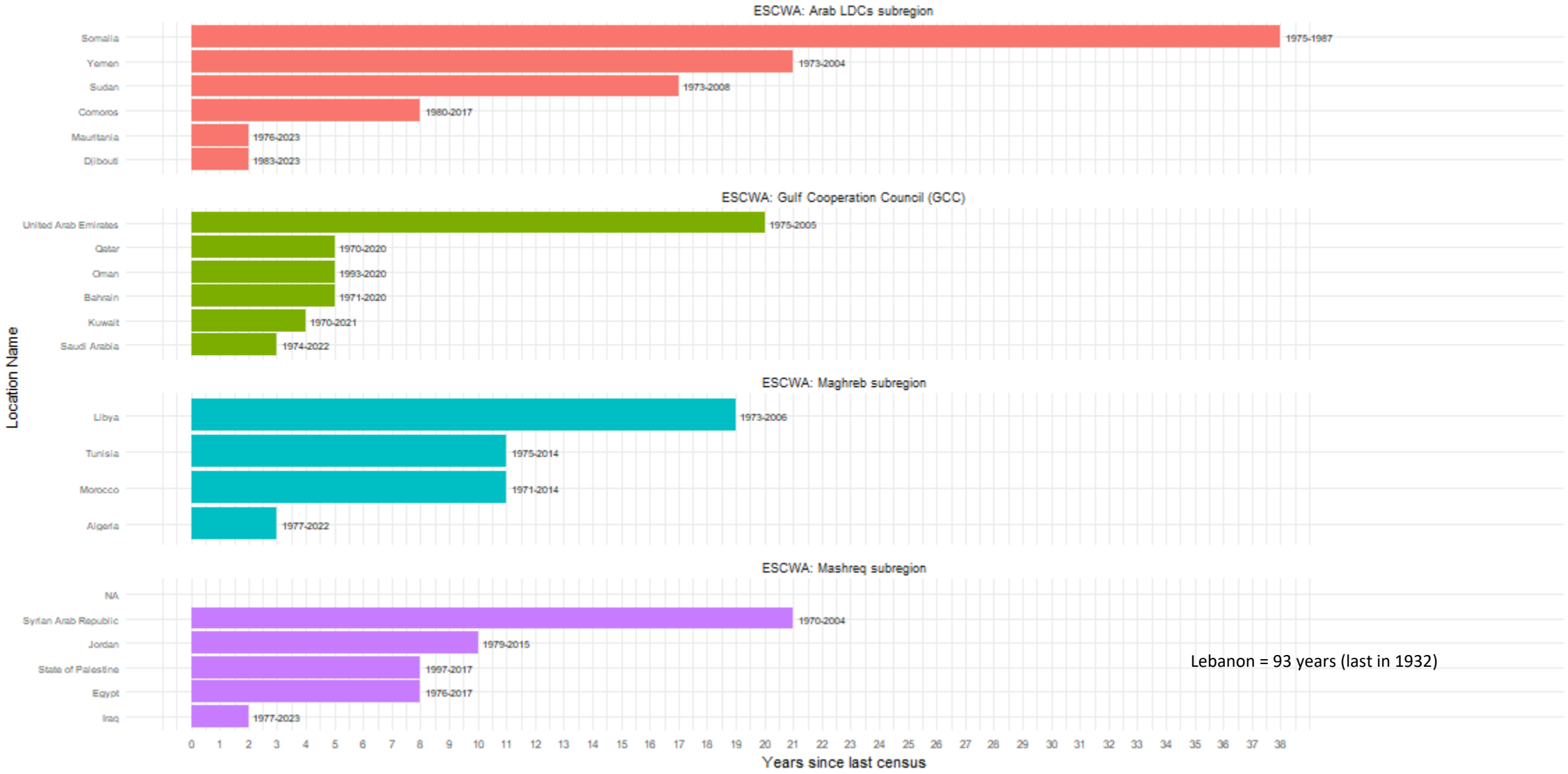
ESCWA region: No. censuses / country since 1970



Lebanon = 0 since 1932

See *WPP 2024 methodological report*, *and* structured metadata for further details

ESCWA region: years since latest pop. census



Lebanon = 93 years (last in 1932)

Data Availability for Population Estimates (WPP 2024)

Latest available mortality statistics vary significantly by region

- 120 countries or areas with good VR statistics for empirical life tables by age and sex, 117 countries or areas with empirical mortality rates by sex and age too sparse or of insufficient quality to estimate the complete annual time series of mortality rates.
- Only 68% of the countries, territories and areas have at least 90% of deaths occurred (UNSD).

Number of countries with data on age- and sex-specific deaths from complete vital registration for the years 2020 to 2023 (as of March 2024)

	Number of countries with 90 per cent or higher death registration	Proportion of total global deaths covered by complete VR (per cent)
2020	106	35
2021	100	30
2022	51	15
2023	5	1

- Under-five mortality: 78% of countries or areas (n=184) have data since 2019, 15% (n=35) from 2014-18, 3% (n=7) from 2009-13, and 2% (n=6) before 2009.
- Adult mortality: 65% of countries or areas (n=154) have data since 2019, 25% (n=60) from 2014-18, 6% (n=15) from 2009-13, 3% (n=6) before 2009, and 1% (n=2) no data.

See *WPP 2024 methodological report*, *and* structured metadata for further details

Data Availability for Population Estimates (WPP 2024)

Great variability in data sources and reliability over time and locations for (net) international migration estimates

- Official annual estimates (e.g., NSOs, Eurostat, n=54)
- Estimates of migrant flows (e.g., Eurostat, OECD, n=56)

n=67 distinct locations

- Foreign-born stocks (UN estimates) and implied annual change
- Administrative data (e.g., work permits issued/renewed)
- Literature review for major forced migration historical events
- Intercensal net residual migration for countries with good VR or residual from cohort-component projection compared to census or population register
- UNHCR estimates of refugee stocks (and implied annual change)

See *WPP 2024 methodological report*, *and* structured metadata for further details

Data sources (used for WPP 2024)

- National statistical sources (tabulations and/or microdata) either taken as-is or adjusted after in-depth evaluation:
  - **2,040 censuses** and 320 post-enumerations surveys
  - **2,980 surveys** (440 since 2015)
  - **Vital registration systems** from **169 countries or areas**
  - Official statistics reported to the **Demographic Yearbook** of the United Nations
  - Global or regional databases (WHO, Eurostat, HMD, HFD, etc.)
  - Public-use micro-datasets (IPUMS International, DHS, MICS, etc.)
  - Published reports and NSO web sites
  - **Population registers, other administrative sources, on international migration statistics, education statistics, immunizations, electoral rolls, etc.** providing independent sources of population headcounts for selected age groups

See *WPP 2024 methodological report*, *and* structured metadata for further details

Data sources (continued)

- Refugee statistics from the Office of the UN High Commissioner for Refugees
- Estimated time series of adult HIV prevalence and coverage of antiretroviral treatment from UNAIDS
- Estimated time series of infant and under-five mortality from the UN Inter-Agency Group for Child Mortality Estimation
- Estimates of international migration flows and stocks of foreign-born persons from the UN
- Various other series of international estimates produced by international and regional organizations and academic research institutions

With so many data available, why estimates are necessary…

- **To fill-in gaps in missing data**: most information often available only for some countries and/or dates, or not sufficiently disaggregated by age
- **To reconcile differences** between (a) data sources and/or estimation method(s) for a specific date and (b) within sources over time
- **To ensure international comparability** using similar definitions/concepts, methodology and assumptions across countries

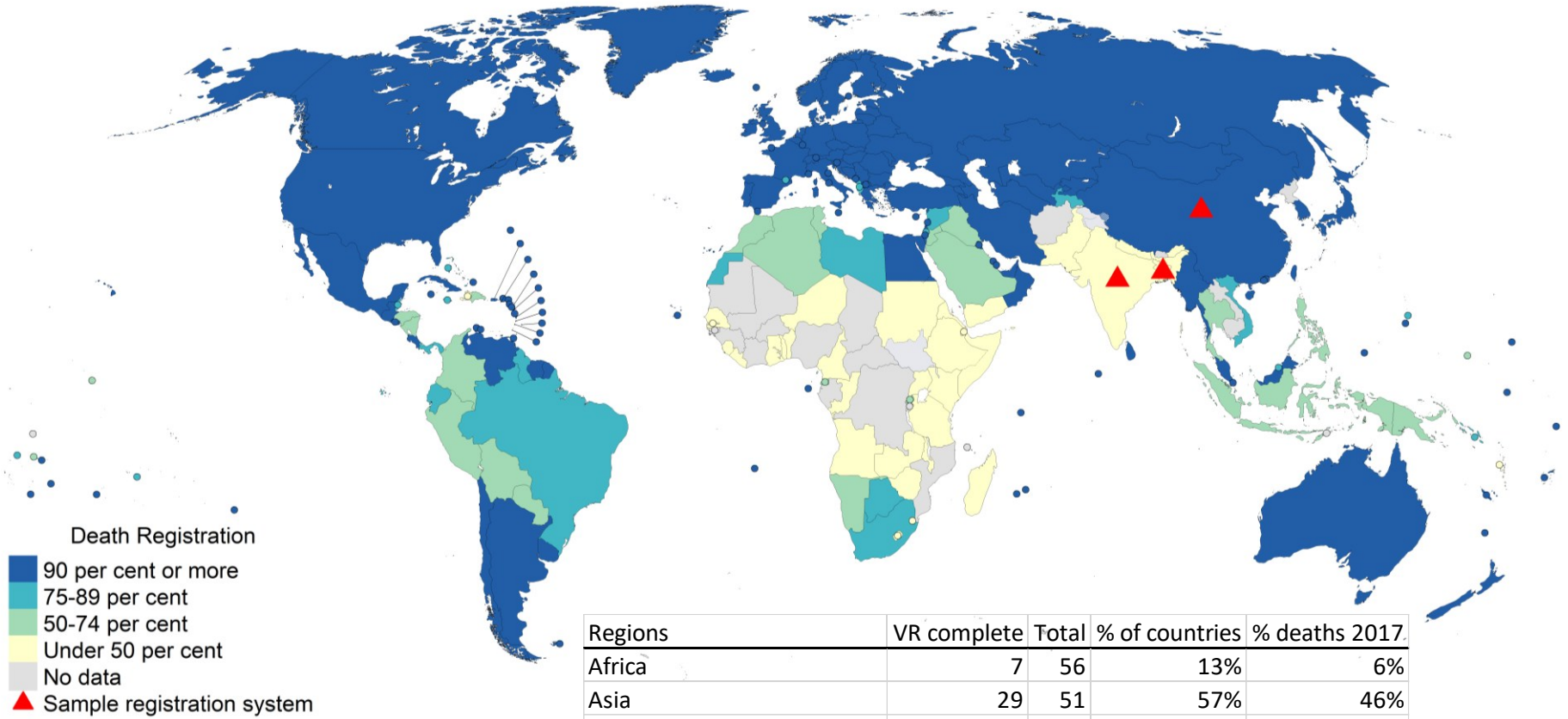
UN population estimates and projections are used in the calculation of many development indicators used by the United Nations system, including about one-quarter of the indicators used to monitor global progress towards the achievement of the SDGs.

Estimation of robust time series for demographic rates (e.g., fertility and mortality)

- For many countries, data available vary greatly in quantity, frequency, quality, reliability and consistency.
- Not all data points are as informative and can be trusted equally…
- Estimates can vary based on the **type of data sources** (census, surveys, vital registration), the **type of survey** itself (national survey vs. international survey programs), the **estimation methods** (direct or indirect estimates) and by **various biases affecting reporting of retrospective data (e.g., birth histories or lifetime fertility, parental or siblings' survival)**.

See *WPP 2024 methodological report* for further details

Completeness of Death Registration
(UNSD, last updated December 2016)



Regions	VR complete	Total	% of countries	% deaths 2017
Africa	7	56	13%	6%
Asia	29	51	57%	46%
Europe, N. America, AUS/NZ	48	50	96%	100%
Latin America & Caribbean	31	46	67%	41%
Oceania	11	21	52%	13%
Sum	126	224	56%	46%

United Nations | Department of Economic and Social Affairs

Part 4

WPP analytical workflow, estimation strategy and statistical modelling for components of population change (fertility, mortality, and international net migration)

WPP estimation workflow

1. **TFR**: bayesTFR with empirical series and bias adjustments
2. **ASFR**: B3 with empirical series and bias adjustments + graduation into single age using (re)calibrated splines.
3. **nQx**: B3 for 45q15 empirical series and bias adjustments
4. **LifeTables**: abridged/complete empirical LT and/or MLT / hybrid time series + graduation into single ages (ungroup)
5. **Population Census protocol**: abridged/complete population age distribution standardization, evaluation/adjustment, extension, graduation
6. **Intercensal residual migration**: by age/sex and overall
7. **Net migration by Age/Sex**: annual time series based on mixed methods
8. **CCMPP for population reconstruction**

See *WPP 2024 methodological report* for further details

WPP workflow process with vital rates

- Compile and compute direct and indirect estimates from as many empirical data sources as possible for each country since 1950
- Review and assess the various series
- Generate an initial robust time trend for overall summary indicators (e.g., TFR, 5q0, 45q15) and age-specific rates (e.g., ASFR, life tables)
- Use this initial set of estimates within the full cohort-component population reconstruction by age and sex since 1950
- Compare and assess the reconstructed population cohorts with those enumerated across the various censuses
- Revise and adjust the set of WPP estimates to reconcile the various demographic components (e.g., TFR) that satisfy the demographic balancing relationships over time, age and cohorts

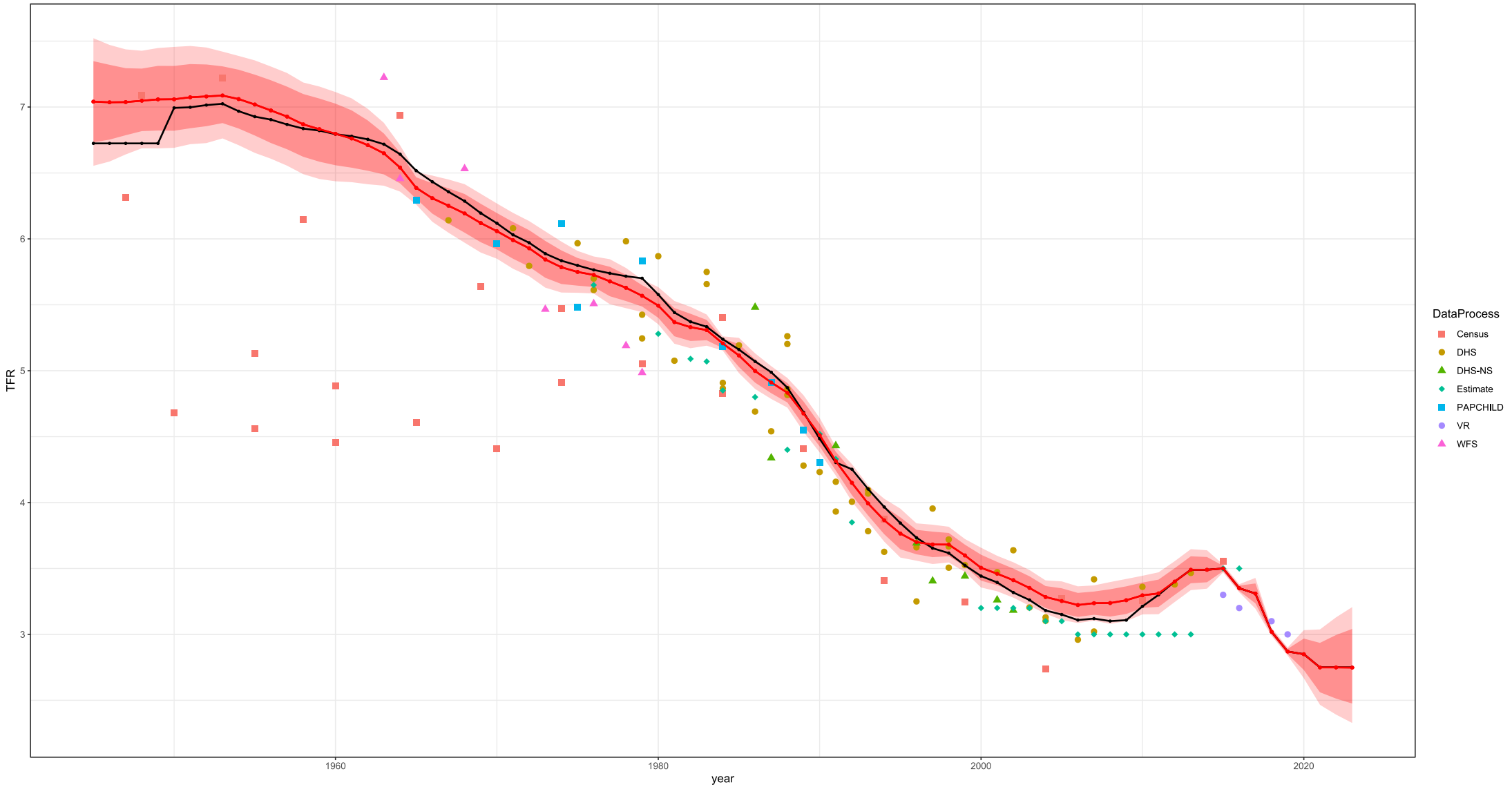
See *WPP 2024 methodological report* for further details

Sources of fertility data and estimation methods

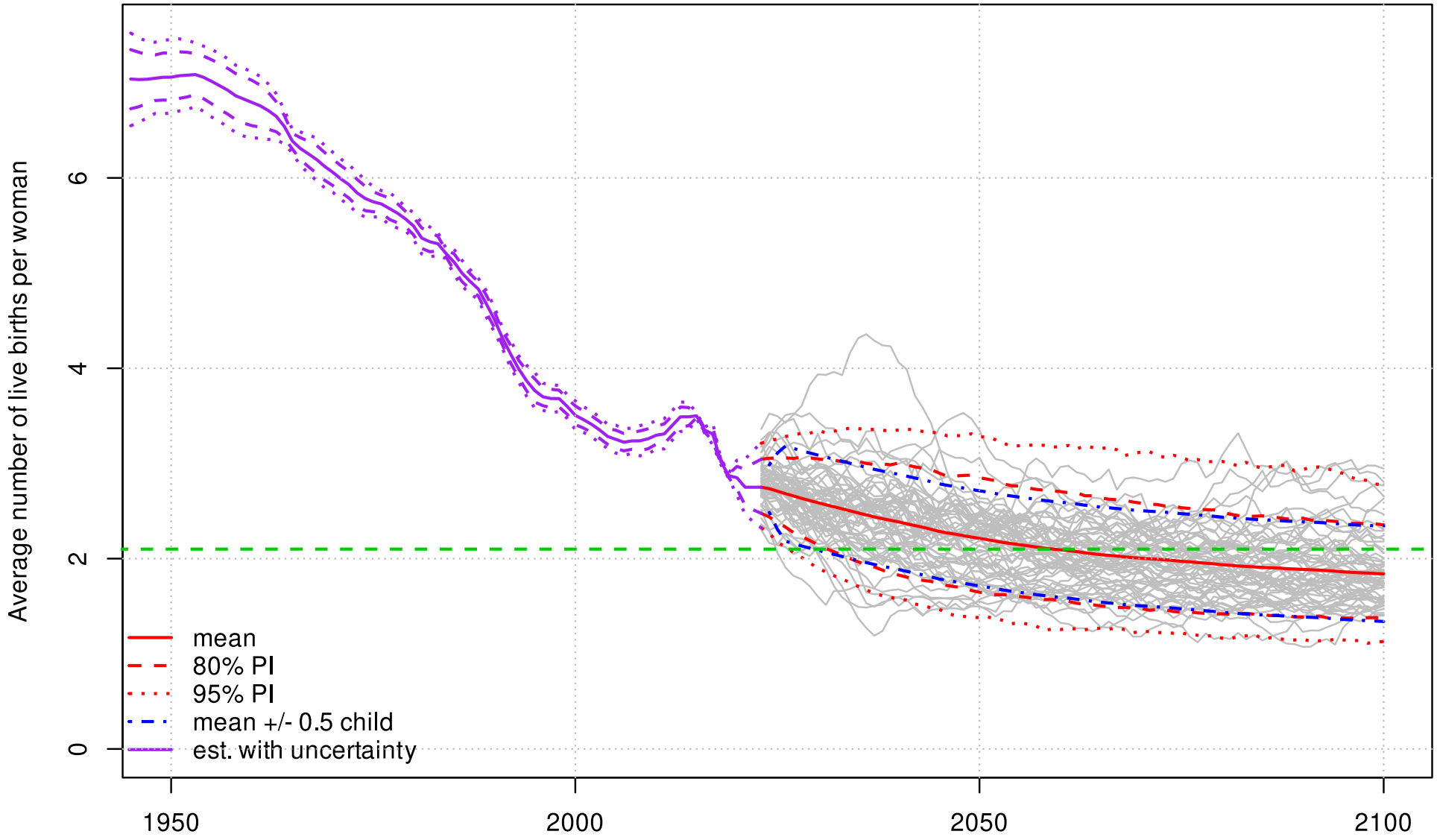
Source	Method	Time period	TFR	ASFR
Official figures	Estimates	**Annual**	✓✓✓✓	✓✓✓✓
Vital statistics from civil registration	Computed rates from DYB-NSO	**Annual**	✓✓✓✓	✓✓✓✓
Surveys	Birth histories (and extrapolations)	**Prior 15-35 years**	✓✓✓✓	✓✓✓✓
Censuses/Surveys	Recent births	**Prior 12-24 months**	✓✓✓	✓✓✓
Censuses/Surveys	Recent births and average parity methods	**Prior 12-24 months**	✓✓✓	✓✓✓
Censuses/Surveys	Children ever born methods	**15-45 years before**	✓✓	
Censuses/Surveys/ admin. stats	Population methods	**Prior 15 years**	✓✓[1,2]	✓✓[1]
Model-based	Other methods[3]	**Prior 5-15 years**	✓	

(1) Own-children method, (2) Reverse survival method, Education stats
(3) Intercensal estimates, regression methods, stable population

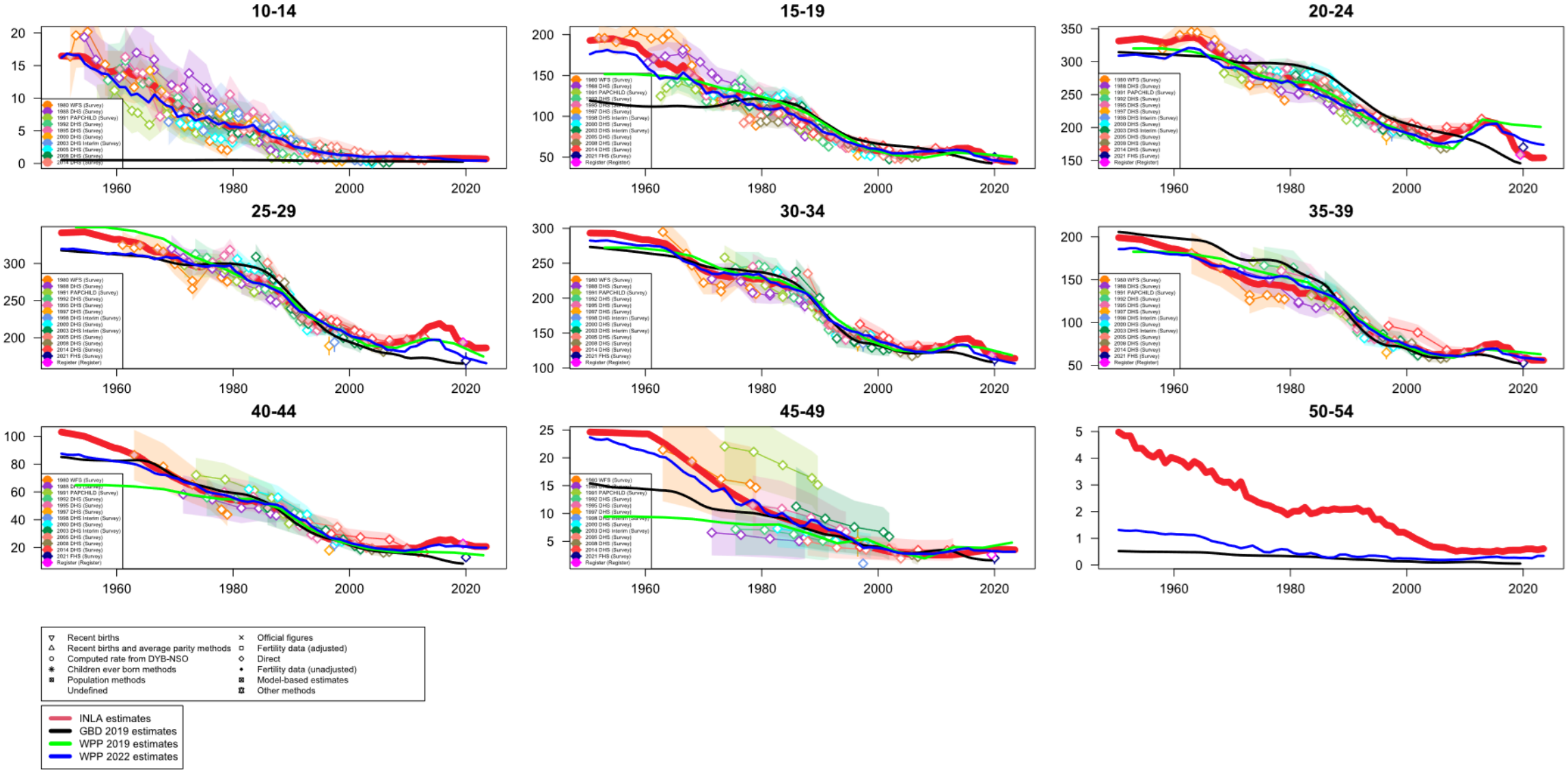
Egypt TFR: 1950-2023 time trend



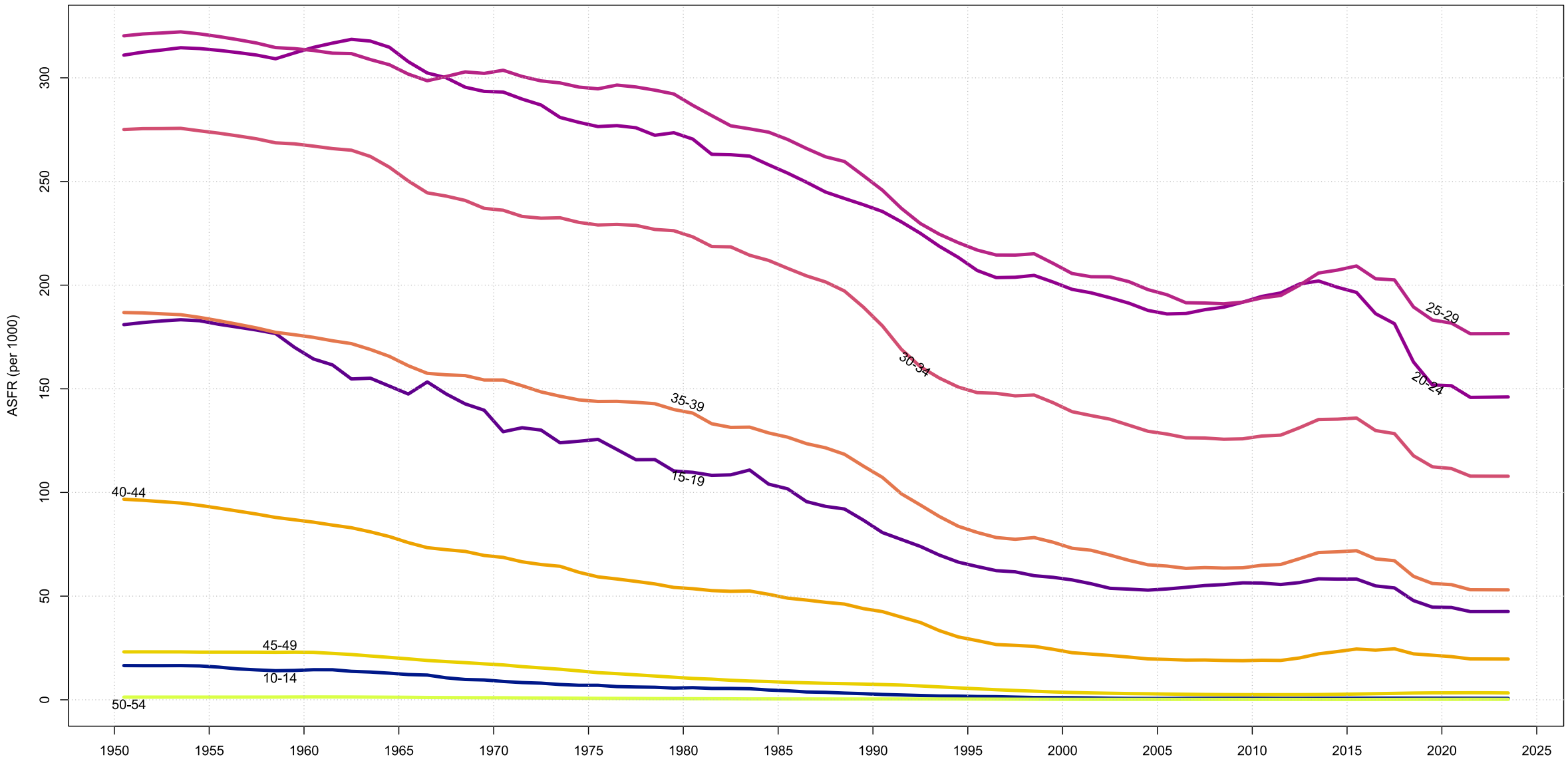
Egypt TFR: 1950-2100 time trend



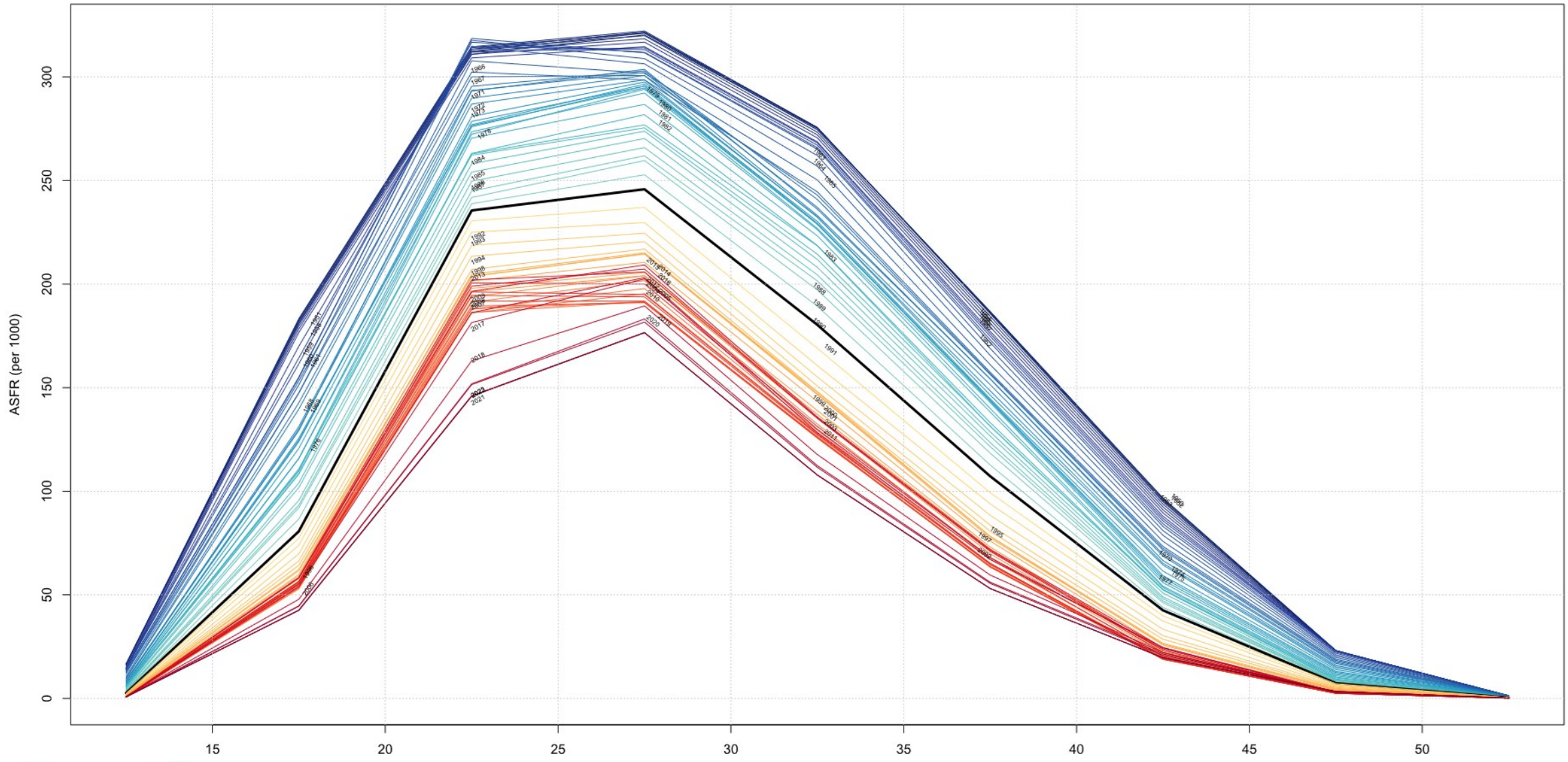
Egypt fertility rates 5-year age groups: 1950-2023



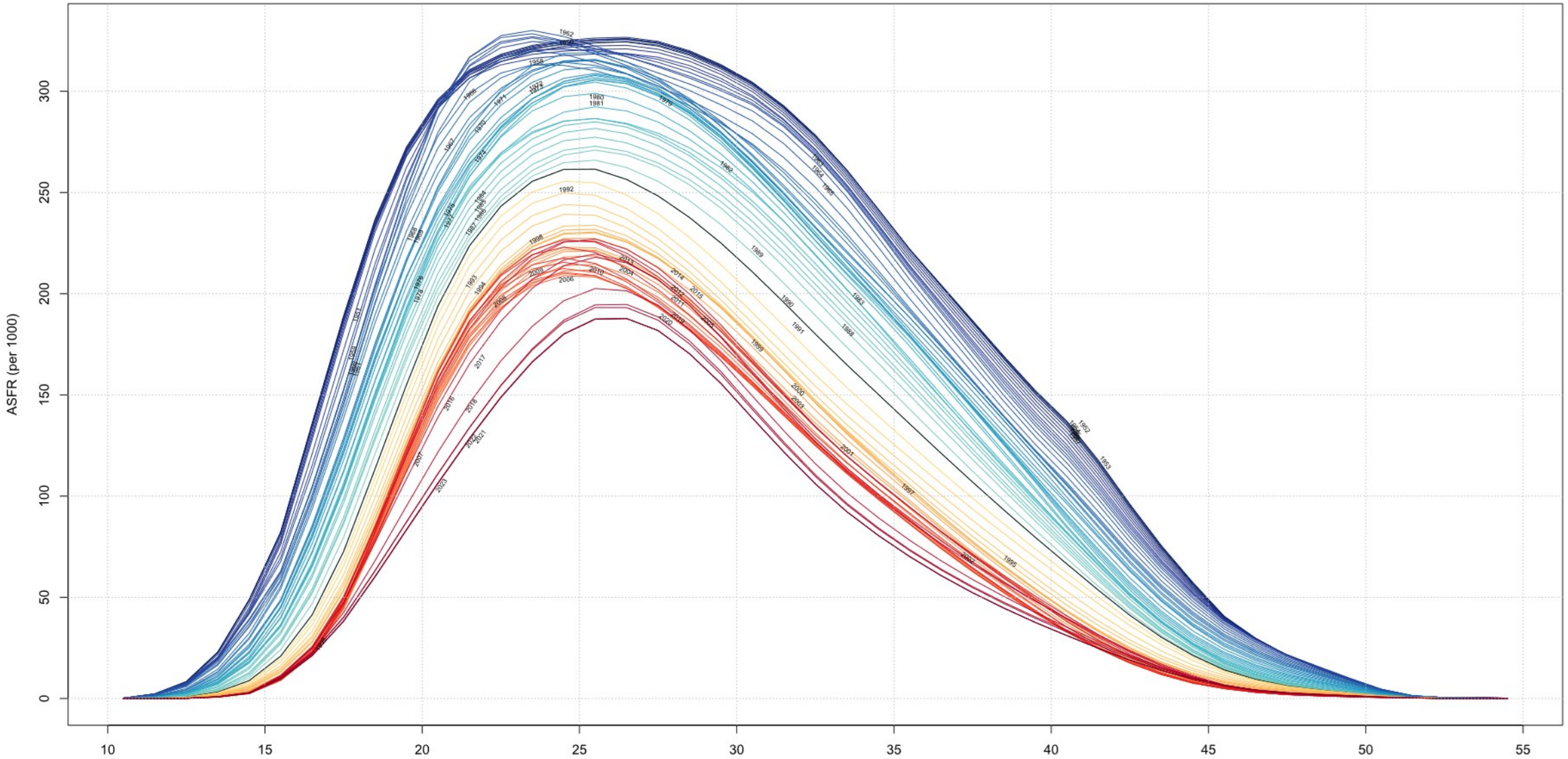
Egypt fertility rates 5-year age groups: 1950-2023



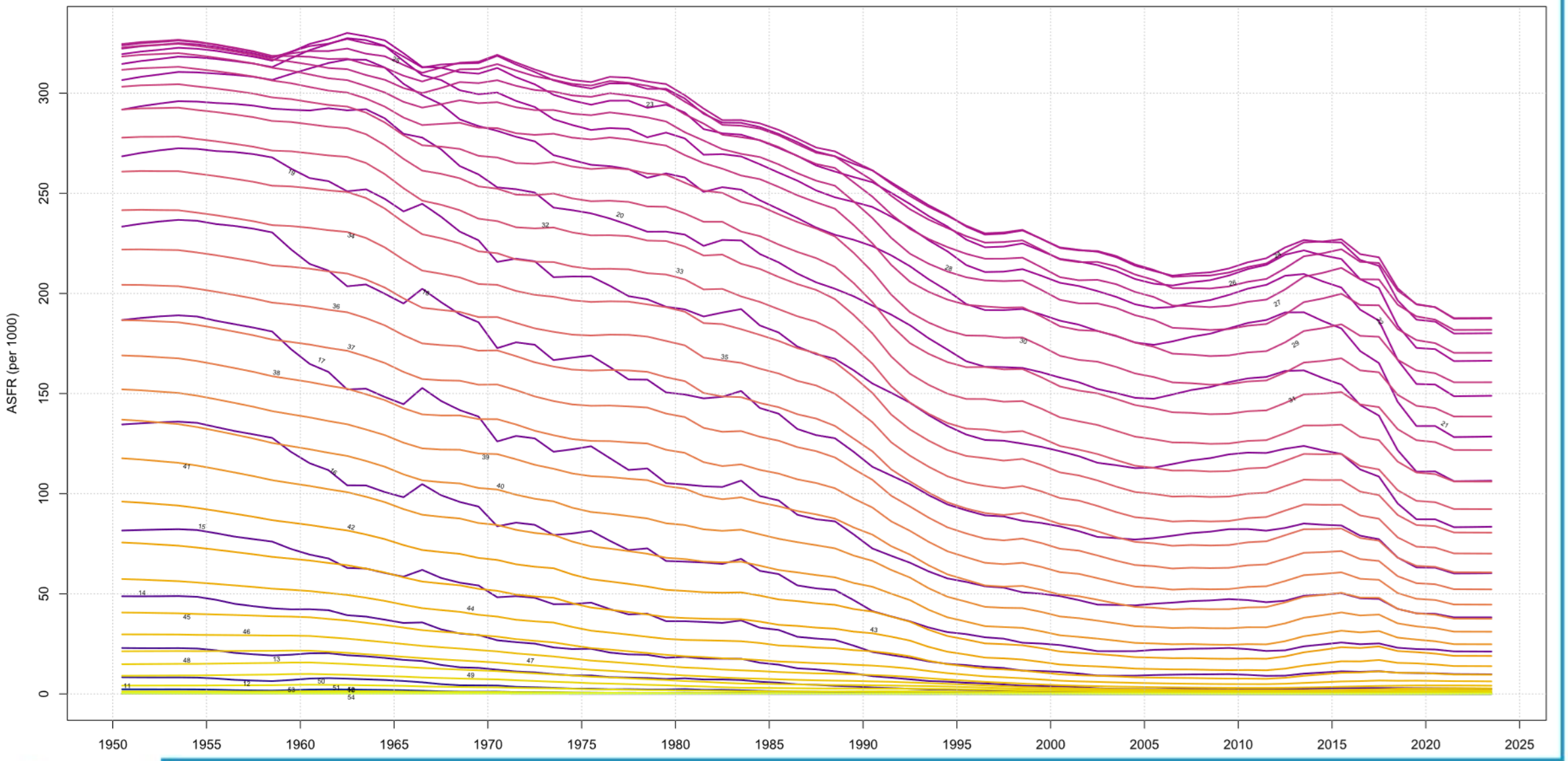
Egypt fertility rates 5-year age groups: 1950-2023



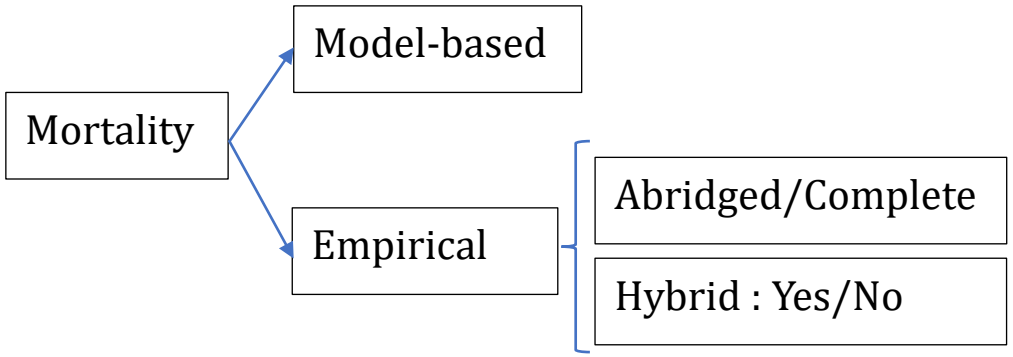
Egypt fertility rates graduated by age: 1950-2023



Egypt fertility rates graduated by age: 1950-2023



Age-specific type of mortality input data



- 120 locations with empirical life tables based on different types of inputs and data gaps

Type of age pattern	Input data	Locations (n)
Empirical	Complete	44
Empirical	Abridged	76
Model-based	Abridged	117
	Total	*237*

Life table estimation steps: good VR

1. Consolidate all sources of death counts and exposure available
2. Review and standardize for definitional issues both numerators and denominators
3. Compute 5x1 and 1x1 life tables based on HMD-like protocol using death counts and exposure
4. Interpolate/extrapolate over time if necessary

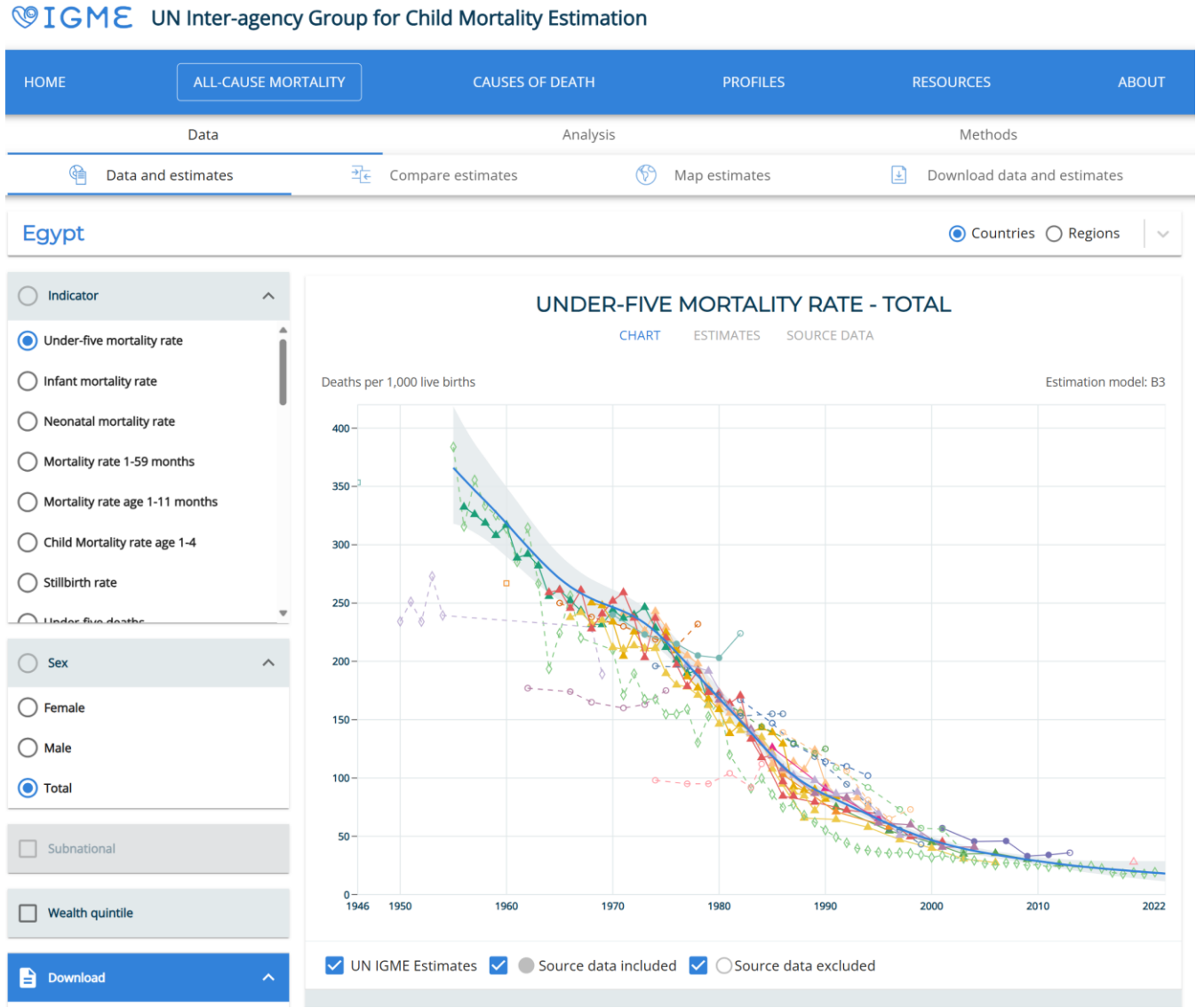
Life table estimation steps: incomplete/deficient vital registration

1. Consolidate all sources of death counts and exposure available
2. Review and standardize for definitional issues both numerators and denominators
3. Evaluate completeness and adjust deaths (age 5+) if necessary
4. Compute (spliced) 5x1 and 1x1 life tables: use 1q0 and 5q0 from IGME, compute 1q0 and 4q1 by sex and mortality rates (age 5+)
5. Interpolate/extrapolate over time if necessary
6. Smooth/adjust old age mortality rates if necessary to insure consistency by (1) age (monotonic increase), (2) sex over time (monotonic decline) and (3) between sex by period (M ≥ F)

Life table estimation steps: no VR, only child (and adult) mortality

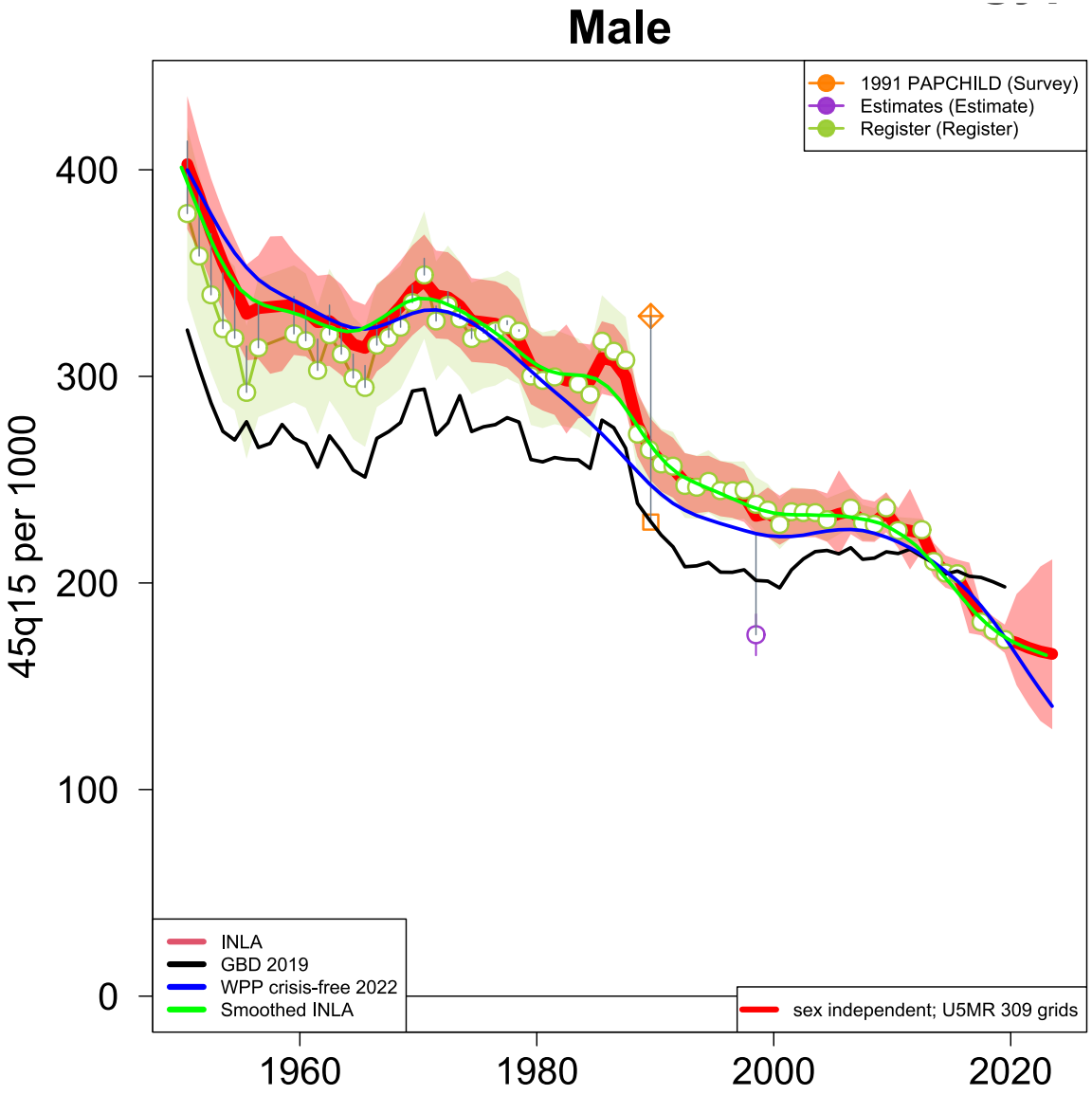
1. < age 5: use 1q0 and 5q0 from IGME (both sexes), compute 1q0 and 4q1 by sex
2. Age 5+: compute robust estimates of adult (and old-age) mortality
    1. Consolidate all mortality rates available for each data source and various estimation methods
    2. Compute summary indices (e.g., 45q15, etc.) using an adaptive relational standard to extrapolate ages if necessary
    3. Compute robust time trend in 45q15 by sex using a Bayesian B-spline Bias reduction model to account for various data biases and non-sampling errors
3. Compute abridged life tables using LogQuad model (n=33) or CD/UN model life table system (n=61) with 2-3 input parameters (e.g., 5q0 and 45q15 + optional 1q0) + for HIV/AIDS mortality countries use of adult HIV prevalence and ART coverage as applicable using SVD-Comp model (n=23).

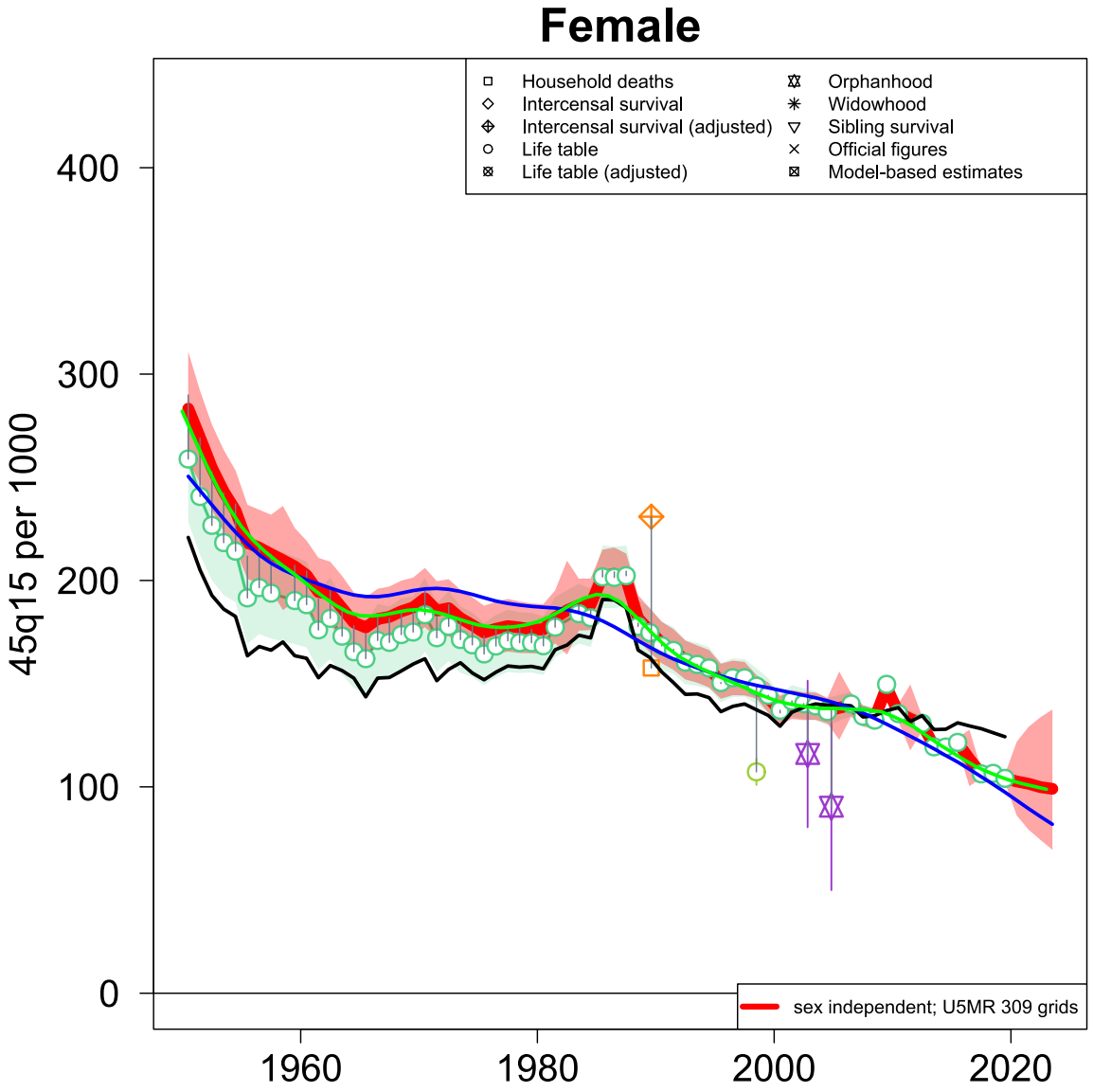
Under-five mortality time series



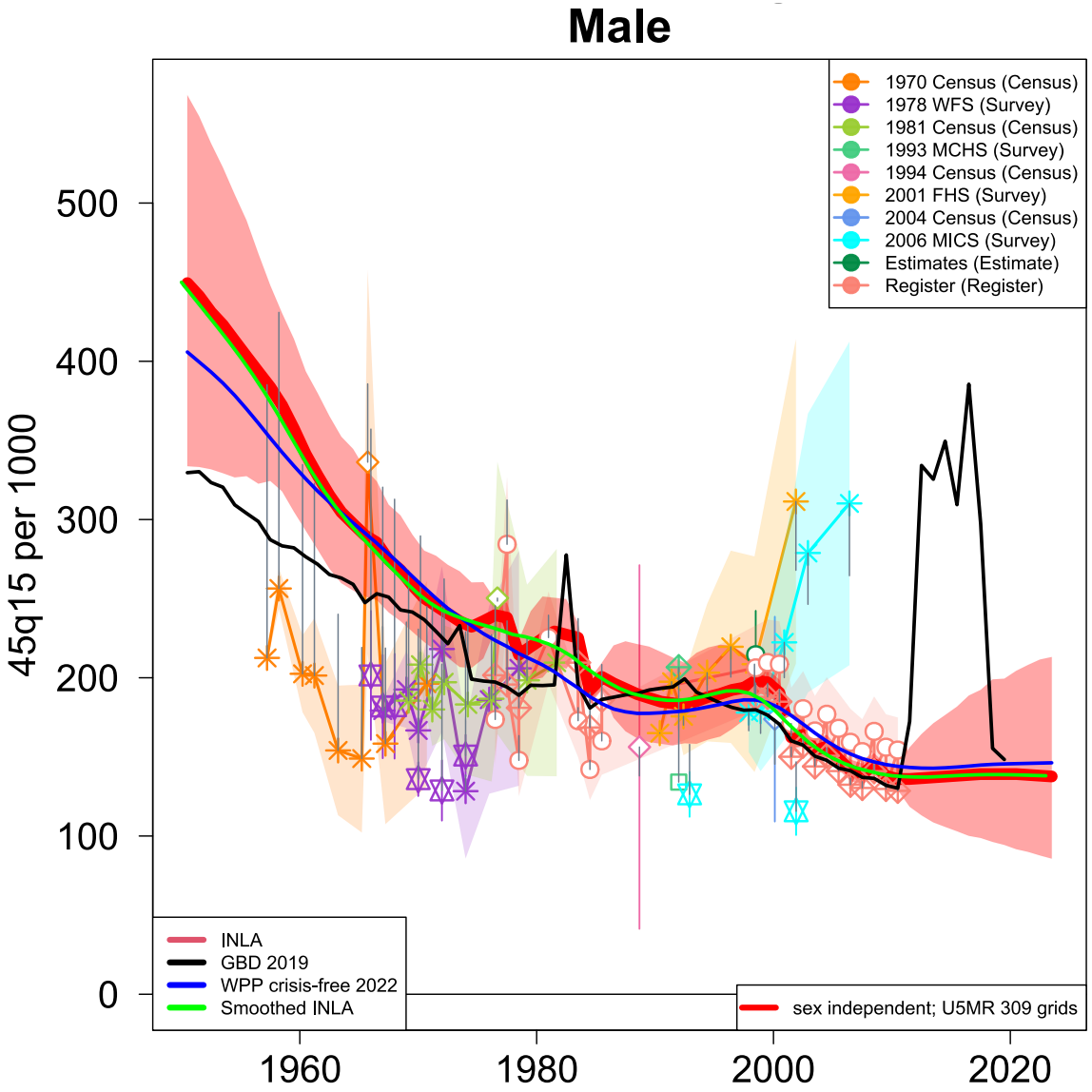
Source: https://childmortality.org/all-cause-mortality/data?refArea=EGY

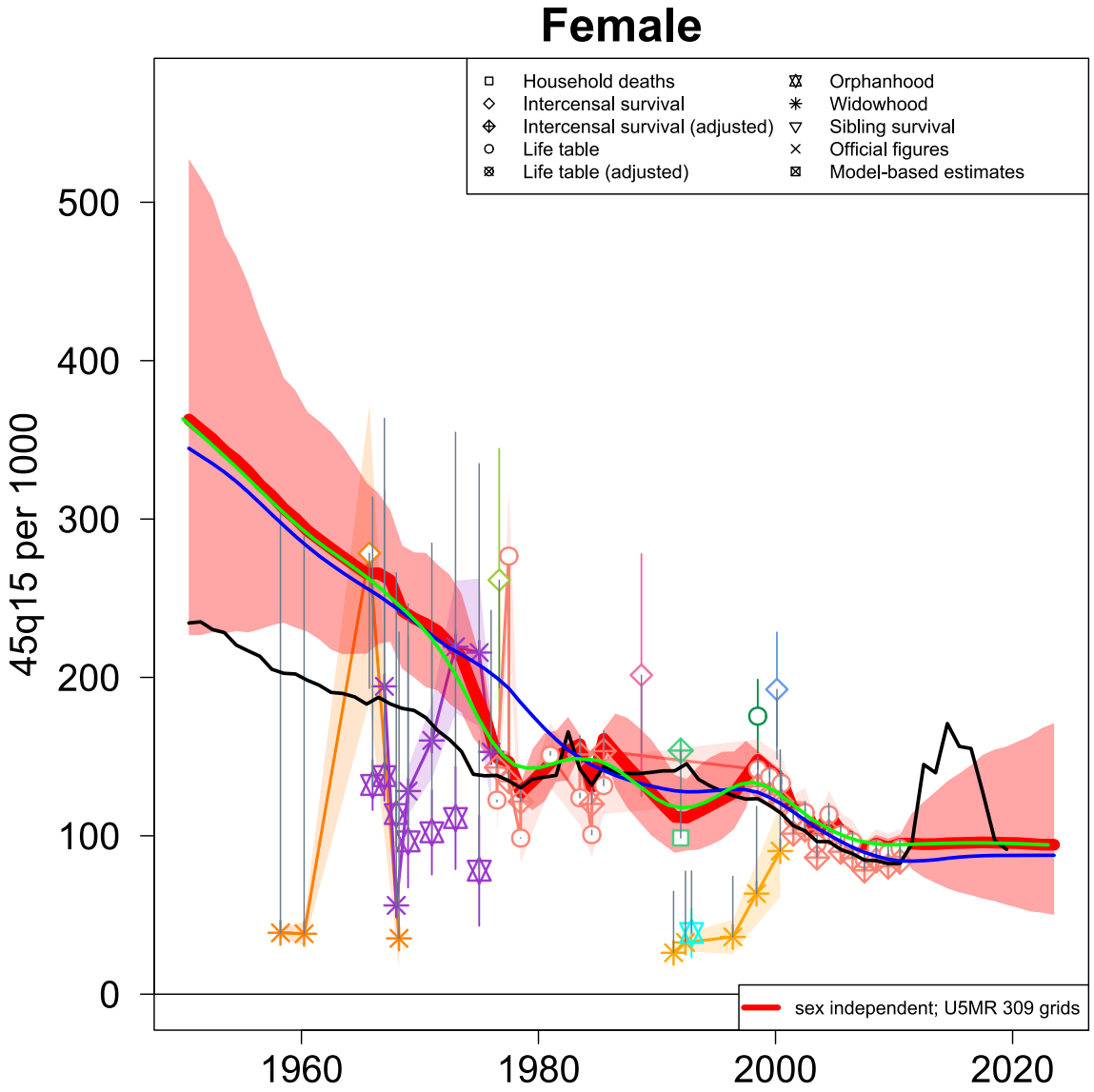
Adult mortality time series: Egypt





Adult mortality time series: Syria





Mortality shocks from conflicts, famines, natural disasters, epidemics, etc

1. **Annual total estimates (time series by country x year)**
2. **Distributions by age and sex (model-based)**

Background information:

- Consolidation of reference datasets for wars, genocides, battle deaths, conflicts, etc.
- Desk review and consolidation of mortality estimates from major famines, epidemics, and other major historical crises since 1950
- Consolidation of reference datasets for major types of natural disasters

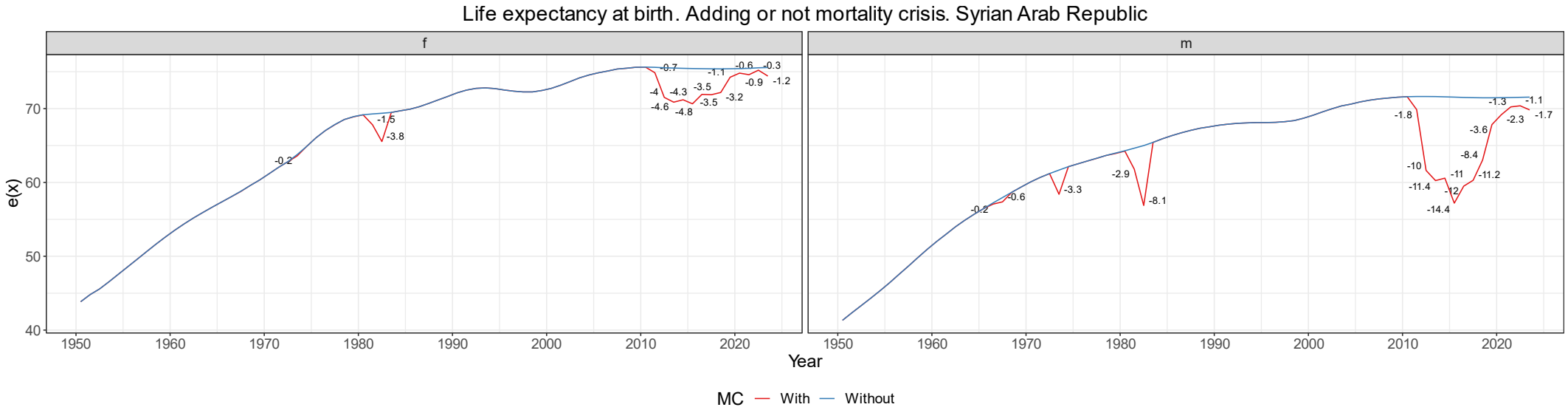
Annual time series of excess mortality due to crises

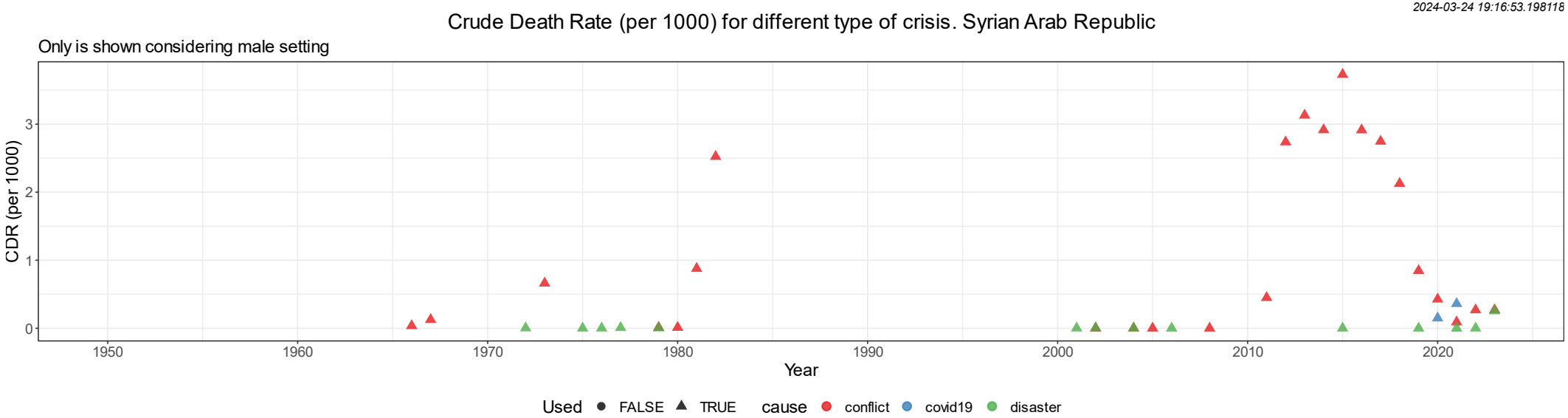
- 1950-2023: **7,306 country/area x years with mortality crises** out of 17,538 locations x years or about 41% (n=237 over 74 years)
- Conflicts and Battle deaths: 3,630 locations x years (n=174)
- Mass killings (including genocide): 82 locations x years (n=17)
- Floodings: 2,444 locations x years (n=163)
- Cyclones: 1,704 locations x years (n=172)
- Epidemics (not including HIV/AIDS and COVID-19): 2010 locations x years (n=151)
- Earthquakes: 1150 locations x years (n=124)
- COVID-19: 642 locations x years (n=229)
- Heat waves: 218 locations x years (n=56)
- Famines/Droughts: 183 locations x years (n=33)
- Tsunami: 45 locations x years (n=25)

About 66 million excess deaths:

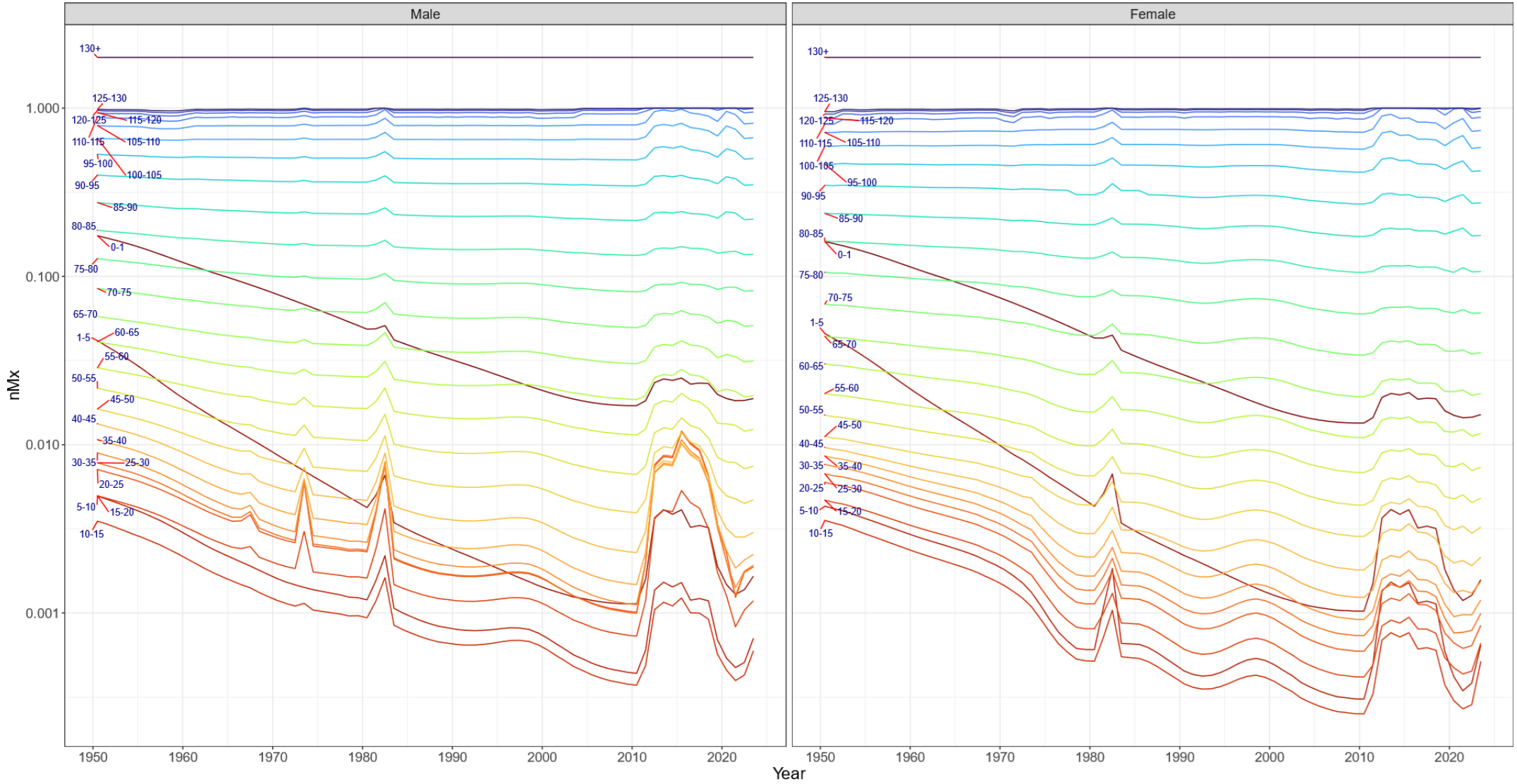
- **20.5 million due to conflicts (31%)**
- **20.6 million due to famines (31%)**
- **6.5 million due to natural disasters (10%)**
- **18.5 million due to COVID-19 (28%)**

Impact of mortality crises on life tables





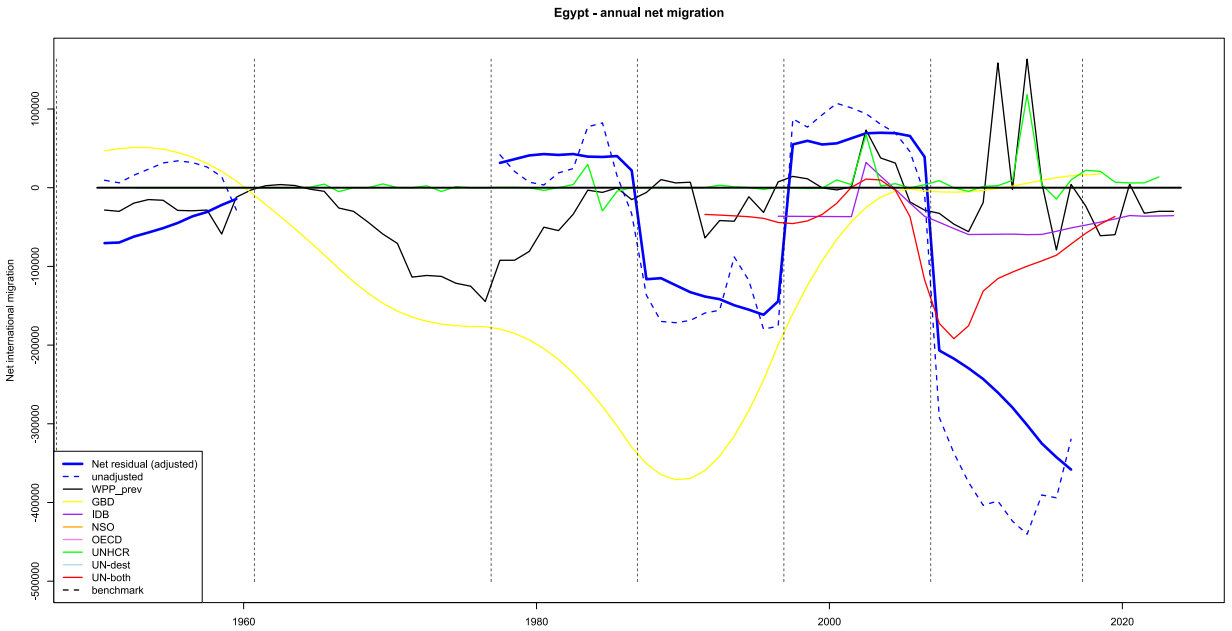
Time series of mortality rates by age groups

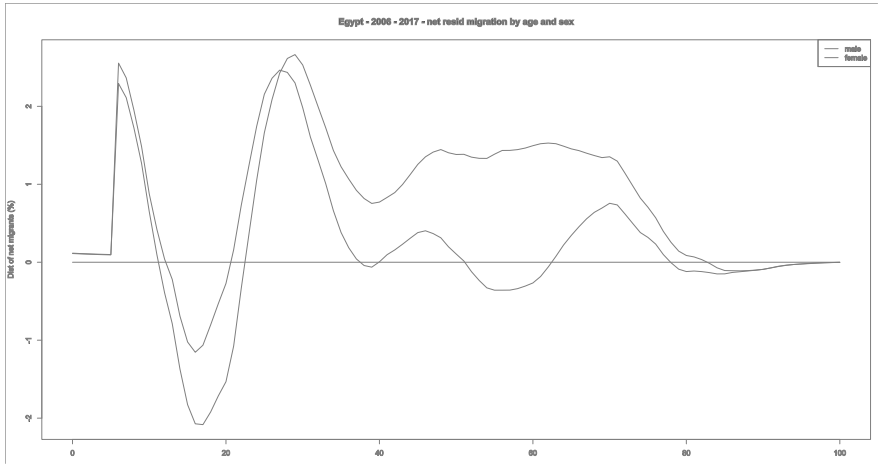


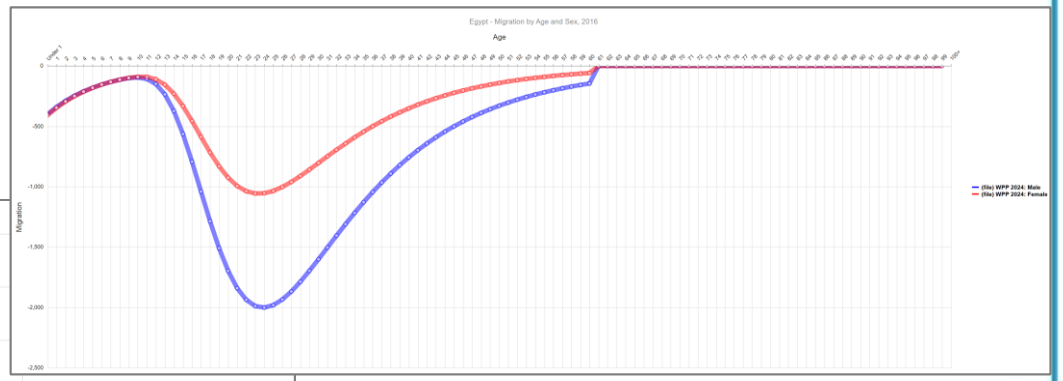
Net international migration estimates

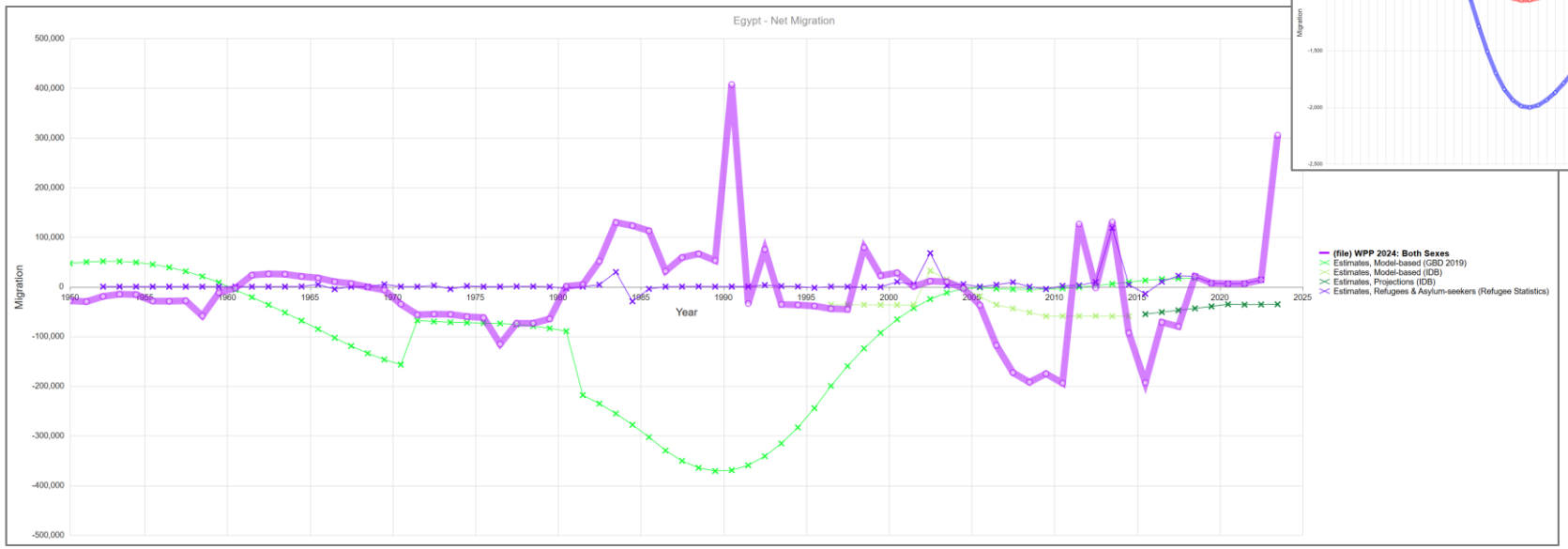
Data sources:

- Official estimates (e.g., NSOs, Eurostat)
- Estimates of migrant flows (e.g., Eurostat, OECD)
- Foreign-born stocks (UN estimates) and implied annual change
- Administrative data (e.g., work permits issued/renewed)
- Literature review for major forced migration historical events
- Intercensal net residual migration for countries with good VR or residual from cohort-component projection compared to census or population register
- UNHCR estimates of refugee stocks (and implied annual change)


Egypt - annual net migration
Net international migration
100000
0
-100000
-200000
-300000
-400000
-500000
1960
1980
2000
2020
Net residual (adjusted)
unadjusted
WPP_prev
GBD
IDB
NSO
OECD
UNHCR
UN-dest
UN-both
benchmark


Egypt - 2006 - 2017 - net resid migration by age and sex
male
female
Dist of net migrants (%)
2
1
0
-1
-2
0
20
40
60
80
100


Egypt - Migration by Age and Sex, 2016
Age
Migration
(file) WPP 2024: Male
(file) WPP 2024: Female


Egypt - Net Migration
Migration
500,000
400,000
300,000
200,000
100,000
0
-100,000
-200,000
-300,000
-400,000
-500,000
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
Year
(file) WPP 2024: Both Sexes
Estimates, Model-based (GBD 2019)
Estimates, Model-based (IDB)
Estimates, Projections (IDB)
Estimates, Refugees & Asylum-seekers (Refugee Statistics)

Net international migration

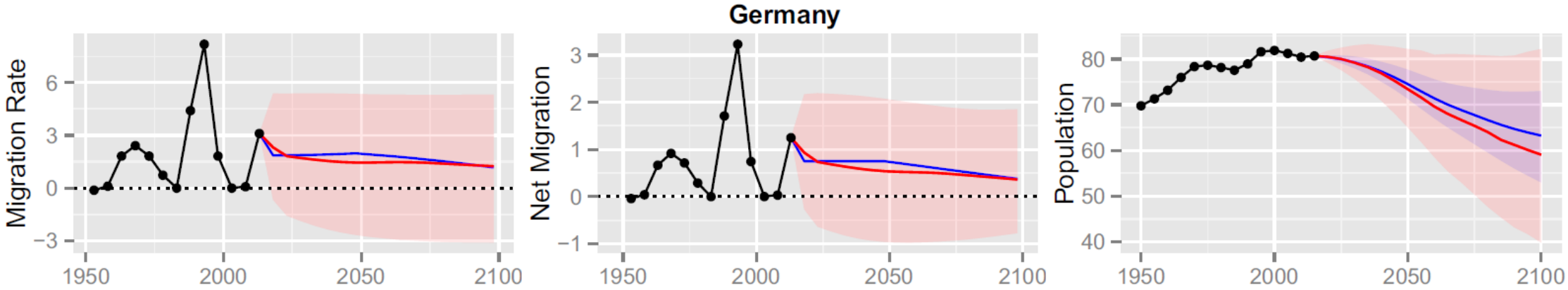
Other issues/challenges:

- End-of-period (migrants not exposed to fertility and mortality) vs. evenly-over-period (and half of the migrants are exposed)
- Annual estimates vs. intercensal periods
- Age/sex patterns:
  - Net residual (as-is or smoothed)
  - Rogers-Castro model-based (based on dominant characteristic: Family, Female Labor, Male Labor) and new model-based patterns for projections (Raymer, 2023)
  - Population distribution
  - Actual data (only for countries with flows data by age/sex)
- World balancing: each location initially estimated independently, but ultimately for each period the sum of all net migrations across all locations must add to 0

Raymer, J., Q. Guan, T. Shen, S. Hertog, P. Gerland (2023). *Modelling the age and sex profiles of net international migration*. United Nations, DESA, Population Division, Technical paper No. 7.

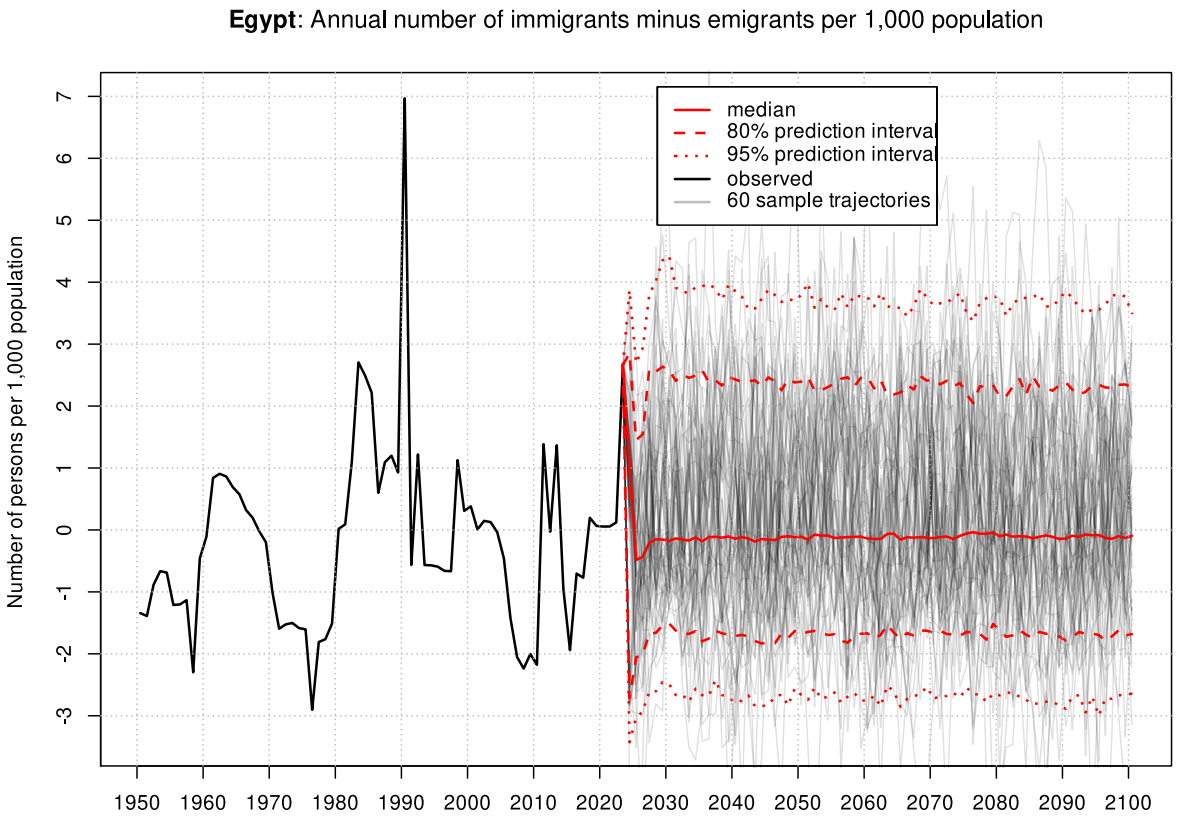
WPP 2024: probabilistic projections of international migrations

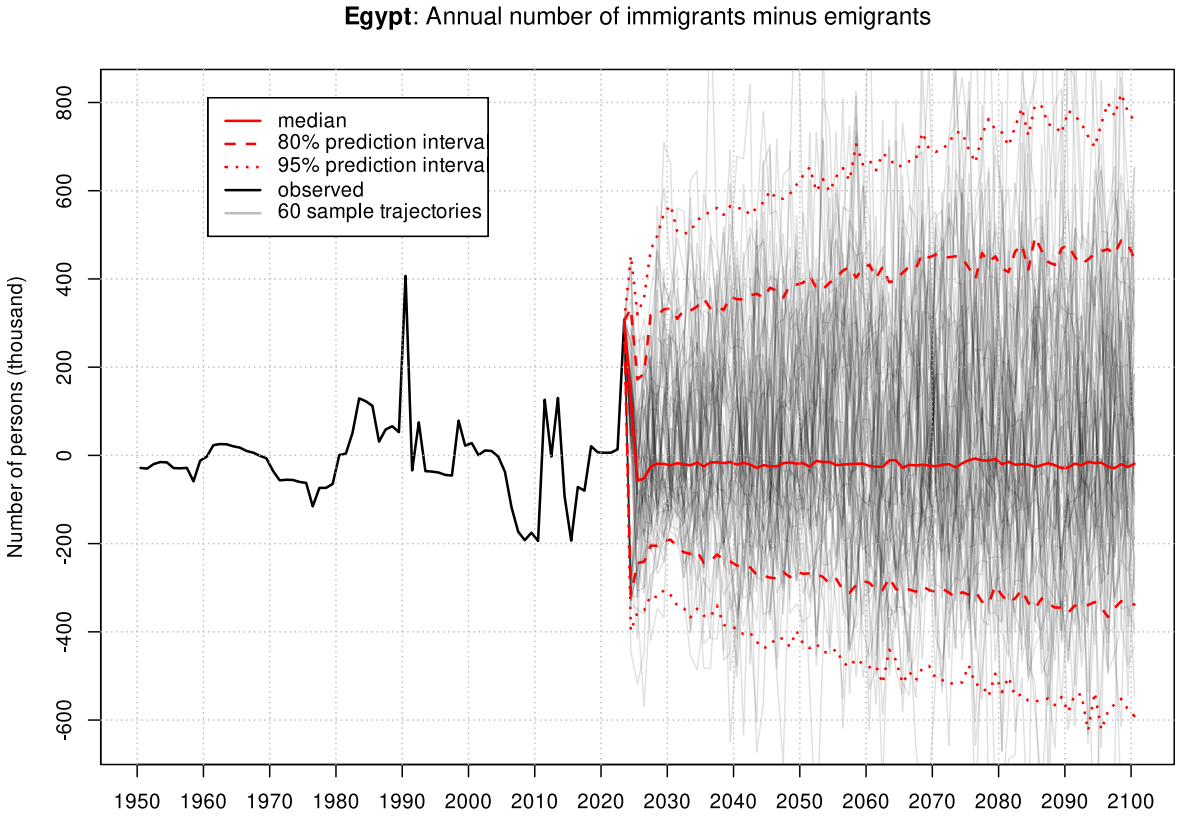
- Building on the methodological work of Azose and Raftery (2015, 2016) adapted to use annual time series and single age population data
- Use past levels and trends, and reflect past variability in the projection of future crude net migration rates, and incorporate uncertainty in future international migrations into probabilistic population projections in addition to uncertainty in future fertility and mortality.



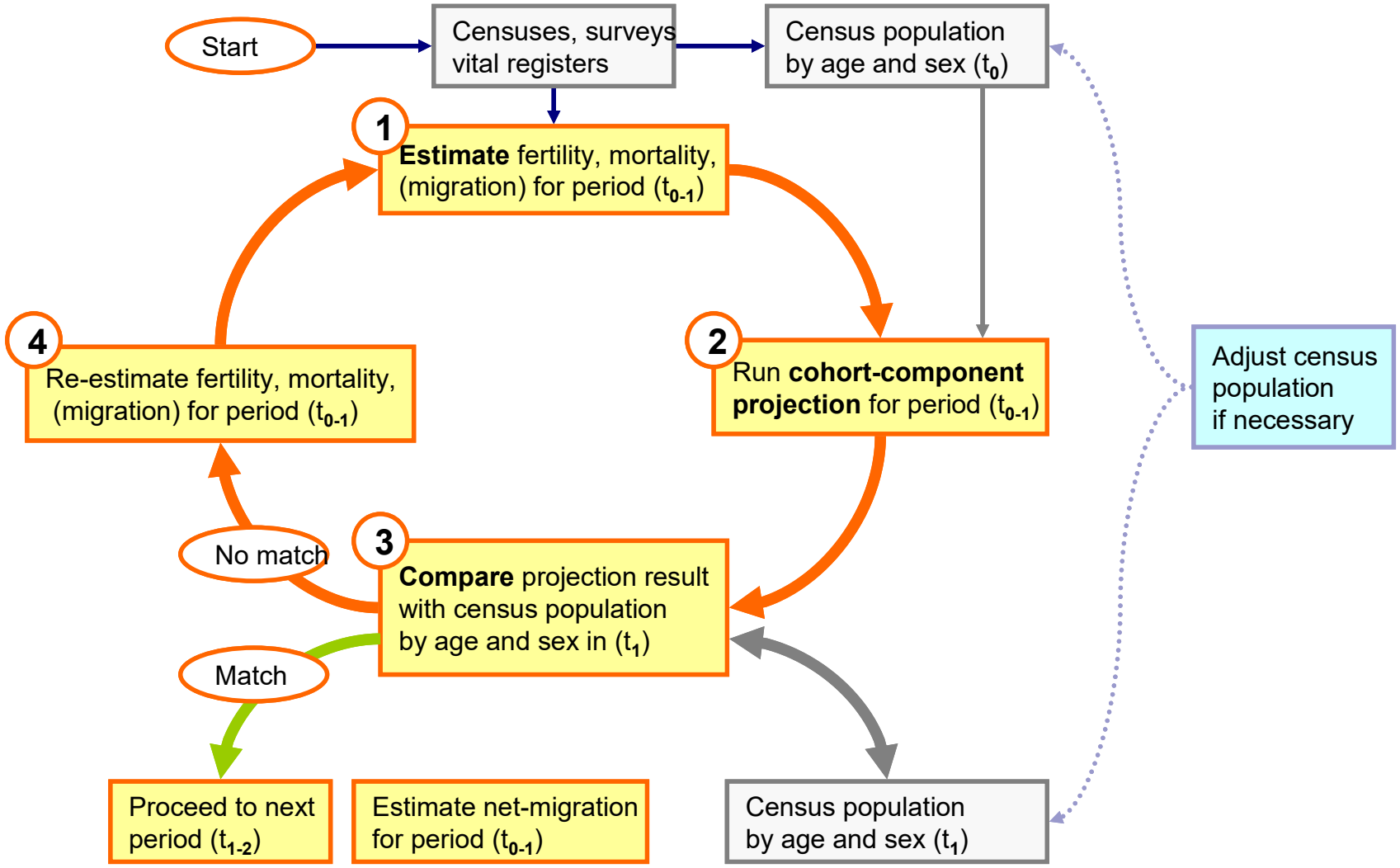
Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences*, *113*(23), 6460-6465.

Example of projection of net migration rate: Egypt





WPP estimation process for each country/area



See *WPP 2024 methodological report* for further details

UN cohort-component projections to cross-validate estimations

Remember… over the last 74 years:

- 87% of countries have ≥ 3 censuses, 77% have ≥ 4 and about half of the countries have ≥ 5.
- Each census can be evaluated by itself, and in the context of the previous ones by following birth cohorts over time.

With at least 3-5 censuses in the last decades…

1. Start cohort-component projections with base year in 1950 for all 237 countries/areas.
2. Insure full internal consistency by age and sex between all demographic components and enumerated populations.

United Nations | Department of Economic and Social Affairs

Part 5

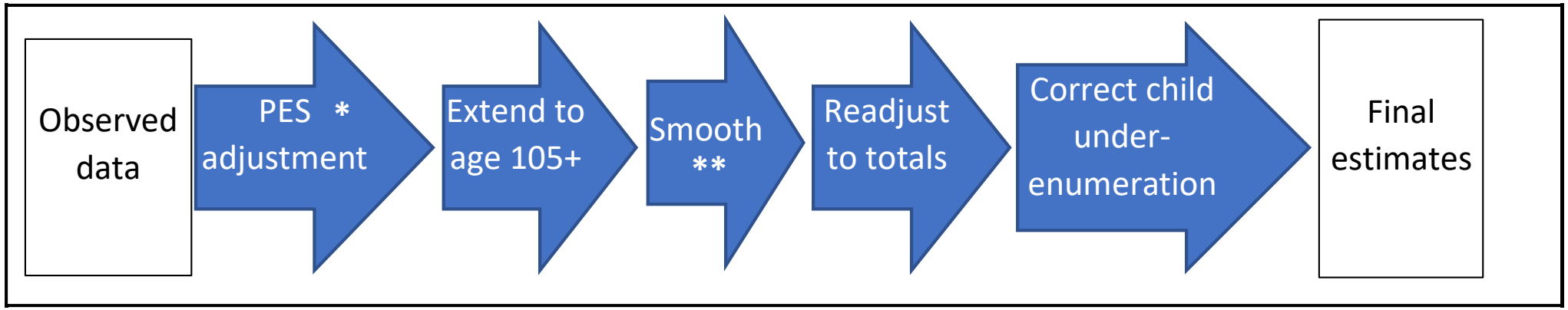
Method protocol for the evaluation of census population data by age and sex

Census population evaluation/adjustment

1. **Territorial and population coverage issues** (e.g., UNSD DYB footnotes and descriptive notes)
2. **Statistical concept**: de-facto vs. de-jure or usual resident
3. **Data status**: provisional, final, adjusted
4. **Completeness evaluation and adjustment** based on Post-Enumeration Surveys or model-based expected under/over count (total and by age/sex)
5. **Age/sex distribution analysis** (for heaping and systematic distortions) and smoothing if required + graduation if applicable and extension to 100 if required
6. **Demographic analysis** (reverse survival based on US Census BASEPOP method) to evaluate undercount of children under age 15, and for adjustment if required [+ intercensal cohort analysis and cohort-component projection]
7. **External validation** for selected age groups based on independent register-based administrative headcounts (e.g., education statistics, immunizations, electoral rolls) and for relative age distributions using household surveys

See *WPP 2024 methodological report*, *and Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

Overview of Population Evaluation Process

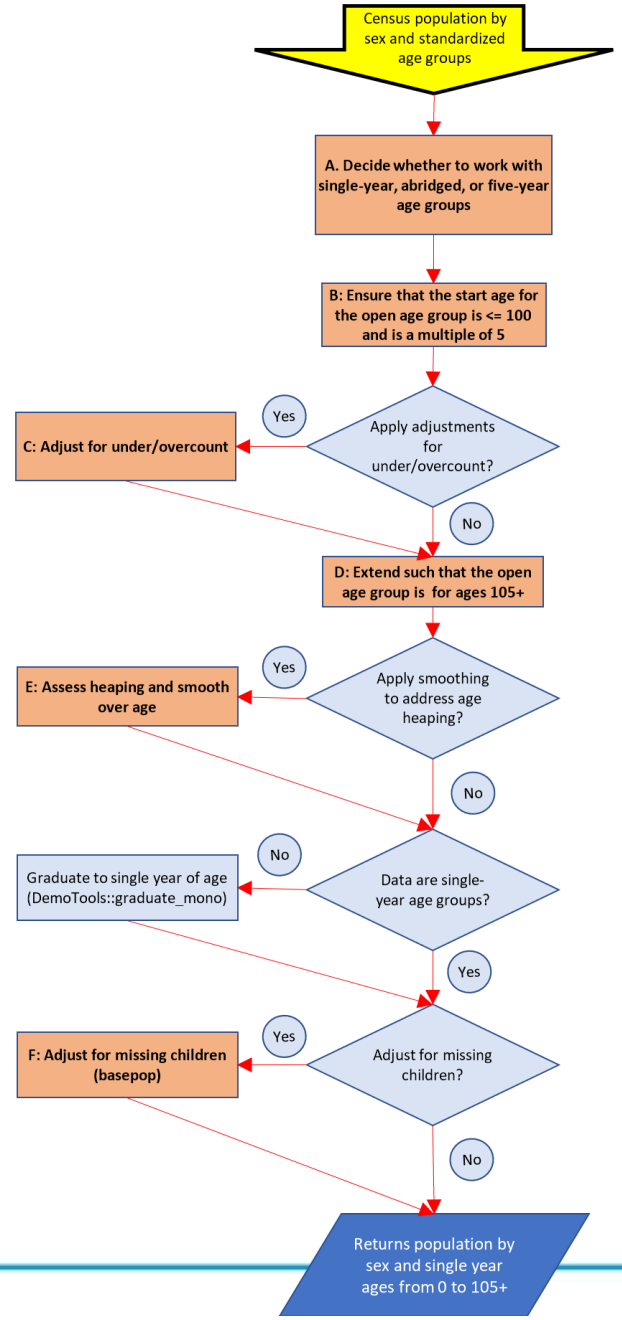


* + additional territorial coverage adjustment as applicable for specific censuses

** adaptative degree of smoothing based on digit preference and age heaping + graduation into single age as applicable

See *WPP 2024 methodological report*, *and Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

Overview of Population Evaluation and adjustment Process



Adjustment of population for under/over count

- **Completeness of enumeration**: Census total population / Adjusted total population (or "true" population, i.e., the population estimated from the PES multiplied by the population from the census after correcting for erroneous inclusions and divided by matched population between the census and the PES)
- **Net Census Error (NCE)**: completeness of enumeration – 1
- Developed a new model for the **overall net enumeration error level** (based on **over 320 PES** and over 120 DA estimates of the net census error for selected censuses covering **130 countries between 1946 and 2022**):

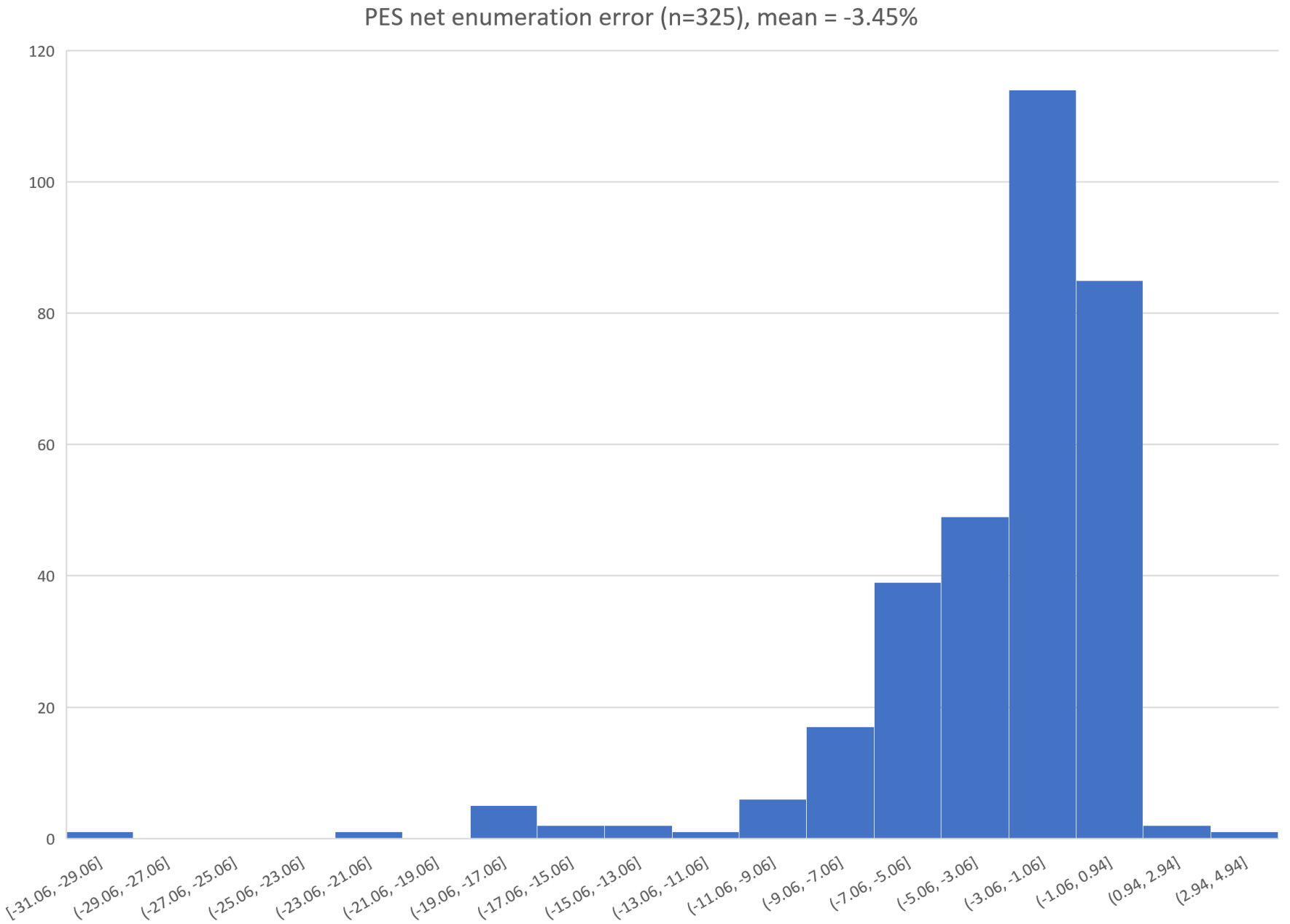
$$NCE_{ijkl} = \beta_0 + \beta_1 PES_{ijkl} + \beta_2 EducYrsM_{ijkl} + \beta_3 EducYrsF_{ijkl} + \beta_4 logLDI_{ijkl} + \beta_5 logQ5_{ijkl} + u_j + v_k + w_l + e_{ijkl}$$

- where ijkl is an I observation nested within country j nested within sub-region k and SDG region l, and

$e_{ijkl} \sim N(0, \sigma_e^2)$, $u_j \sim N(0, \sigma_u^2)$, $v_k \sim N(0, \sigma_v^2)$, $w_l \sim N(0, \sigma_w^2)$

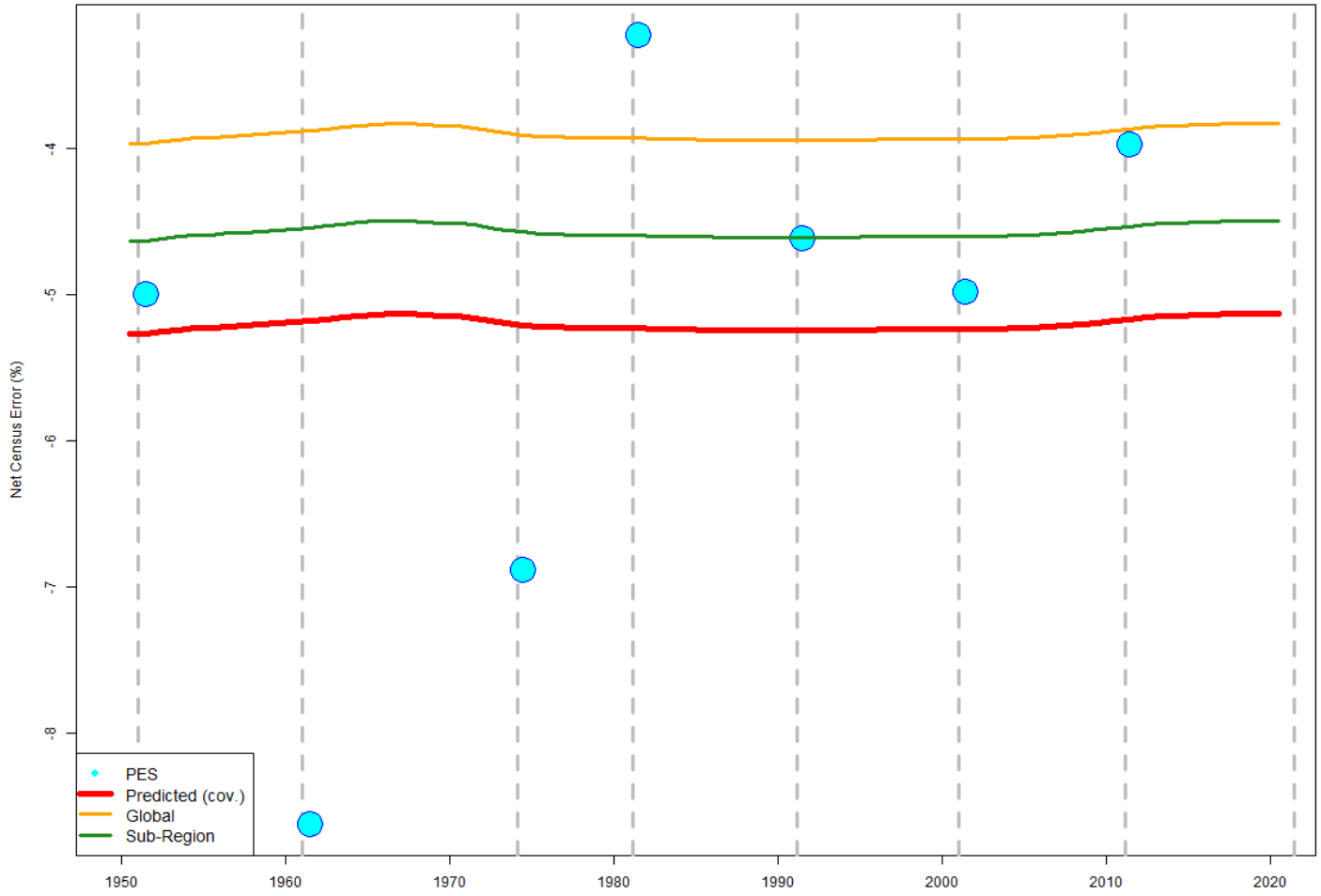
- with PES a dummy variable equal to 1 if the observation is based on PES or 0 otherwise if it is based on Demographic Analysis, EducYrsM and EducYrsF are respectively the male and female average number of years of education, logLDI is the log transformed Lag-distributed Income, and logQ5 is the log transformed Under-Five probability of dying between birth and age 5.

See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details



See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

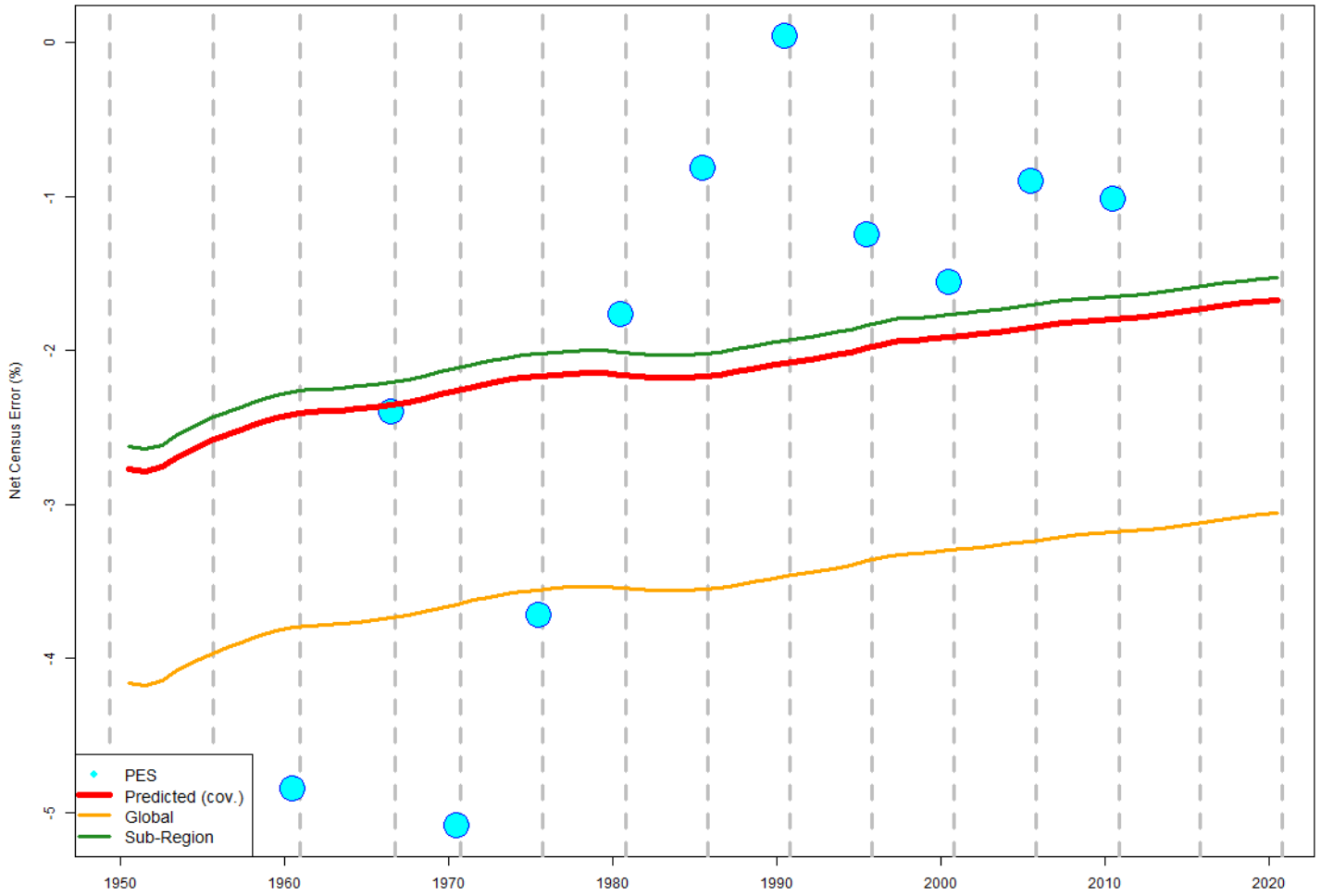
Observed and Predicted PES Net Census Errors for Bangladesh



The respective years with censuses are plotted with vertical gray dash lines, the PES estimates are shown as blue circles, and DA estimates as green squares (upon availability). The predicted (or expected) net census error (NCE) is show for (1) the country-specific expected values as bold red line, and only as baseline reference for informative purpose for (2) the UN sub-region shown as green line, and (3) the overall global model (i.e., world) shown as yellow line.

See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

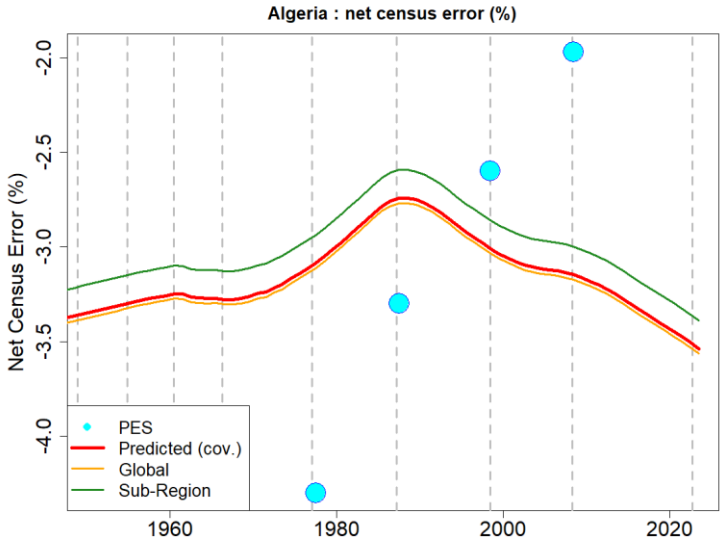
Observed and Predicted PES Net Census Errors for Republic of Korea

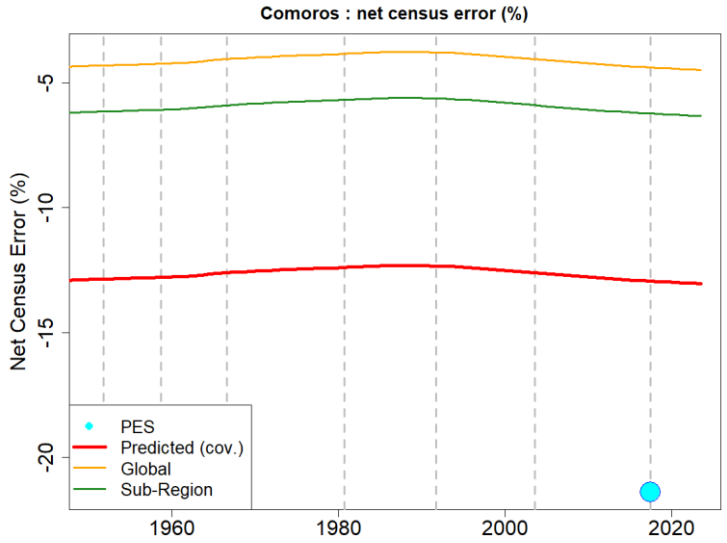


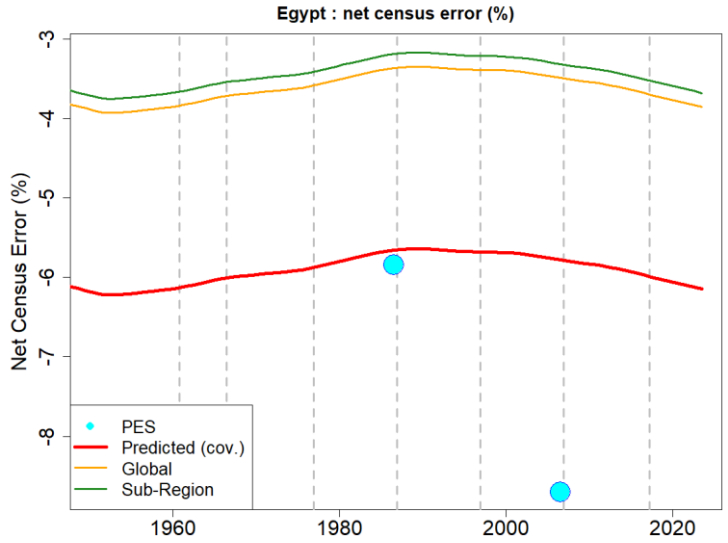
The respective years with censuses are plotted with vertical gray dash lines, the PES estimates are shown as blue circles, and DA estimates as green squares (upon availability). The predicted (or expected) net census error (NCE) is show for (1) the country-specific expected values as bold red line, and only as baseline reference for informative purpose for (2) the UN sub-region shown as green line, and (3) the overall global model (i.e., world) shown as yellow line.

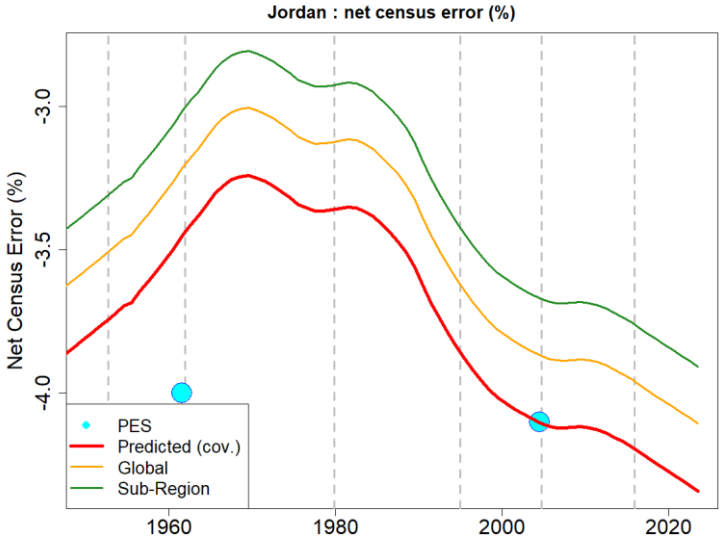
See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

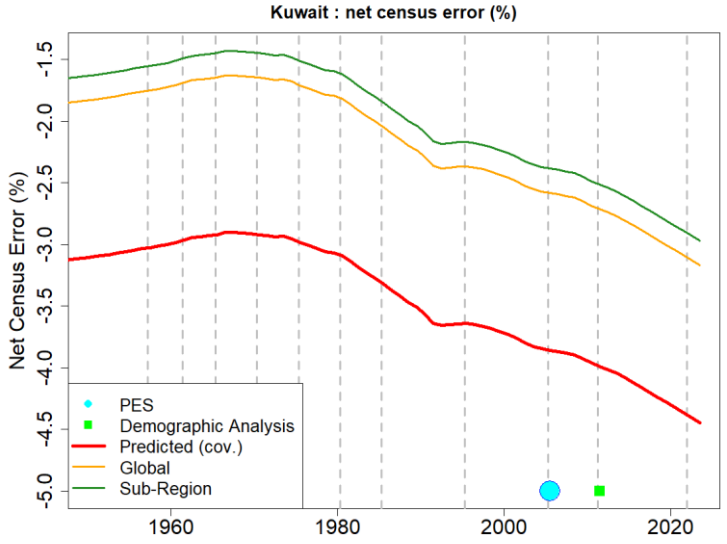
Observed and Predicted PES Net Census Errors

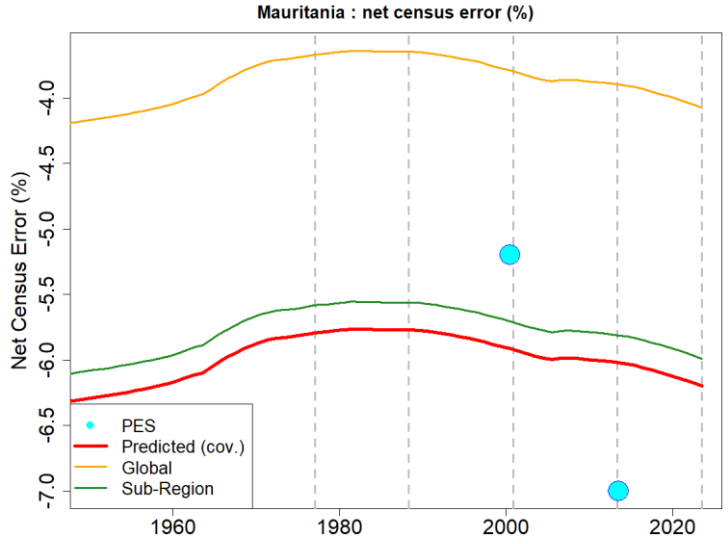






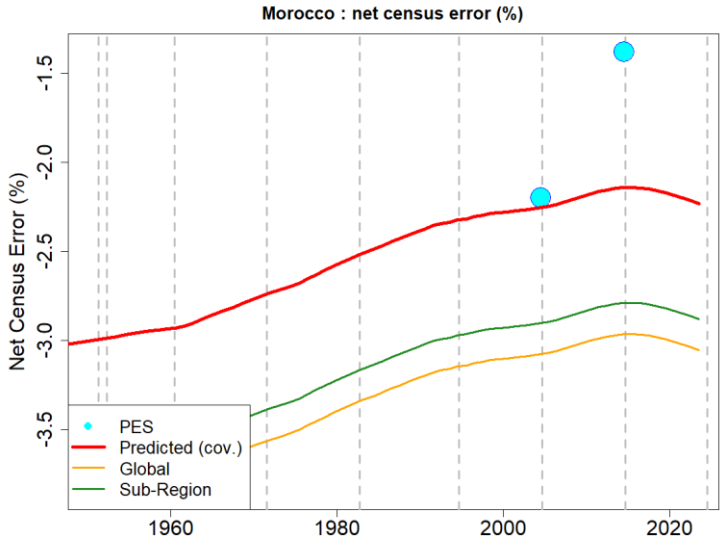


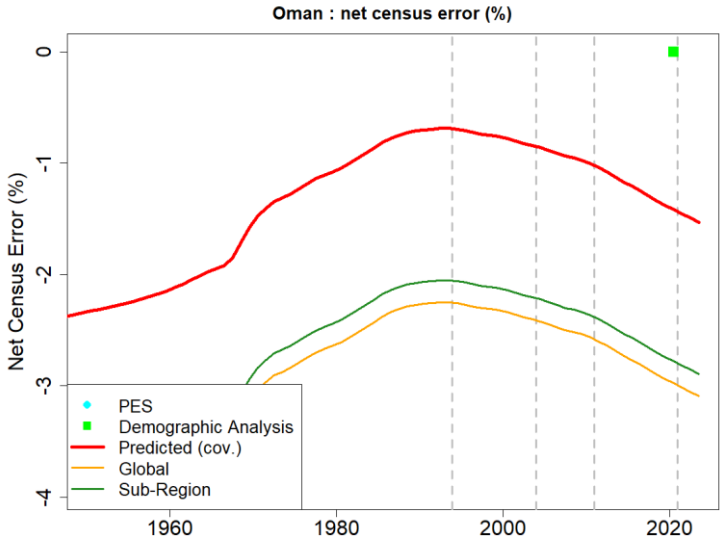


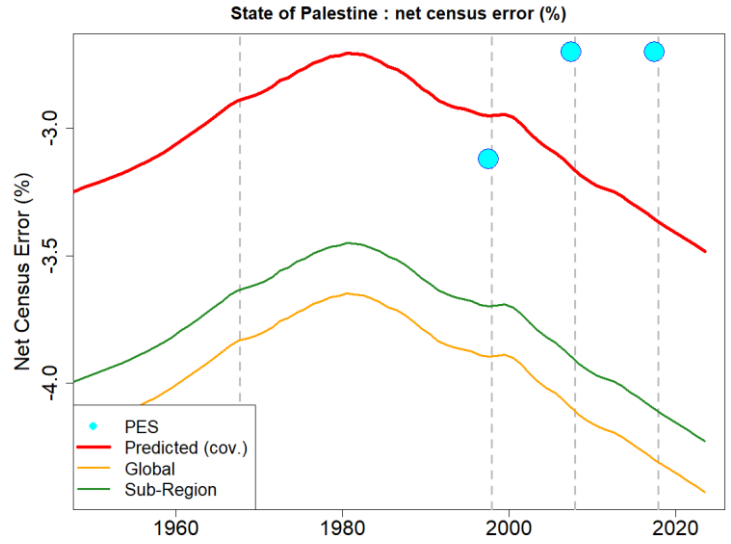


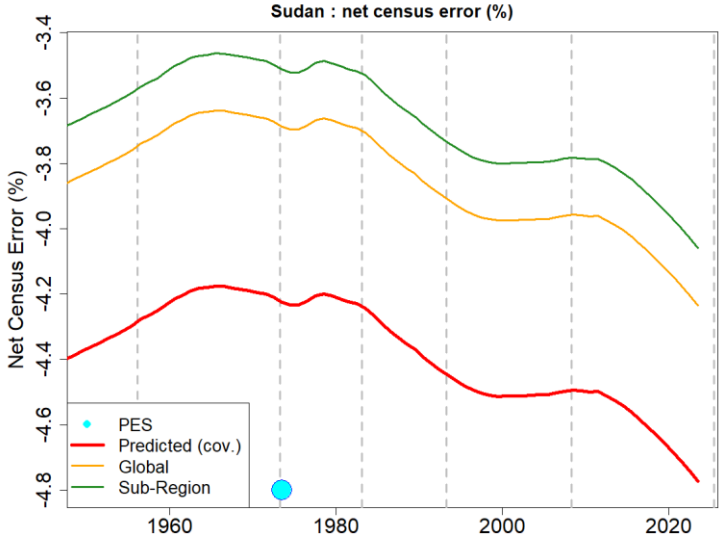
See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

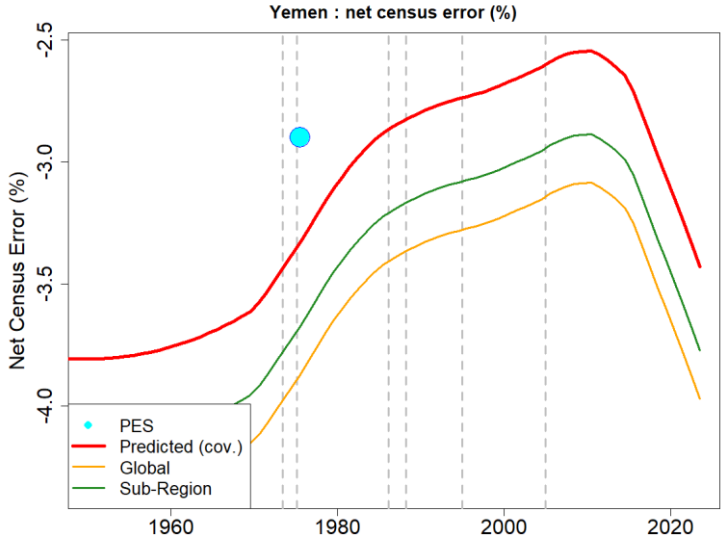
Observed and Predicted PES Net Census Errors











See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

Model of Differences in Net Census Errors from the Overall Level, by Age and Sex

Knowing the **sex-specific NCE** (respectively NCE_M and NCE_F for males and females) for a smaller subset of observations than for the overall total, we can compute the **sex-specific difference** (e.g., NCE_M_Diff = (NCE_M – NCE) for male) for this subset (about 100 censuses), and the following analytical forms were fitted on the data by sex:

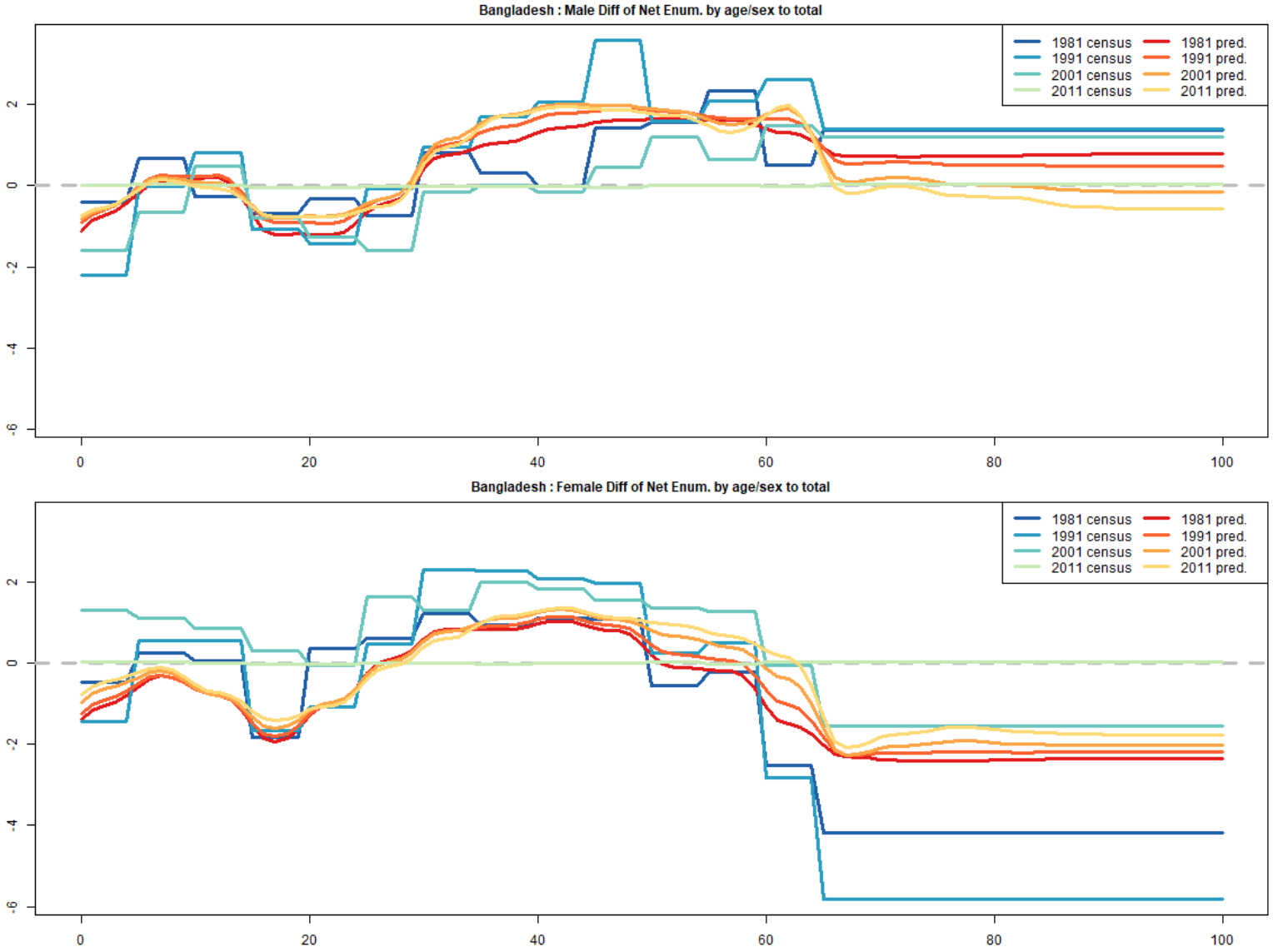
- For males: $\text{NCE\_M\_Diff}_{ijkl} = \beta_0 + \beta_1 NCE_{ijkl} + \beta_2 PES_{ijkl} + \beta_3 EducYrsM_{ijkl} + \beta_4 logLDI_{ijkl} + \beta_5 logQ5_{ijkl} + u_j + v_k + w_l + e_{ijkl}$
- For females: $\text{NCE\_F\_Diff}_{ijkl} = \beta_0 + \beta_1 NCE_{ijkl} + \beta_2 PES_{ijkl} + \beta_3 EducYrsF_{ijkl} + \beta_4 logLDI_{ijkl} + \beta_5 logQ5_{ijkl} + u_j + v_k + w_l + e_{ijkl}$

Finally a model of the differences in the level of **net enumeration by sex and age** (based on PES data from 56 censuses for 28 countries) was fitted on the subset of censuses with such information, and predicted values were estimated for all locations from 1950 up to 2020 based on covariates. The following analytical forms were fitted on the data by sex:

- For males: $\text{NCE\_M\_Diff\_x}_{ijkl} = \beta_0 + \beta_1 NCE_{ijkl} + \beta_2 \text{NCE\_M\_Diff}_{ijkl} + \beta_3 Age5 + \beta_4 PES_{ijkl} + \beta_5 (Age5 * EducYrsM_{ijkl)} + \beta_6 (Age5 * logLDI)_{ijkl} + \beta_5 (Age5 * logQ5)_{ijkl} + u_j + v_k + w_l + e_{ijkl}$
- For females: $\text{NCE\_F\_Diff\_x}_{ijkl} = \beta_0 + \beta_1 NCE_{ijkl} + \beta_2 \text{NCE\_F\_Diff}_{ijkl} + \beta_3 Age5 + \beta_4 PES_{ijkl} + \beta_5 (Age5 * EducYrsM_{ijkl)} + \beta_6 (Age5 * logLDI)_{ijkl} + \beta_5 (Age5 * logQ5)_{ijkl} + u_j + v_k + w_l + e_{ijkl}$

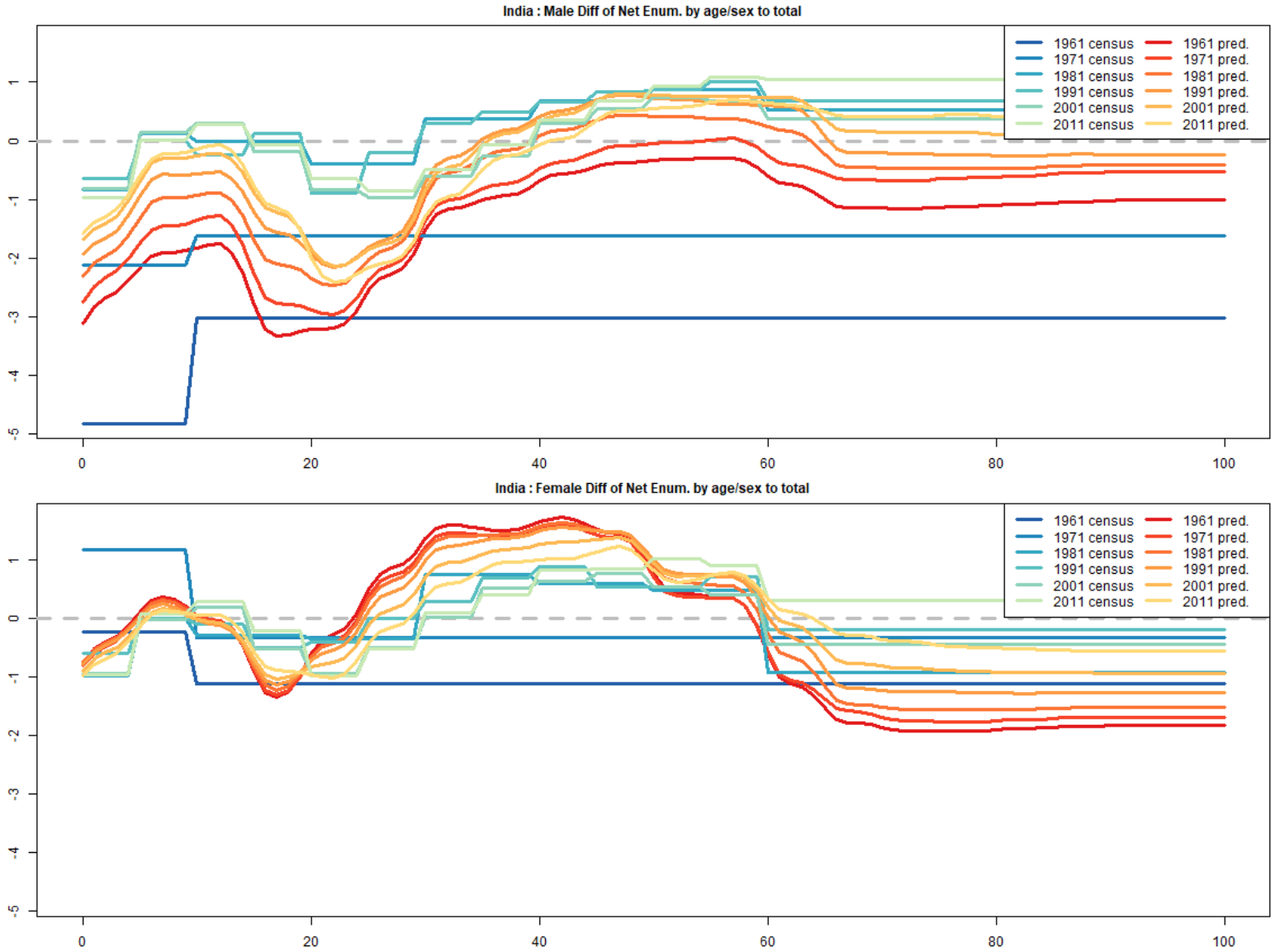
See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

Difference in Net Census Error by Sex and Age for Bangladesh



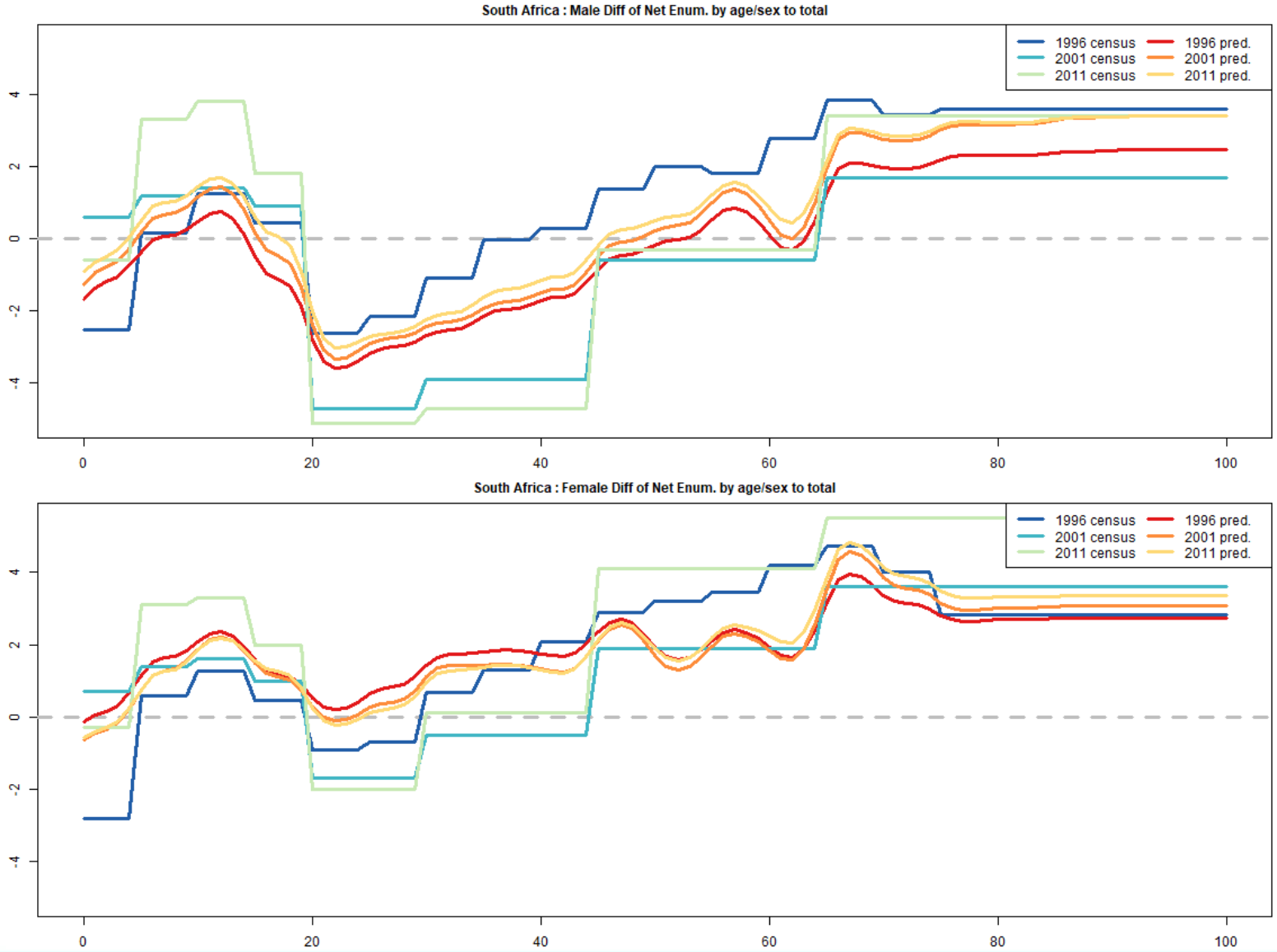
See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

Difference in Net Census Error by Sex and Age for India



See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

Difference in Net Census Error by Sex and Age for South Africa



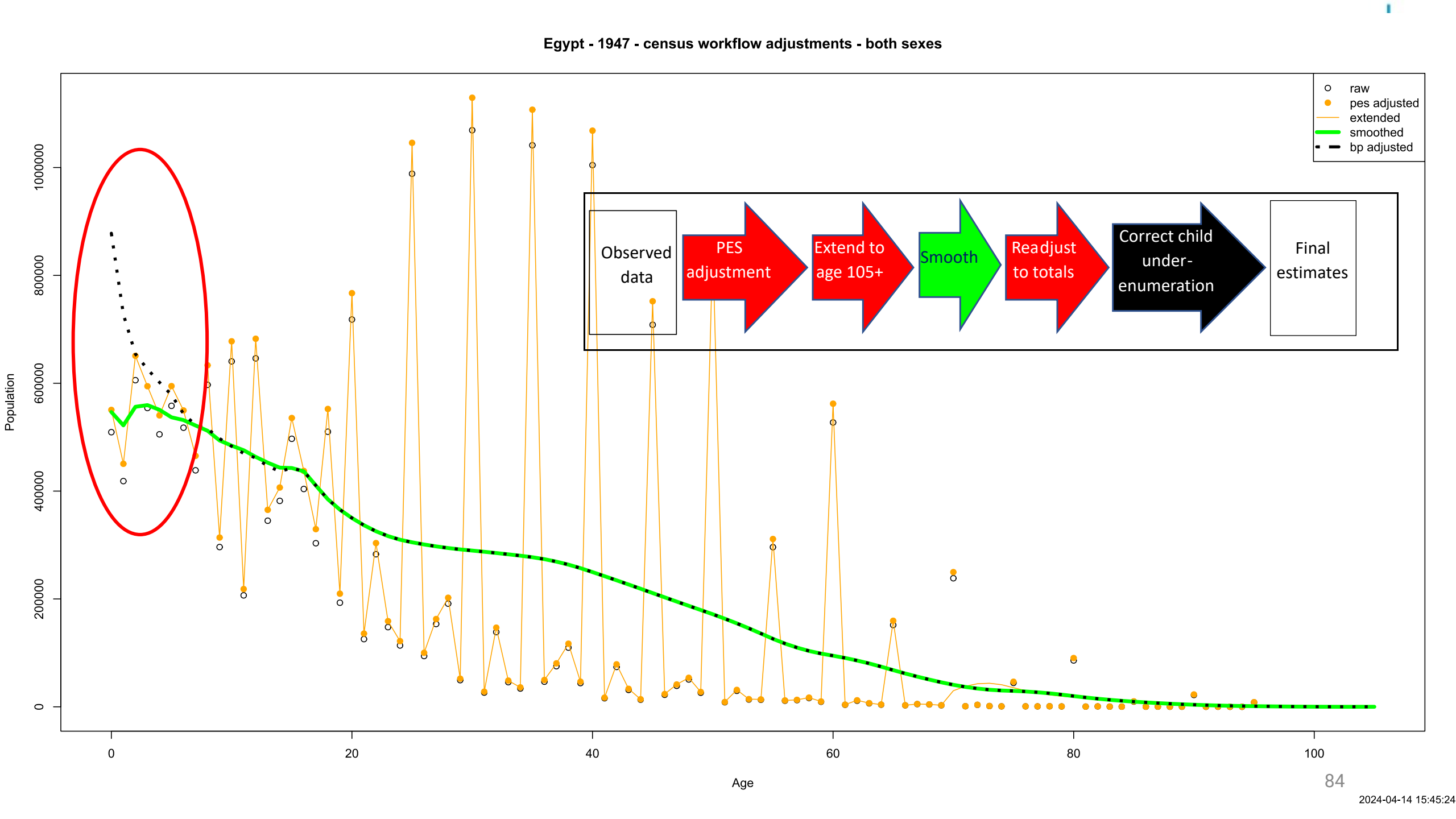
See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

United Nations | Department of Economic and Social Affairs

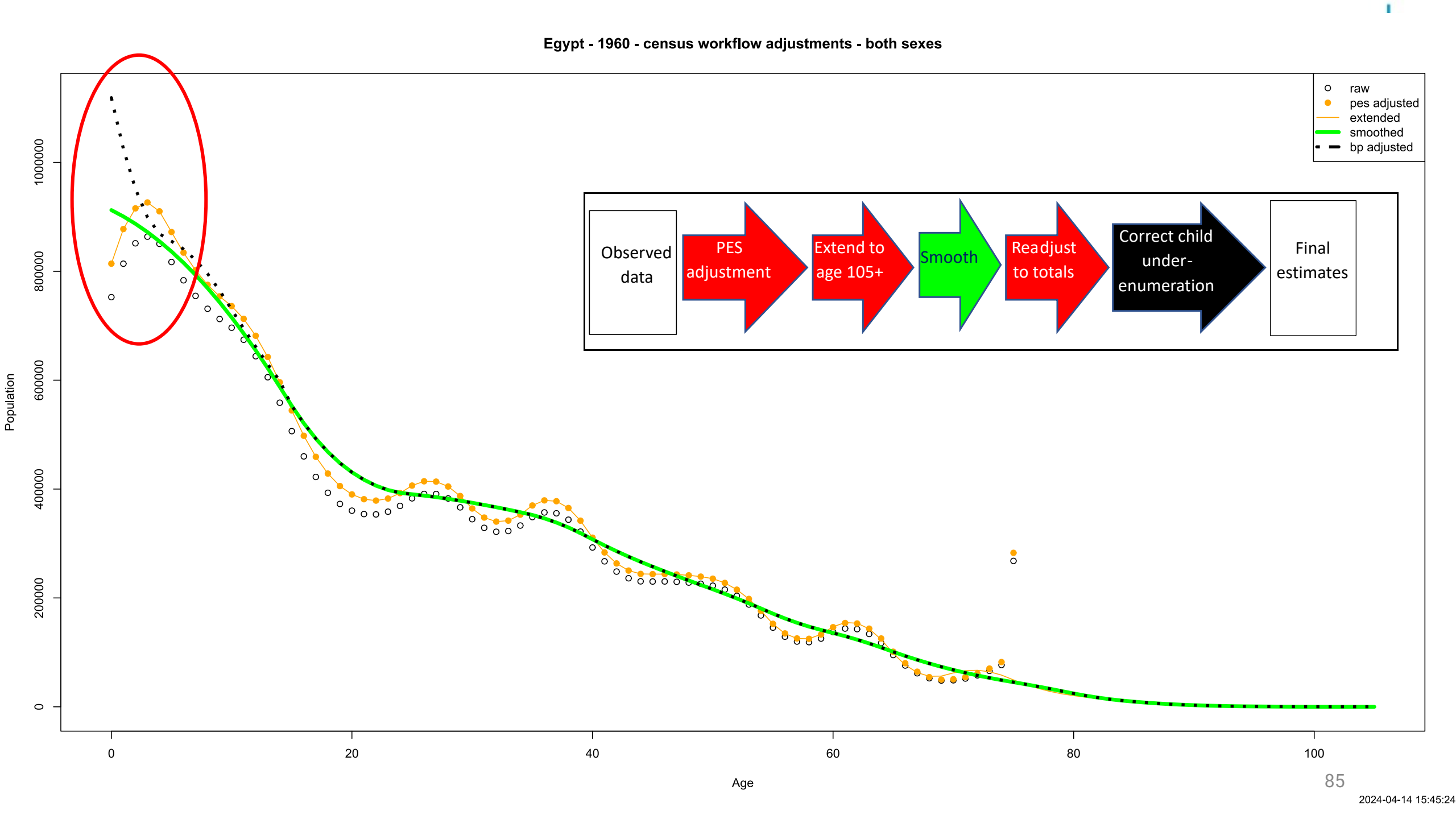
Example of application of Population Census Evaluation Process

Egypt

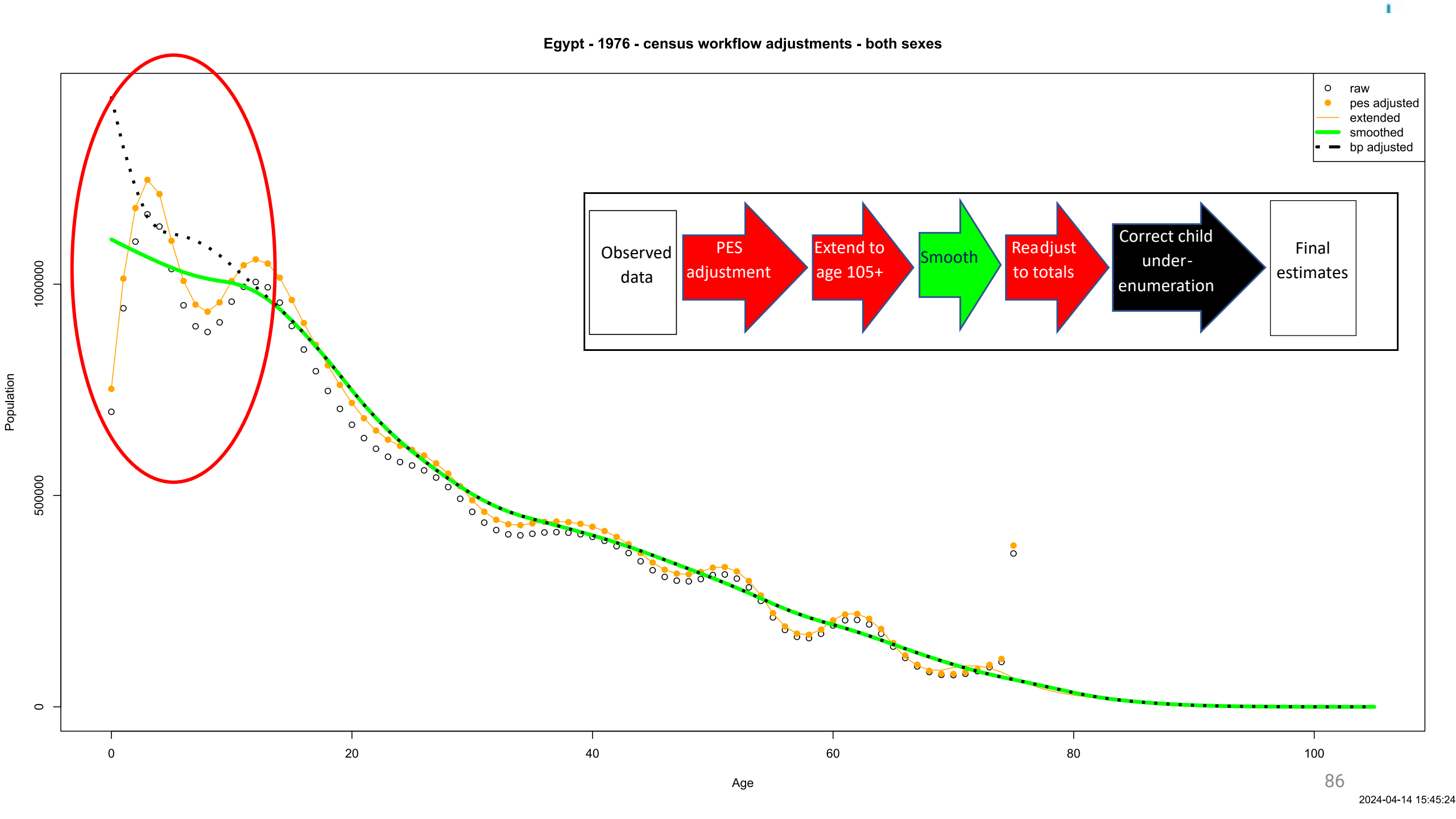
Short Name	Reference Period Start	Reference Period End	Census unadjusted	Census PES adjusted	Census age heaping adjusted	Census children <15 adjusted
2017 Census	2017	2017	Used	Used	Used	Excluded
2006 Census	2006	2006	Used	Used	Used	Used
1996 Census	1996	1996	Used	Used	Used	Used
1986 Census	1986	1986	Used	Used	Used	Used
1976 Census	1976	1976	Used	Used	Used	Used
1966 Census	1966	1966				
1960 Census	1960	1960	Used	Used	Used	Used
1947 Census	1947	1947	Used	Used	Used	Used



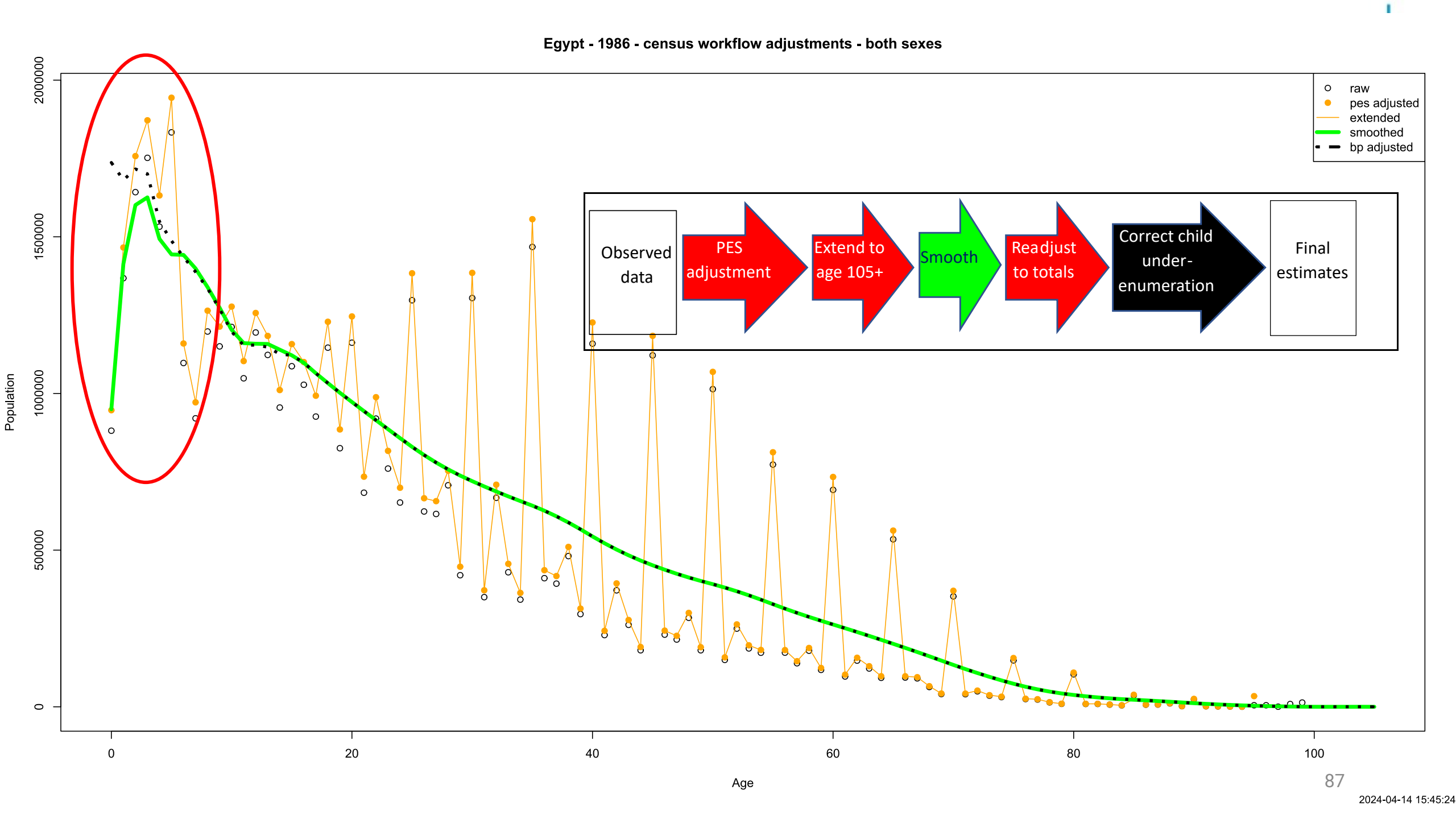
2024-04-14 15:45:24



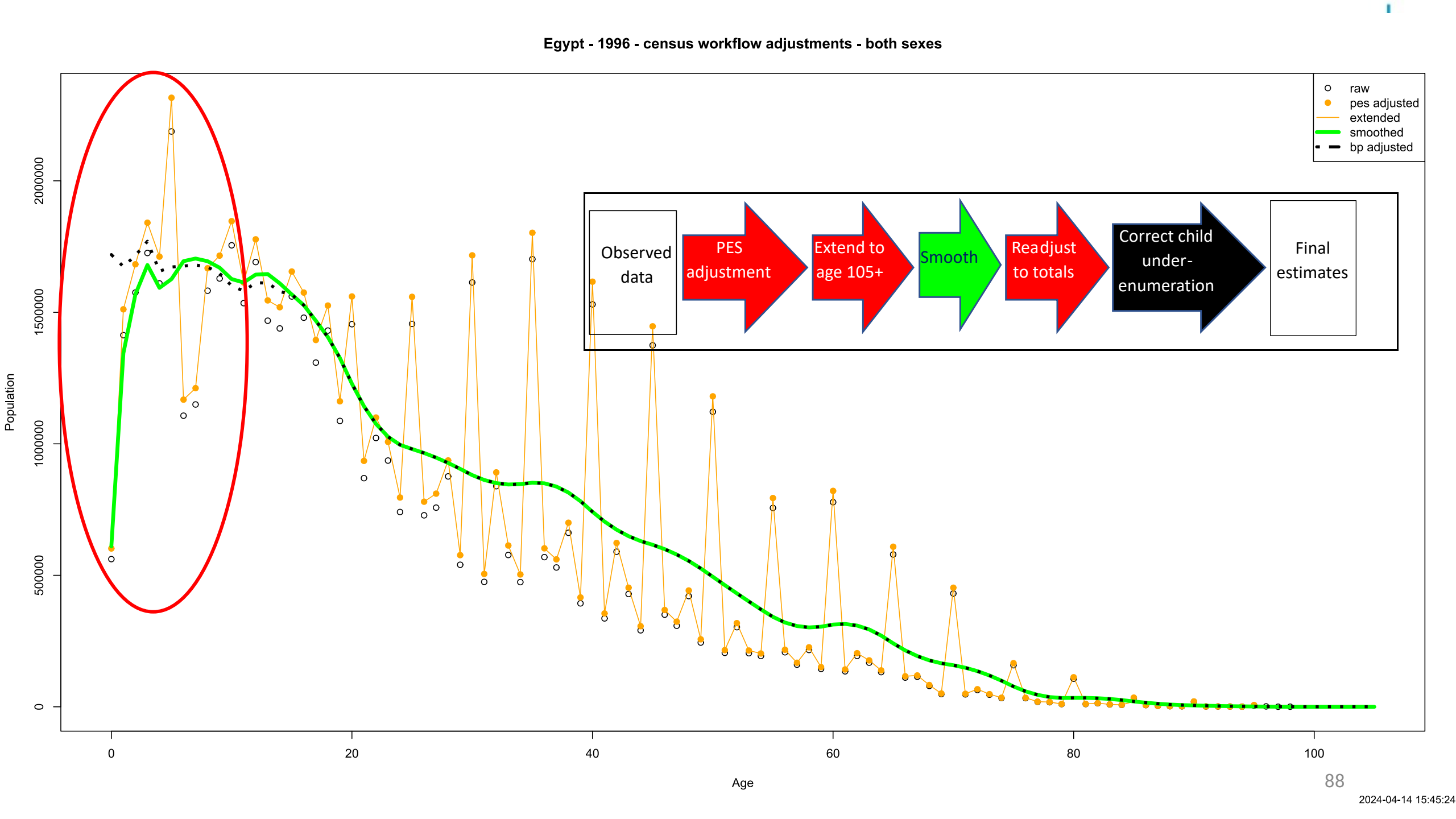
2024-04-14 15:45:24



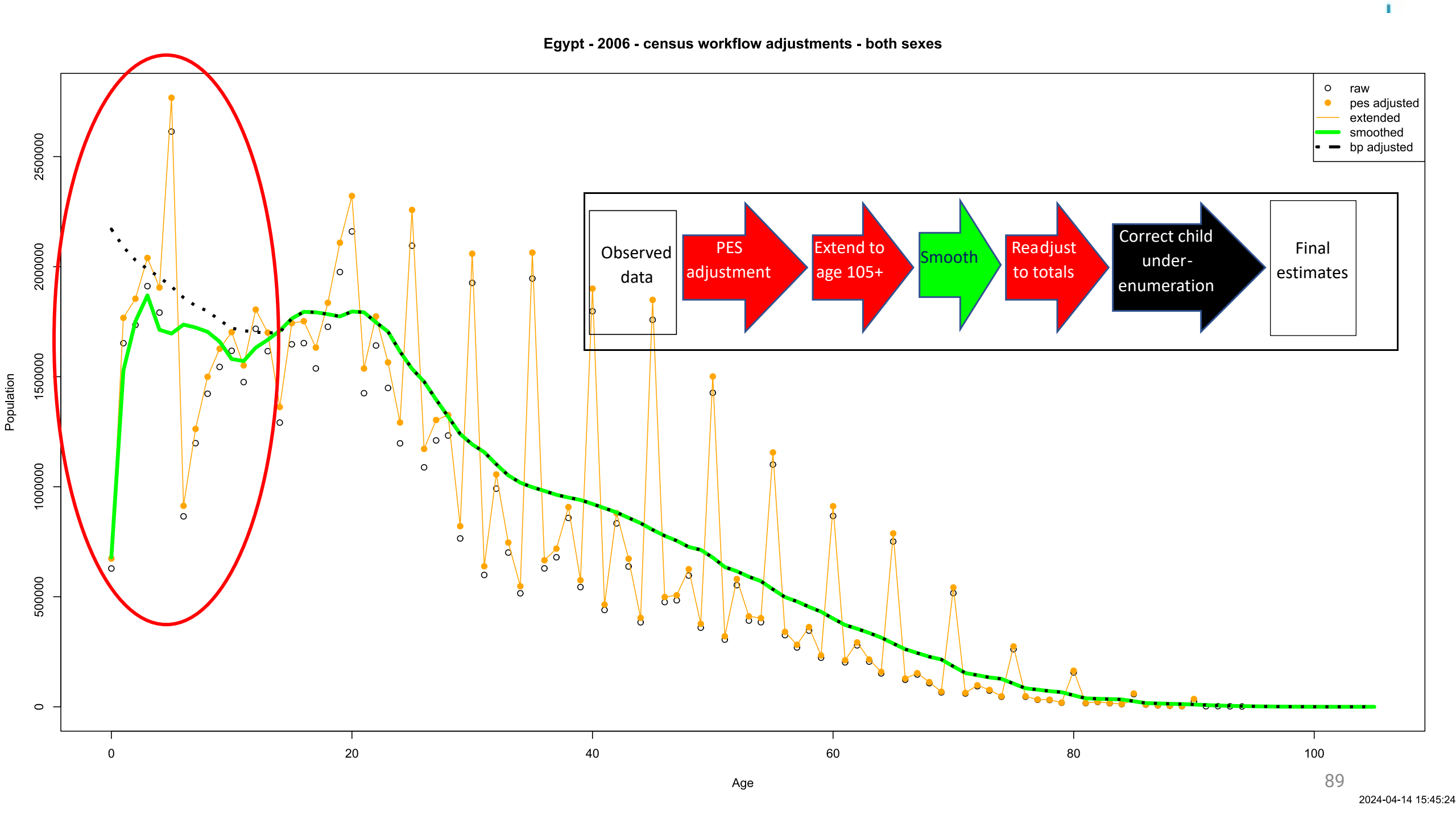
2024-04-14 15:45:24



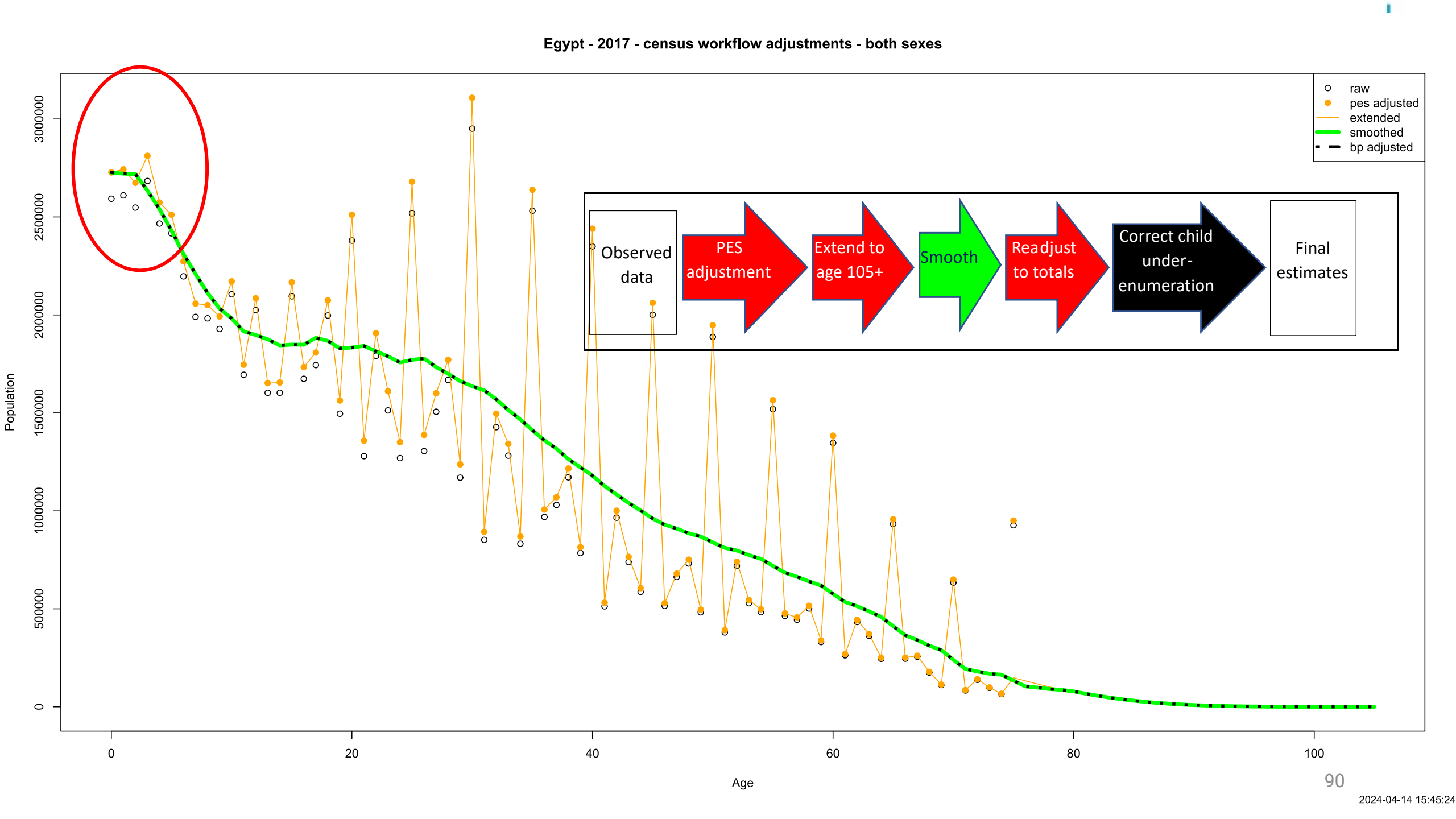
2024-04-14 15:45:24



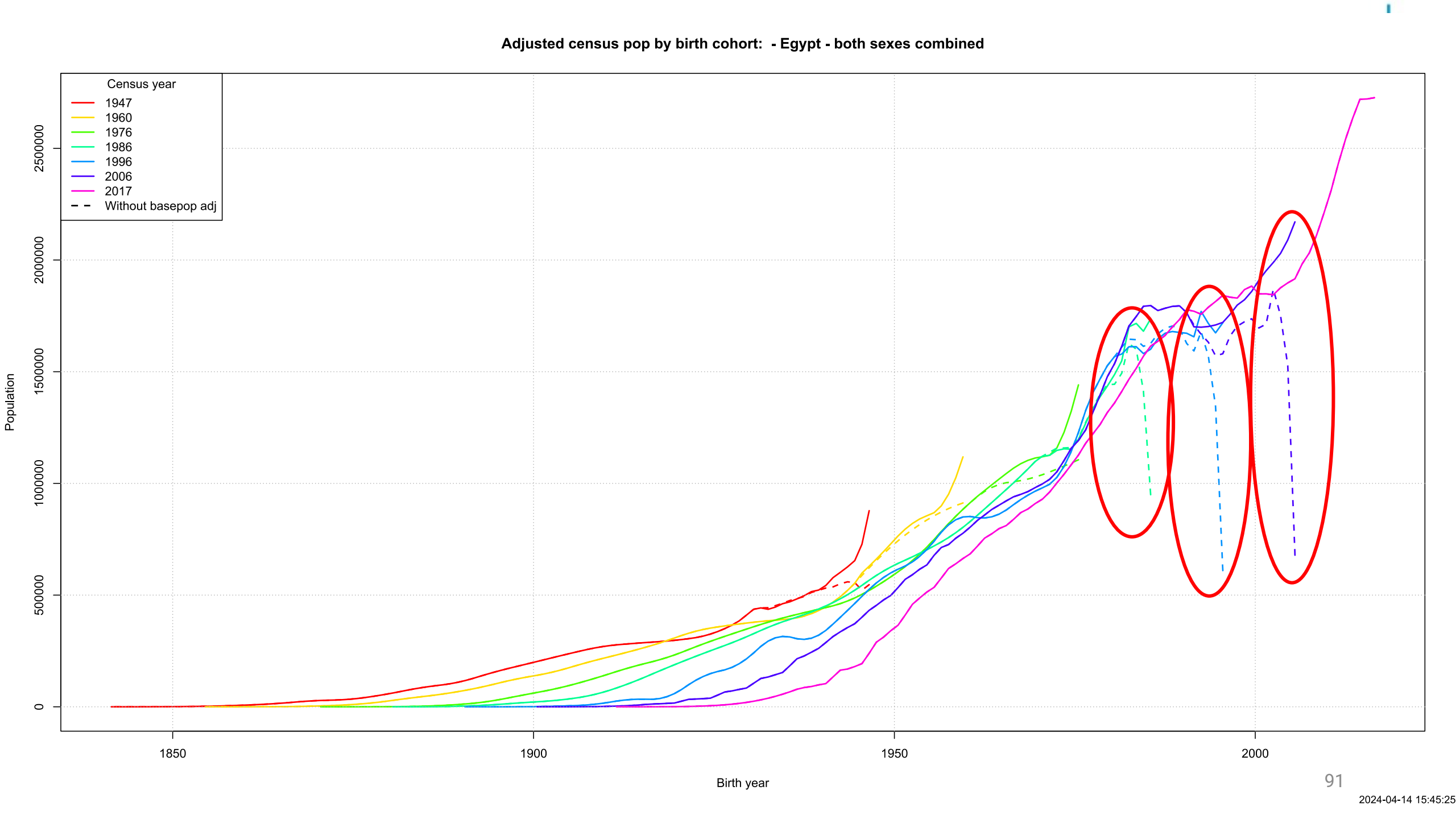
2024-04-14 15:45:24



2024-04-14 15:45:24



2024-04-14 15:45:24



2024-04-14 15:45:25

Egypt

- Example of application of Population Census Evaluation Process for WPP 2024 estimates

Census	adjust pes	pes adjustment	adjust smooth	adjust basepop	input age structure	input max age	age redist start	best smooth adult	best smooth child	bachi adult	bachi child	EduYrs	WPP %diff pes	WPP %diff smooth	WPP %diff basepop	Final WPP %diff
1947 Census	TRUE	-6.11	TRUE	TRUE	single	95	70	bestGrad5 = 2	bestMavN = 10	62.9	12.9	0.3	6.5	6.5	10.7	
1960 Census	TRUE	-6.13	TRUE	TRUE	abridged	75	75	bestGrad5 = 2	bestGrad5 = 2			0.7	6.5	6.5	8.9	4.1
1976 Census	TRUE	-5.89	TRUE	TRUE	abridged	75	75	bestGrad5 = 2	bestGrad5 = 2			1.9	6.3	6.3	10.0	9.6
1986 Census	TRUE	-5.84	TRUE	TRUE	single	99	85	bestGrad5 = 2	bestMavN = 10	32.9	5.0	3.5	6.2	6.2	8.9	9.3
1996 Census	TRUE	-5.68	TRUE	TRUE	single	98	85	bestGrad5 = 1	bestMavN = 10	30.5	4.6	5.6	6.0	6.0	8.6	13.9
2006 Census	TRUE	-5.78	TRUE	TRUE	single	94	85	bestMavN = 10	bestMavN = 10	26.4	5.7	8.0	6.1	6.1	11.2	14.5
2017 Census	TRUE	-4.00	TRUE	TRUE	single	75	75	bestMavN = 10	bestMavN = 6	22.8	3.3	10.8	4.2	4.2	4.2	8.9

See *Method protocol for the evaluation of census population data by age and sex*, as well as structured metadata for further details

United Nations | Department of Economic and Social Affairs

Part 6

Global perspective on challenges and opportunities for internationally comparable estimates

WPP data sources, metadata and empirical data

United Nations | Department of Economic and Social Affairs
Population Division

World Population Prospects 2024

WPP Home | Data | Graphs / Profiles | Documentation | World Urbanization Prospects | Population Division | Contact Us

Data Sources

A description of the empirical data used and the methods applied in revising past estimates of population and components of demographic change (fertility, child, adult and overall mortality, international migration) is available here for each country or area for the period 1950 to 2023. For the countries with less than 1,000 inhabitants in 2023, only the data sources for total population are made available.

Select below a country or area to see the respective information, or download the whole set of metadata in PDF format for all countries or areas or a tabular version for each demographic component under Documentation.

Select a Country or Area

Egypt

Egypt

Population

Total population and distribution by age and sex estimated to be consistent with the population by age and sex of the (a) 1947, 1960, 1976, 1986, 1996, 2006, 2017 censuses; (b) adjusted for under/over count; (c) adjusted for age heaping; (d) adjusted for under enumeration of children under age 15; (e) official estimates through 2021; (f) only total population for 1966 Census; and with estimates of the subsequent trends in fertility, mortality and international migration.

Graphs:

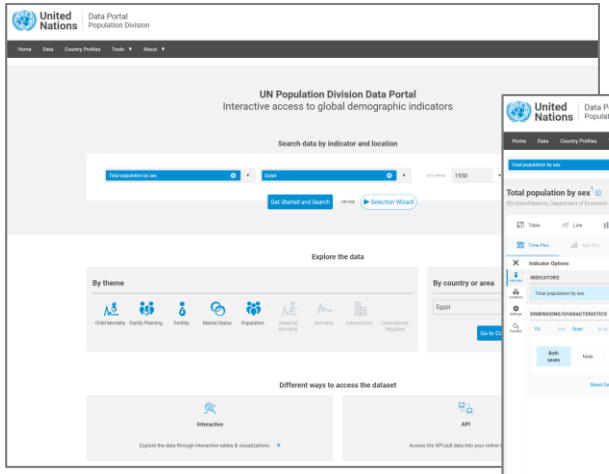
Total Population | 5-year age groups | Single age distribution

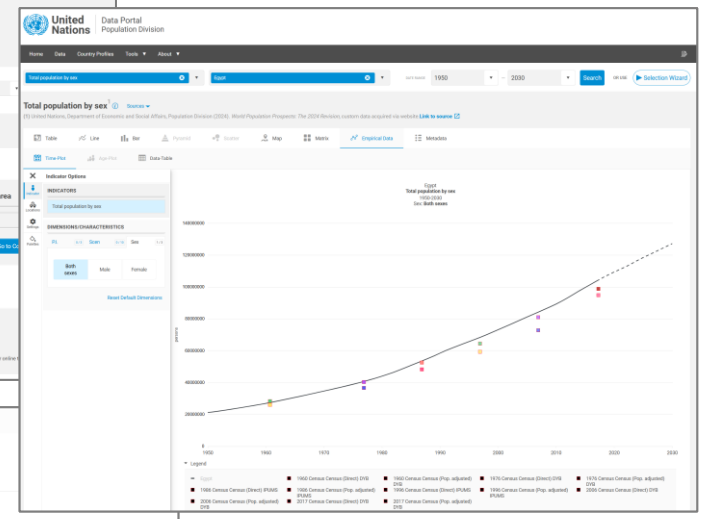
Fertility

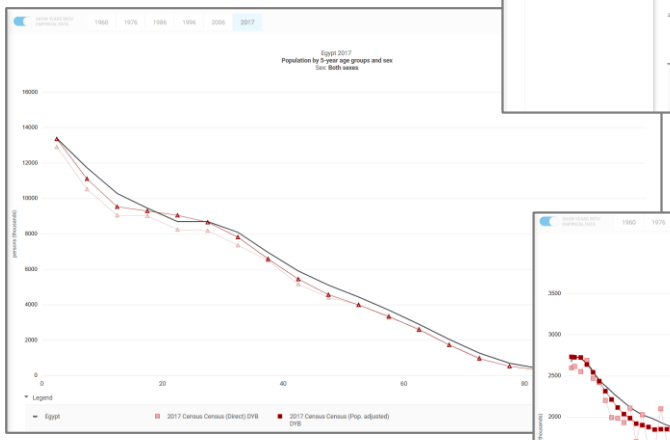
Total fertility rate and age pattern of fertility based on: (a) official estimates of age-specific fertility rates through 2016; (b) registered births classified by age of mother and the underlying female population by age through 2020; (c) birth-histories data from the 1980 WFS and 1988, 1992, 1995, 1997, 2000, 2005, 2008, 2014 DHS and 1991 PAPCHILD and 1998, 2003 DHS Interim; (d) births in the household in the preceding 12 (or 24) months classified by age of mother from the 1947 Census; (e) indirect estimates obtained from the application of the reverse survival method to the 1956-2011 Education Stats and 1960, 1976, 1986, 1996, 2006, 2017 censuses; (f) cohort-completed fertility backdated by the mean age of childbearing from the 1976, 1986 censuses and 1980 WFS and 1988, 1992, 1995, 2000, 2005, 2008, 2014 DHS and 1991 PAPCHILD and 2003 DHS Interim. In addition, In addition, the births in the household in the preceding 12 (or 24) months classified by age of mother from the 1960, 1976 censuses and 1984 CPS have been considered.

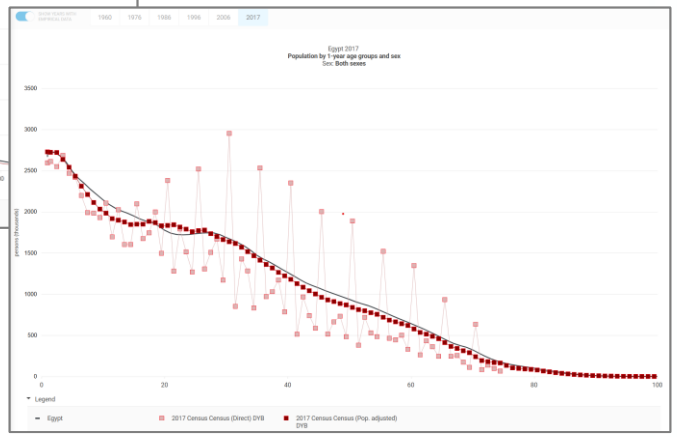
Graphs:

Total fertility rate | 5-year age groups





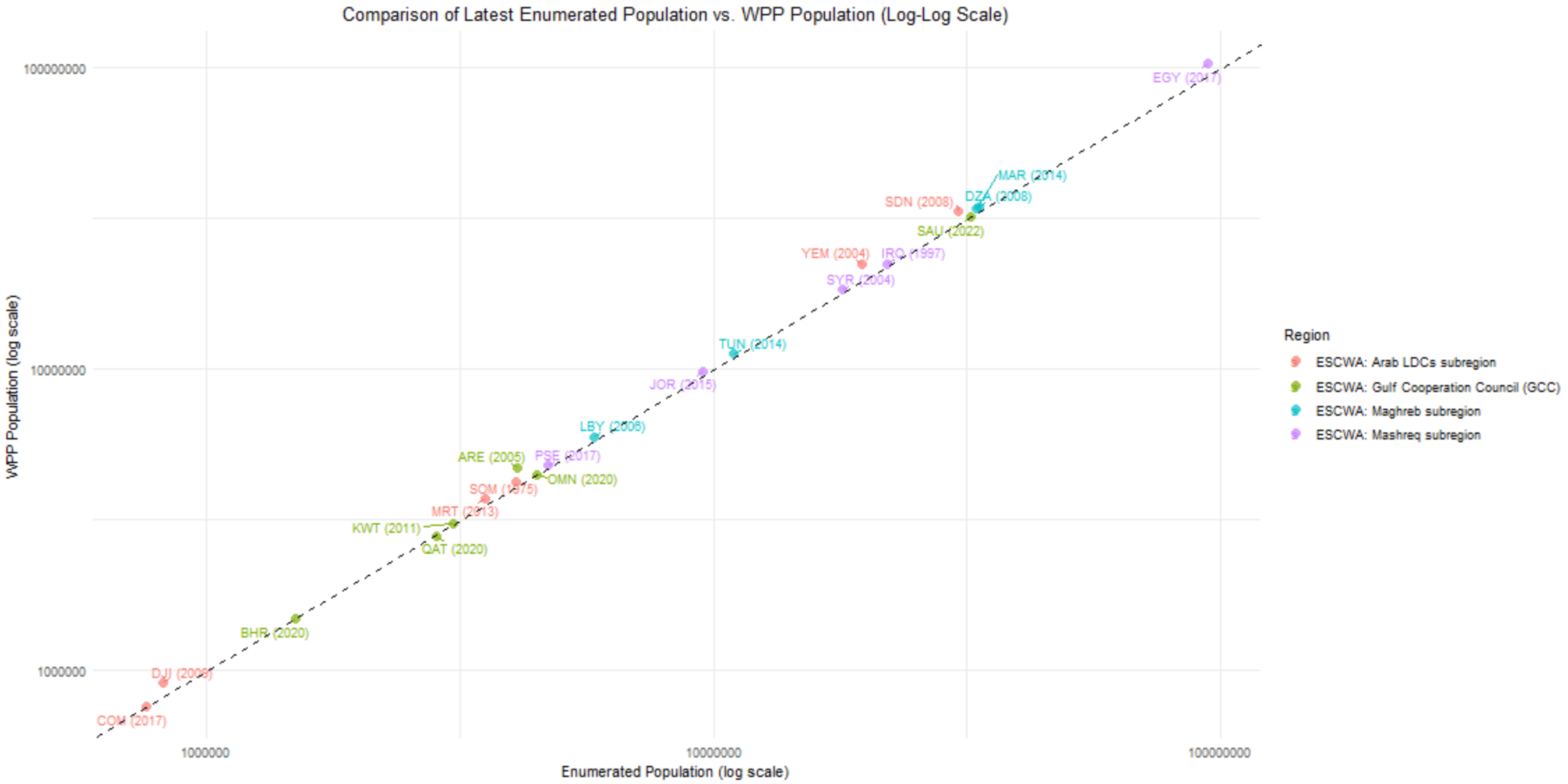




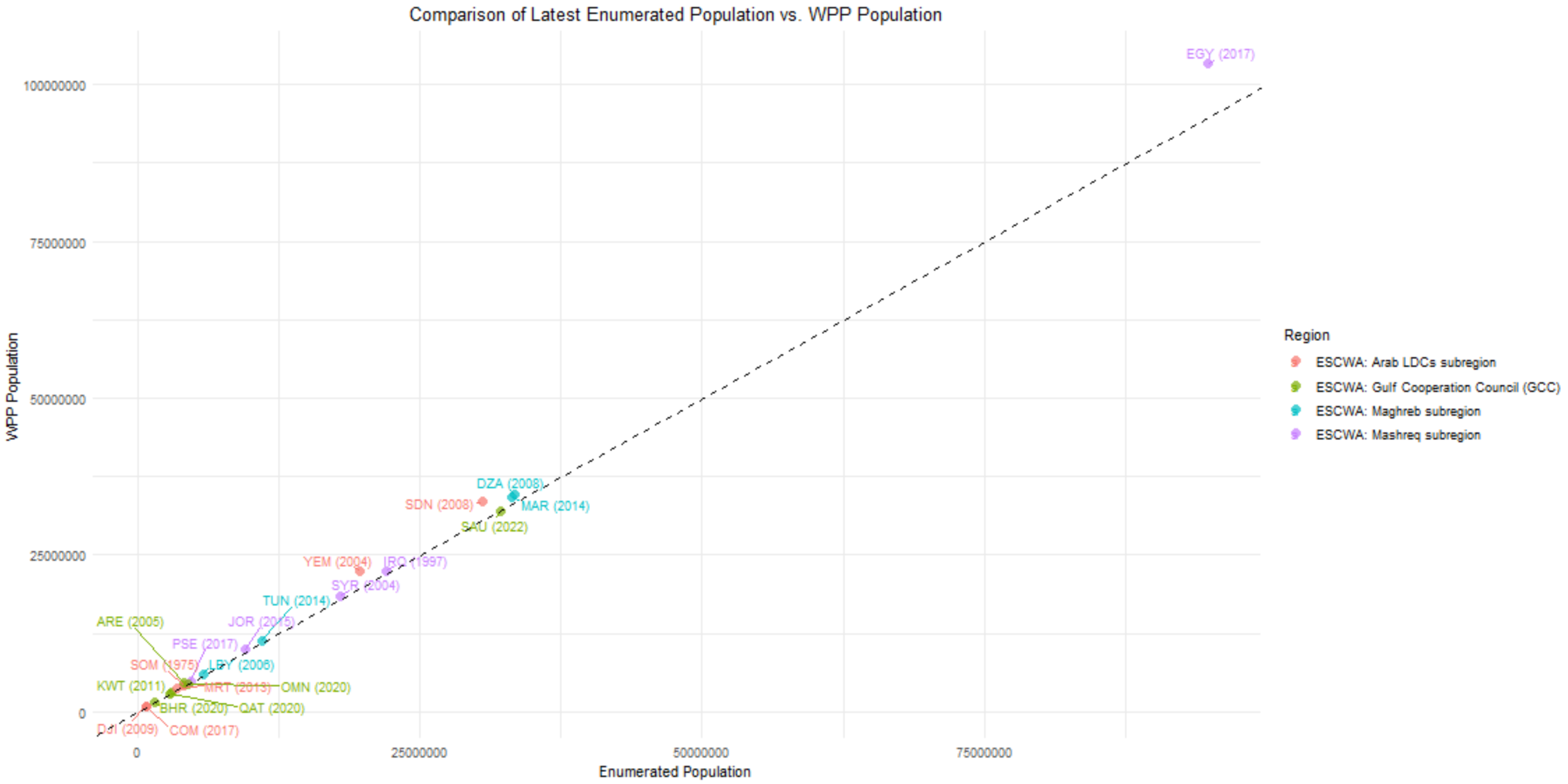
[Online data portal]

Source: https://population.un.org/wpp/data-sources?location=818

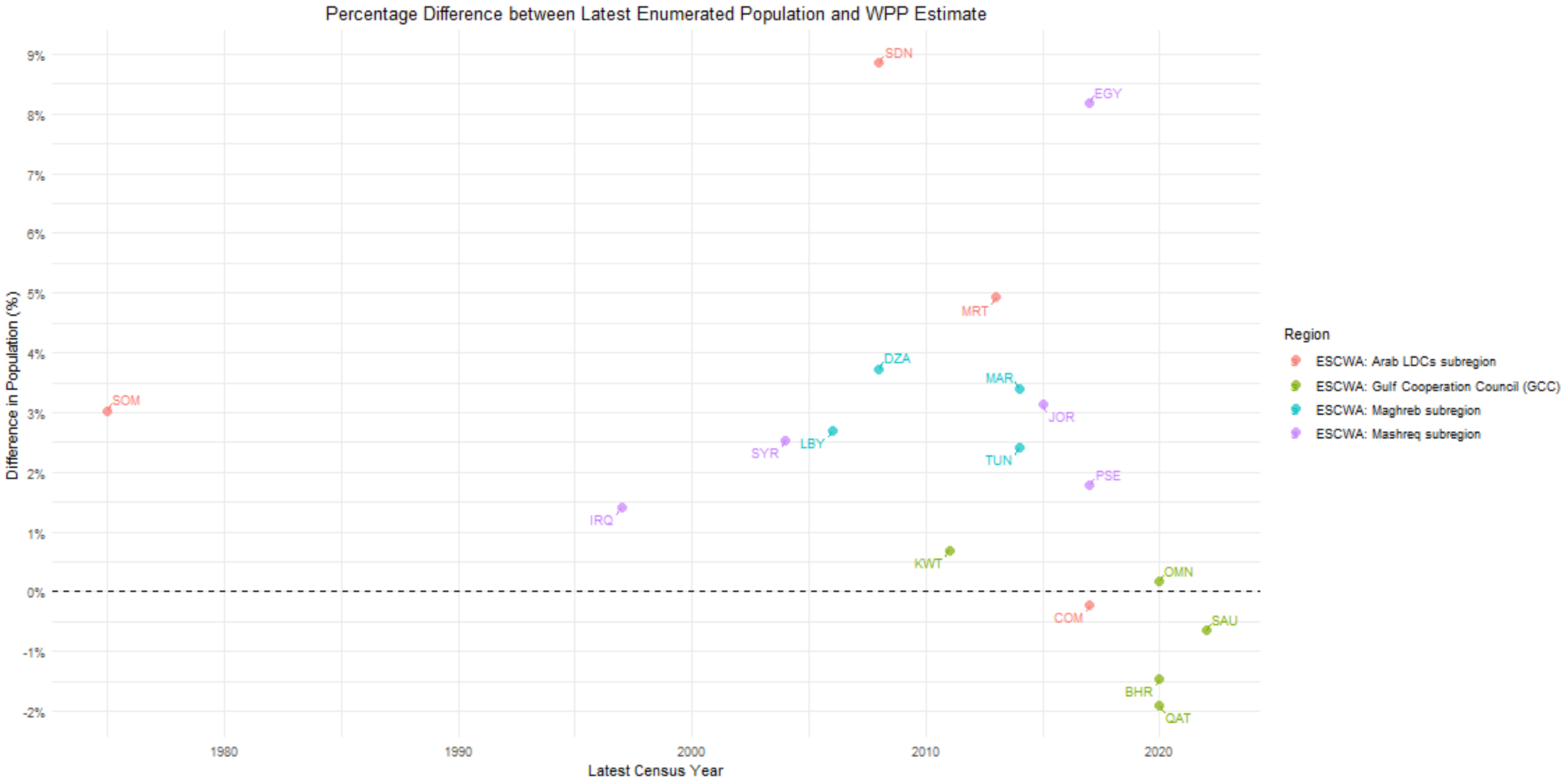
ESCWA region: latest pop. census vs. WPP



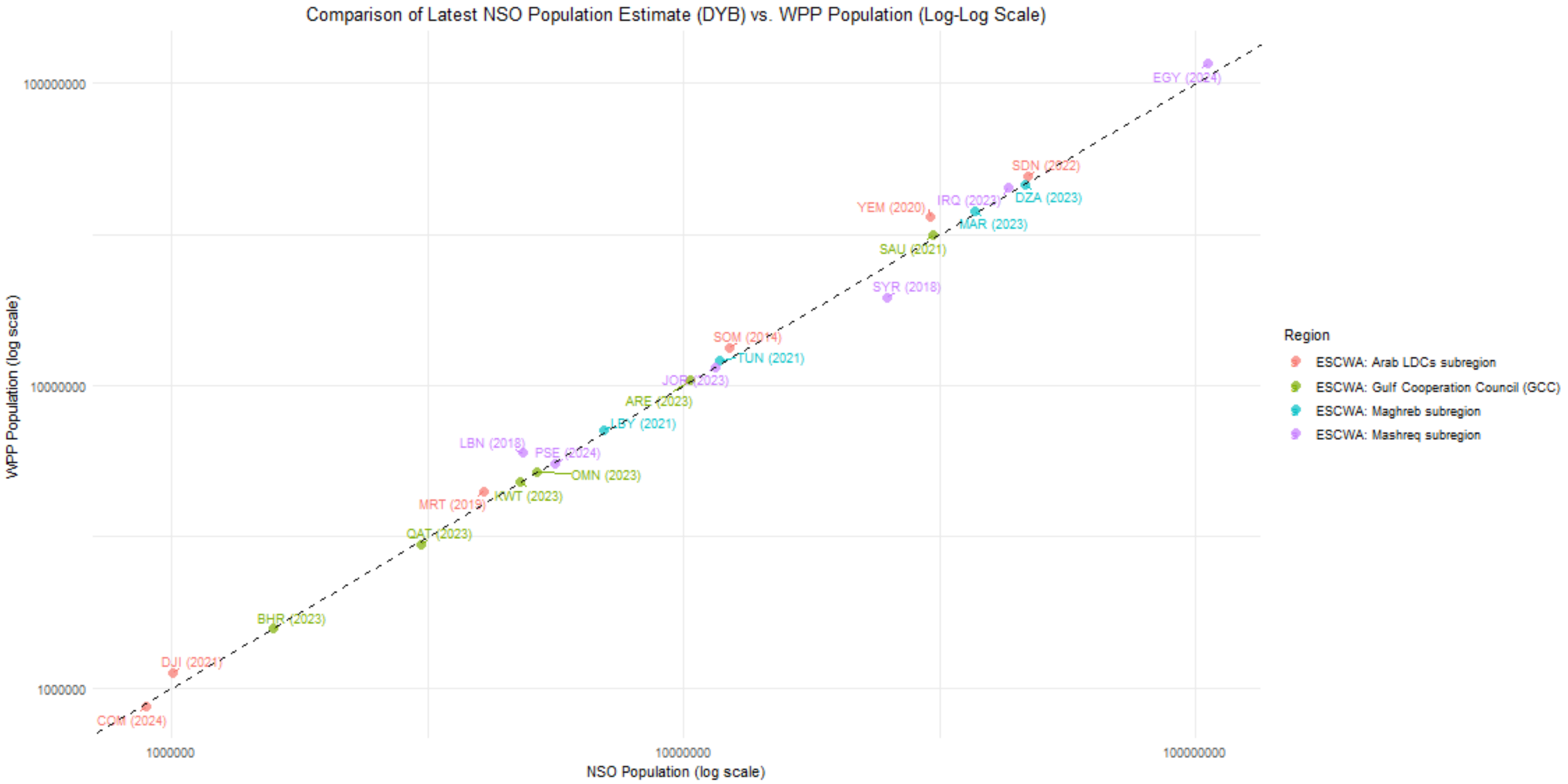
ESCWA region : latest pop. census vs. WPP



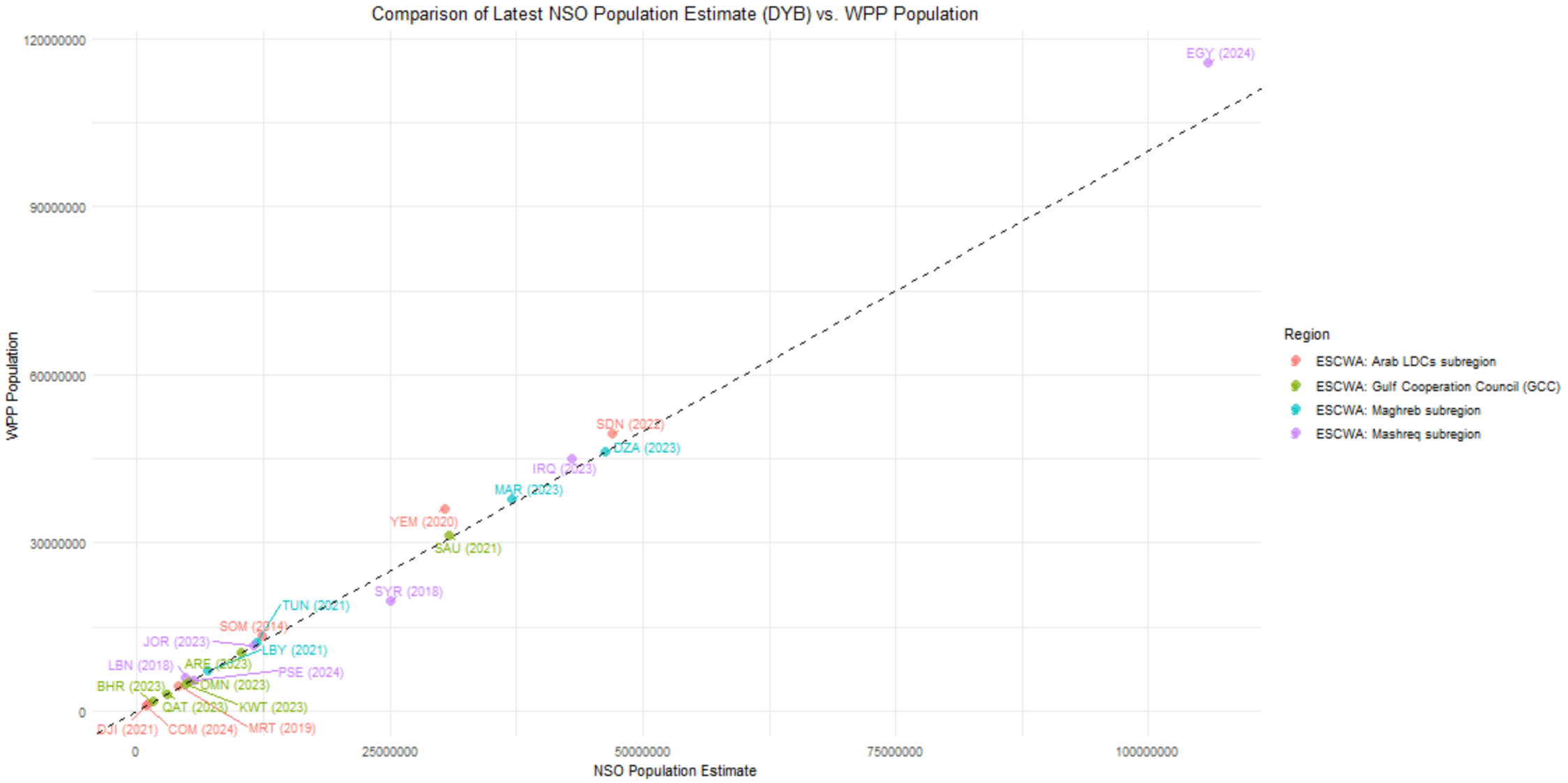
ESCWA region : latest pop. census vs. WPP



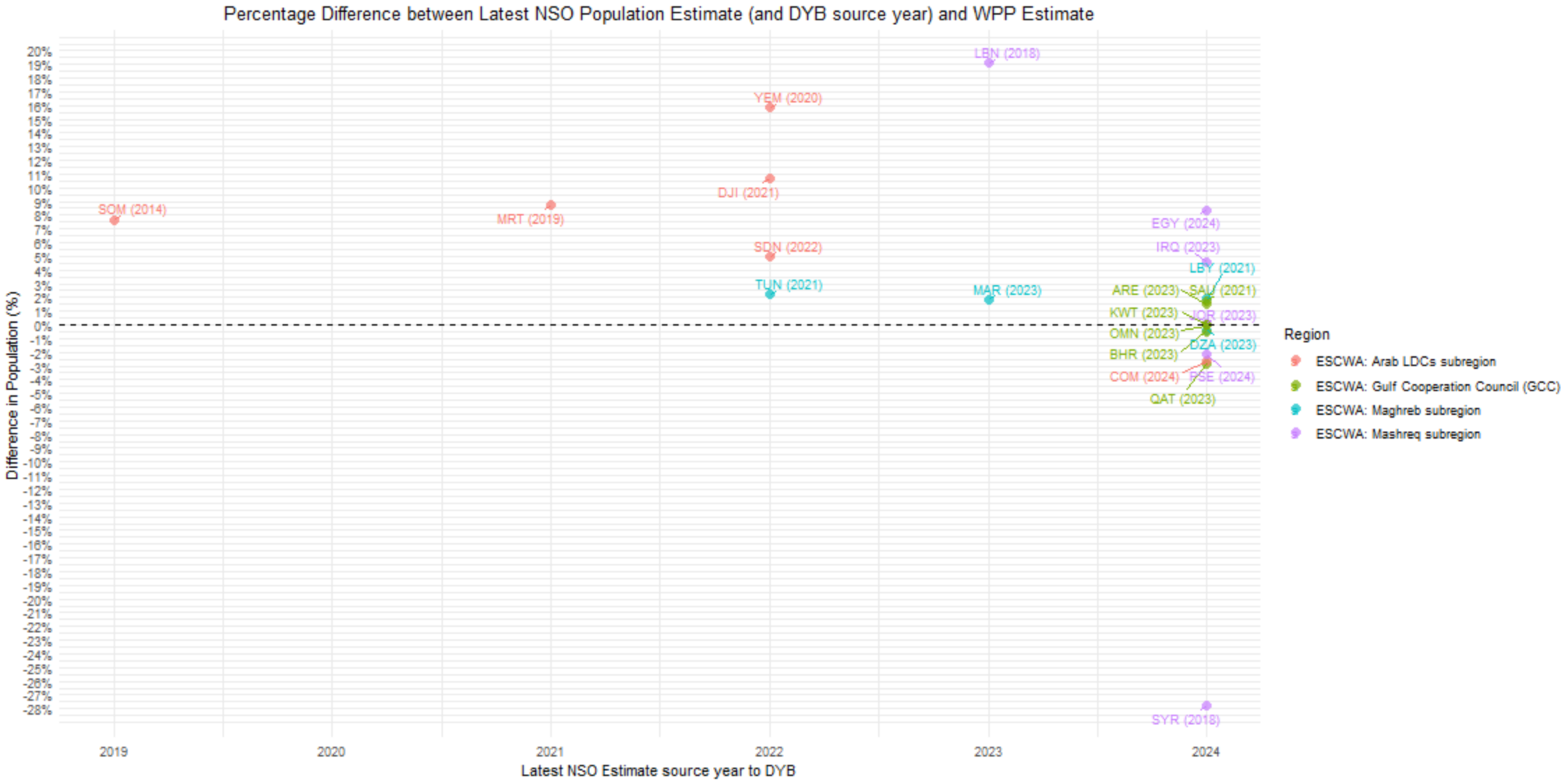
ESCWA region : latest pop. estimate vs. WPP



ESCWA region : latest pop. estimate vs. WPP



ESCWA region : latest pop. estimate vs. WPP

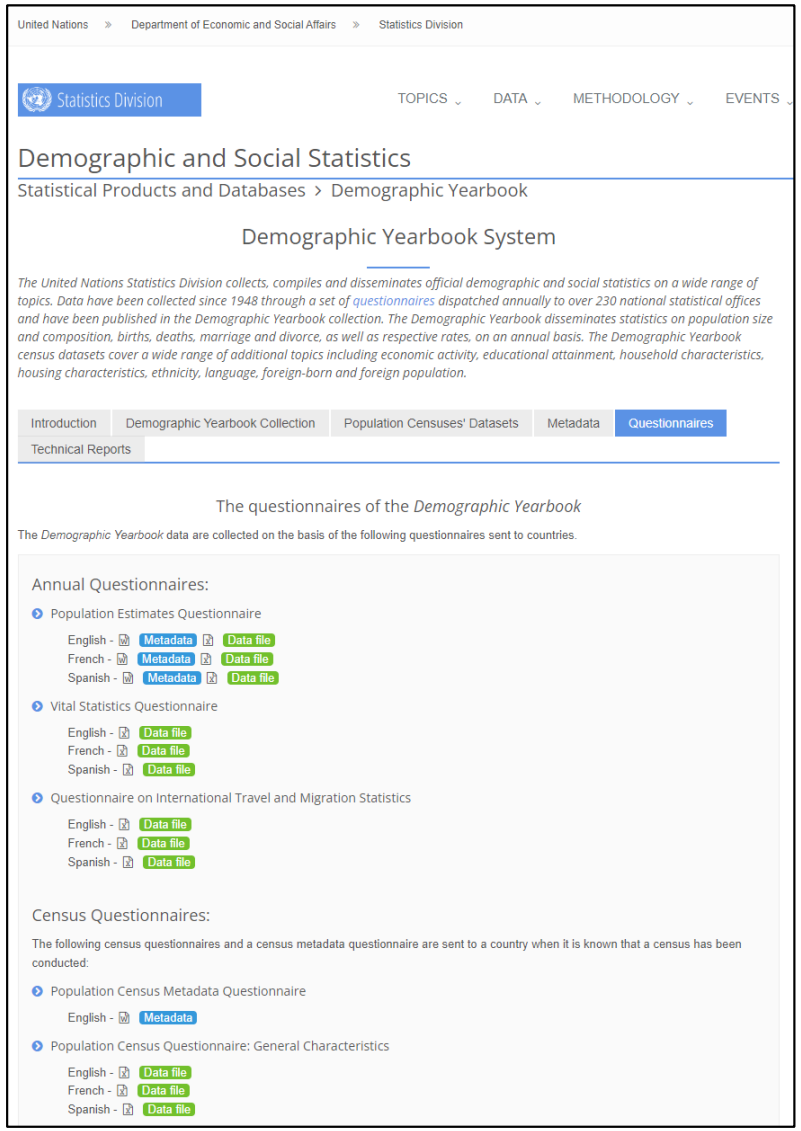


Latest empirical data sources used for WPP 2024 estimates

Location	Pop. Census	Fertility (TFR, ASFR) Census	Fertility (TFR, ASFR) Survey	Fertility (TFR, ASFR) Register	Child mortality (5q0) Census	Child mortality (5q0) Survey	Child mortality (5q0) Register	Adult mortality (45q15) Census	Adult mortality (45q15) Survey	Adult mortality (45q15) Register
ESCWA: Arab LDCs subregion										
Comoros	2017	2017	2012		2018	2012		2017	2012	
Djibouti	2009	2012	2012			2012		2009	2011	
Mauritania	2013	2020	2020		2013	2019	2013	2013	2020	
Somalia	1975	2018	2018			2016			2018	
Sudan	2008	2015	2014	2015	2008	2014		2008	2014	
Yemen	2004	2023	2023		2004	2022		2001	1991	
ESCWA: Gulf Cooperation Council (GCC)										
Bahrain	2020	2020	1995	2020	1997	1995	2020			2020
Kuwait	2011	2022	1996	2022	1977	1996	2022			2022
Oman	2020	2022	2014	2022	2003	2014	2023	2020		2022
Qatar	2020	2021	2012	2021	2000	1998	2022			2021
Saudi Arabia	2022	2022	2016	2022	2007	2017	2022	2010	2017	
United Arab Emirates	2005	2022	1995	2022	1978	1995	2023	1980		
ESCWA: Maghreb subregion										
Algeria	2008	2019	2019	2019		2019	2021	2008	2019	2019
Libya	2006	2014	2014	2014	1971	2014	2017	2006	1995	2011
Morocco	2014	2021	2011		2010	2018	2018	2014	2017	
Tunisia	2014	2021	2018	2021	2009	2018	2021	2004	2018	2016
ESCWA: Mashreq subregion										
Egypt	2017	2021	2021	2021	1985	2021	2022	2017	2014	2021
Iraq	1997	2018	2018		1994	2018	2021	1997	2018	2019
Jordan	2015	2023	2023	2023	2015	2017	2020	2015	2017	2018
Lebanon	1942		2018			2008	2023		2007	2018
State of Palestine	2017	2020	2020	2020	2016	2020	2021	2017	2020	2021
Syrian Arab Republic	2004	2009	2009		2004	2018	2011	2004	2006	2010

Latest available data
2019-2023 (last 5 years)
2014-2018 (5-9 years ago)
2009-2013 (10-14 years ago)
< 2009 (more than 15 years ago)

UNSD: Demographic Yearbook questionnaires



United Nations » Department of Economic and Social Affairs » Statistics Division

Statistics Division

TOPICS DATA METHODOLOGY EVENTS

Demographic and Social Statistics

Statistical Products and Databases › Demographic Yearbook

Demographic Yearbook System

The United Nations Statistics Division collects, compiles and disseminates official demographic and social statistics on a wide range of topics. Data have been collected since 1948 through a set of questionnaires dispatched annually to over 230 national statistical offices and have been published in the Demographic Yearbook collection. The Demographic Yearbook disseminates statistics on population size and composition, births, deaths, marriage and divorce, as well as respective rates, on an annual basis. The Demographic Yearbook census datasets cover a wide range of additional topics including economic activity, educational attainment, household characteristics, housing characteristics, ethnicity, language, foreign-born and foreign population.

Introduction | Demographic Yearbook Collection | Population Censuses' Datasets | Metadata | Questionnaires | Technical Reports

The questionnaires of the *Demographic Yearbook*

The *Demographic Yearbook* data are collected on the basis of the following questionnaires sent to countries.

Annual Questionnaires:

- Population Estimates Questionnaire
  - English - Metadata Data file
  - French - Metadata Data file
  - Spanish - Metadata Data file
- Vital Statistics Questionnaire
  - English - Data file
  - French - Data file
  - Spanish - Data file
- Questionnaire on International Travel and Migration Statistics
  - English - Data file
  - French - Data file
  - Spanish - Data file

Census Questionnaires:

The following census questionnaires and a census metadata questionnaire are sent to a country when it is known that a census has been conducted:

- Population Census Metadata Questionnaire
  - English - Metadata
- Population Census Questionnaire: General Characteristics
  - English - Data file
  - French - Data file
  - Spanish - Data file

Introduction | Demographic Yearbook Collection | Population Censuses' Datasets | Metadata | Questionnaires | Technical Reports

Demographic Yearbook Metadata

Below are made available the completed Demographic Yearbook metadata questionnaires of the recent reporting years, received from national statistical offices along with the respective completed Demographic Yearbook data questionnaires.

AFRICA

Country or area	Population estimates metadata	Vital statistics metadata	Population and housing census metadata
Algeria	2016 -		2008 -
Angola	2015 - 2018 - 2019 - 2021 - 2022 - 2024 -	2018 -	
Benin	2011 -		
Botswana	2005 - 2009 - 2011 - 2017 - 2020 - 2023 -	2009 - 2013 - 2019 - 2020 - 2021 -	1991 - 2022 -
Burkina Faso	2020 - 2021 - 2022 - 2024 -	2021 -	2006 -
Burundi	2015 - 2016 - 2017 - 2019 -	2016 - 2020 -	
Cameroon	2017 - 2019 - 2020 - 2022 - 2023 - 2024 -	2017 -	
Cape Verde			1990 -
Central African Republic			
Chad	2020 -		
Comoros			2003 -

Source: https://unstats.un.org/unsd/demographic-social/products/dyb/index.cshtml#questionnaires

Metadata Availability for Population Estimates

Essential information on data sources, estimation methods, adjustments, and assumptions.

Importance of metadata: crucial for transparency, data interpretation, and to enable users to assess data quality and comparability, understand limitations, and make informed decisions.

Global Availability:

- UN DYB questionnaire on data and methods used for official estimates
- Metadata availability and completeness vary significantly by country.
- Some regions provide complete metadata, but others lack details on adjustments, and methods used.

For further details, see Spoorenberg (2020)

Challenges for Internationally Comparable Estimates

Common issues:

- Lack of detailed documentation on population data and methods
- Differences in definitions (e.g., de-facto, de-jure, usual resident), and methods across countries
- Varying data availability and quality gaps
- Differences in coverage of population registers and civil registration systems
- Difficulties in including migration data
- Metadata gaps make it difficult to assess the accuracy of population estimates for international comparison

Opportunities:

- Growing use of administrative data and registers
- Greater use of geospatial data and digital-first census methods to fill data gaps
- Technological advancements, data integration, and improved statistical methods
- Potential for more timely, granular and comparable estimates
- Need for more harmonized approaches while respecting national contexts

Opportunities for Improving Population Estimates

- Countries must modernize data systems to produce reliable, timely, and internationally comparable population estimates.
- Emphasis on the integration of different data sources (censuses, registers, administrative data) for better estimates from individual level to various aggregated level (geographical and subgroups).
- Increased use of technology, administrative data, and geospatial techniques can provide more frequent and accurate population estimates.
- Produce and disseminate more comprehensive and accessible metadata for greater transparency.
- Importance of international collaboration: sharing best practices, capacity building, and harmonization of standards.

United Nations | Department of Economic and Social Affairs

Part 7

Annexes

References

- European Commission (2003). Basic Methodology for the Recalculation of Intercensal Population Estimates, Luxembourg: Office for Official Publications of the European Communities, Working Papers and Studies (3/2003/E/no 27).
- Johnson, P., and others (2022). *Method protocol for the evaluation of census population data by age and sex*. UN Population Division, Technical Paper, No. UN DESA/POP/2022/TP/No.5, New York, United Nations.
- Pelletier, F. (2020). *Census Counts, Undercounts and Population Estimates: The Importance of Data Quality Evaluation*, New York, United Nations, DESA, Population Division, Technical Paper No. 2020/02.
- Spoorenberg, T. (2020). *Data and Methods for the Production of National Population Estimates: An Overview and Analysis of Available Metadata*, New York, United Nations, DESA, Population Division, Technical Paper No. 2020/01.
- Statistics Canada. Demography Division (2016). *Population and Family Estimation Methods at Statistics Canada*. No. 91-528-X, 1100231609, Ottawa, Statistics Canada.
- United Nations (2024). *World Population Prospects 2024: Methodology of the United Nations population estimates and projections*. Working Paper, No. UN DESA/POP/2024/DC/No. 10, New York, Population Division, Dept. of Economic and Social Affairs

WPP 2022-24 upgrade summary (1)

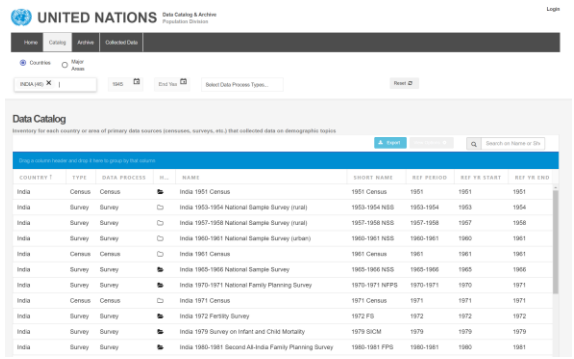
1. Upgrade production system to **single year and single age data model**
2. Improve capacity to **use annual time series** (upon data availability and reliability).
3. Improve capacity to **use single age data** (upon availability and reliability): use for good VR countries, for the rest **use 5-year age groups graduated into single age**.
4. Streamline/harmonize steps used to prepare country data and WPP estimates to enhance transparency, reproducibility, accessibility and linkages between empirical data and WPP estimates
5. Greater documentation and explanations of the various methods used to derive demographic estimates for each demographic components and the reconciliation with population estimates **->** WPP methodological report **+** WPP method protocol
6. Harmonized methods and tools (e.g., R packages like DemoTools), and open APIs and databases for input/output data
7. Extra fertility projection scenarios and probabilistic projections of net migrations

WPP 2022-24 upgrade summary (2)

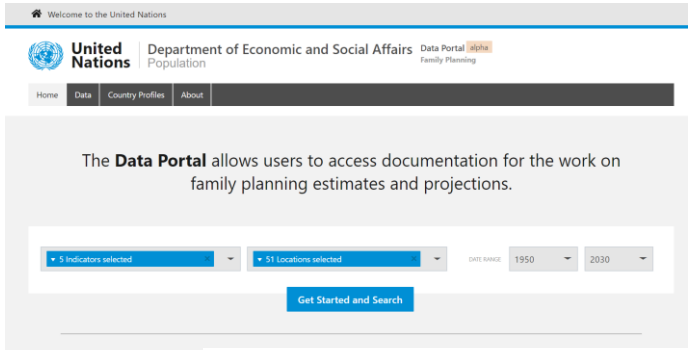
8. Improved visualization and dissemination of key indicators (e.g., new DataPortal)
9. Provide access to both WPP estimates and underlying empirical data for key demographic indicators -> Data Portal + Demo Data + Data Archive
10. More GATHER compliant, including comprehensive metadata on data sources and estimation methods for all demographic components and locations (in textual and structured database format)
11. Data available interactively and for bulk download in Excel and ASCII formats, online plots and country profiles
12. R packages implementing all probabilistic projection methods + wpp R data package used for replication
13. Results of all previous revisions available online in data format since 1992 (Excel) and 1998 (ASCII), and PDF reports (since 1950s)

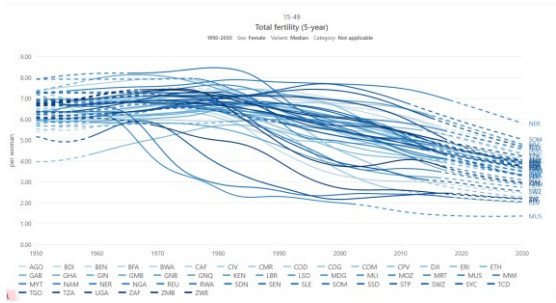
DataCatalog, DataArchive and DemoData

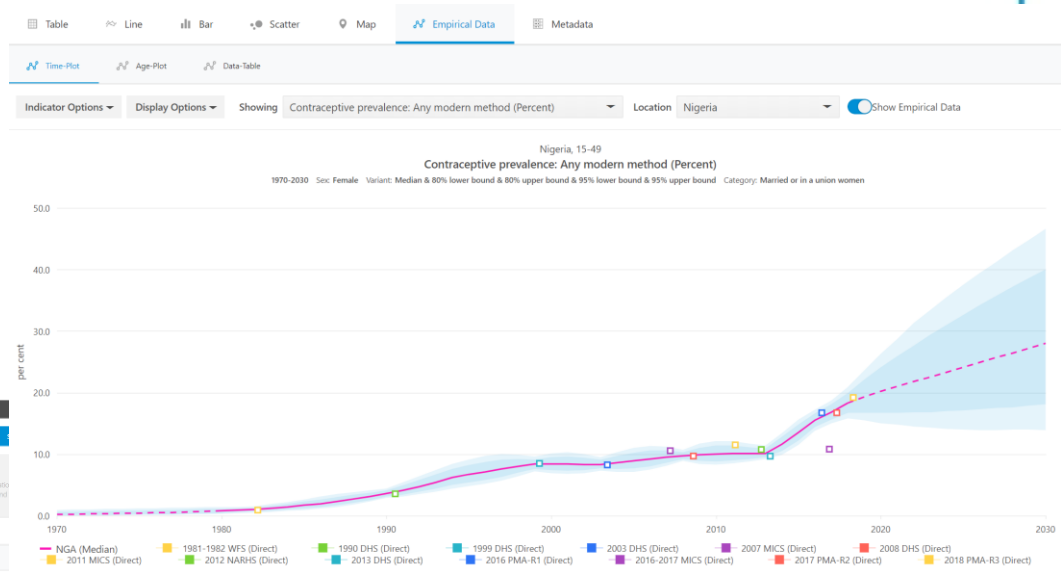
- DataCatalog: comprehensive **inventory for each country of primary data sources** (censuses, demographic surveys, etc.) providing data on demographic processes (fertility, mortality, population structure and dynamics, marital status and family planning) for all countries and areas, as of June 2024, about 8,310 entries
- DataArchive: a **virtual repository** of documents, tabular datasets and reports (potentially) for each data source, as of June 2024, more than 34,200 files
- DemoData: **SQL database to store in a structured and standardized way empirical data and demographic estimates** (with meta-information) on population, fertility, mortality and (net) international migration data from as many sources as possible

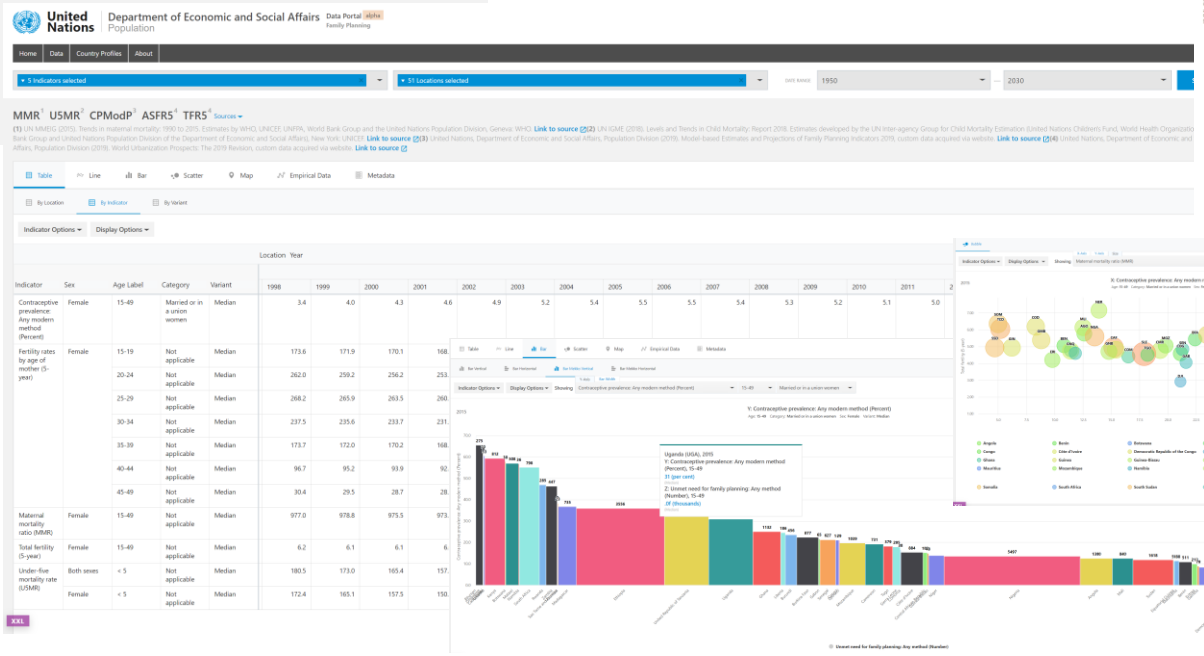


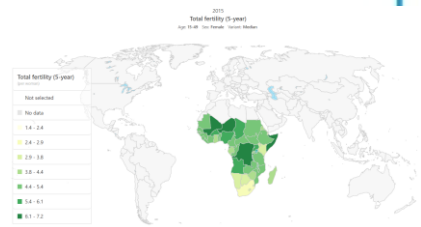
Data Portal: dissemination of estimates/projections and empirical data

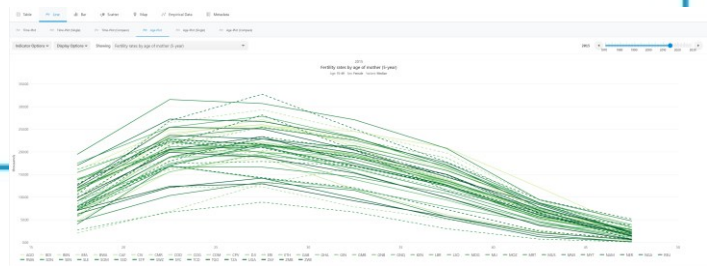












DESA | Population Division https://population.un.org/dataportal/

An open suite of R packages and functions

- New 1x1 cohort-component population projection computational engine designed to work with a standard set of 1x1 inputs and outputs
  - R implementation – for deterministic projections/simulations
  - C implementation – for probabilistic projections/simulations
- B3-type of robust time trend modelling for TFR/ASFR, and adult mortality for non-VR countries
- DDSQLTools: set of functions to query DemoData SQL database with API
- DemoTools: set of functions to evaluate, transform and adjust counts or rates
- DDM and FertEstR: set of functions to evaluate and adjust mortality and fertility data
- Ungroup and MortalityLaws: mortality graduation and extension at older ages
- SVDcomp: new expanded set of model life tables (including HIV and ART)
- Calibrated Splines: graduation of fertility age patterns
- Additional packages/functions to operationalize WPP method protocol (ddHarmony, ccmppWPP, etc.)
- popReconstruct: probabilistic demographic estimation and population reconstruction
- BayesTFR, BayesLife/BayesLifeHIV, MortCast, bayesMig, BayesPop: probabilistic projections

Acknowledgments

- The 2024 Revision of the *World Population Prospects* benefited of the inputs and support of the following individuals and teams:
- *Core team*: Srikanth Athaluri, Helena Cruz Castanheira, Fernando Fernandes, Patrick Gerland, Sara Hertog, Yumiko Kamiya, Vladimíra Kantorová, Pablo Lattes, Kyaw Kyaw Lay, Joseph Molitoris, Suryanarayana Murthy Palacharla, José H. C. Monteiro da Silva, Mark Wheldon, Iván Williams, Venkata Yamarthy and Lubov Zeifman;
- *Additional support team*: Dennis Butler, Fengqing Chao, Adrian Raftery, James Raymer, Hana Ševčíková, Carl Schmertmann, Bruno Schoumaker.
- United Nations Inter-agency Group for Child Mortality Estimation (UN IGME)
- WHO-UNDESA Technical Advisory Group for COVID Mortality Assessment