



Wittgenstein Centre

FOR DEMOGRAPHY AND  
GLOBAL HUMAN CAPITAL



# Relational spline-model for interpolating demographic data and population projections

**Dalkhat M. Ediev**

([ediev@iiasa.ac.at](mailto:ediev@iiasa.ac.at); [dalkhat@hotmail.com](mailto:dalkhat@hotmail.com))

The work was funded by the Russian Foundation for Basic Research  
(Grants №18-01-00289; №18-010-01169 )

Work Session on Demographic Projections  
Belgrade, 25–27 November 2019

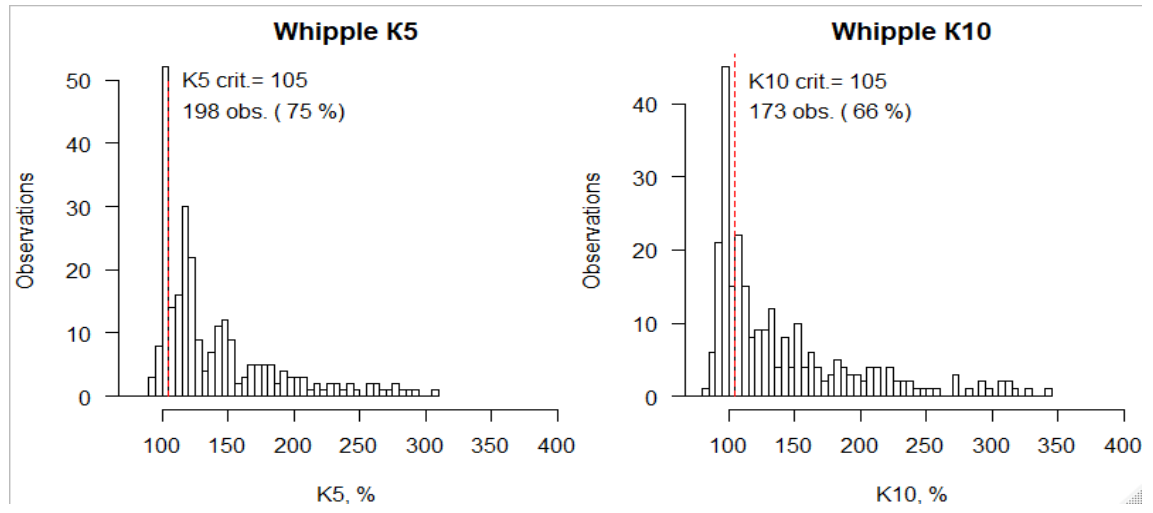
# Need for Interpolation of Demographic Data

- Correcting data
- Data preparation
- Enhancement of population projection's output

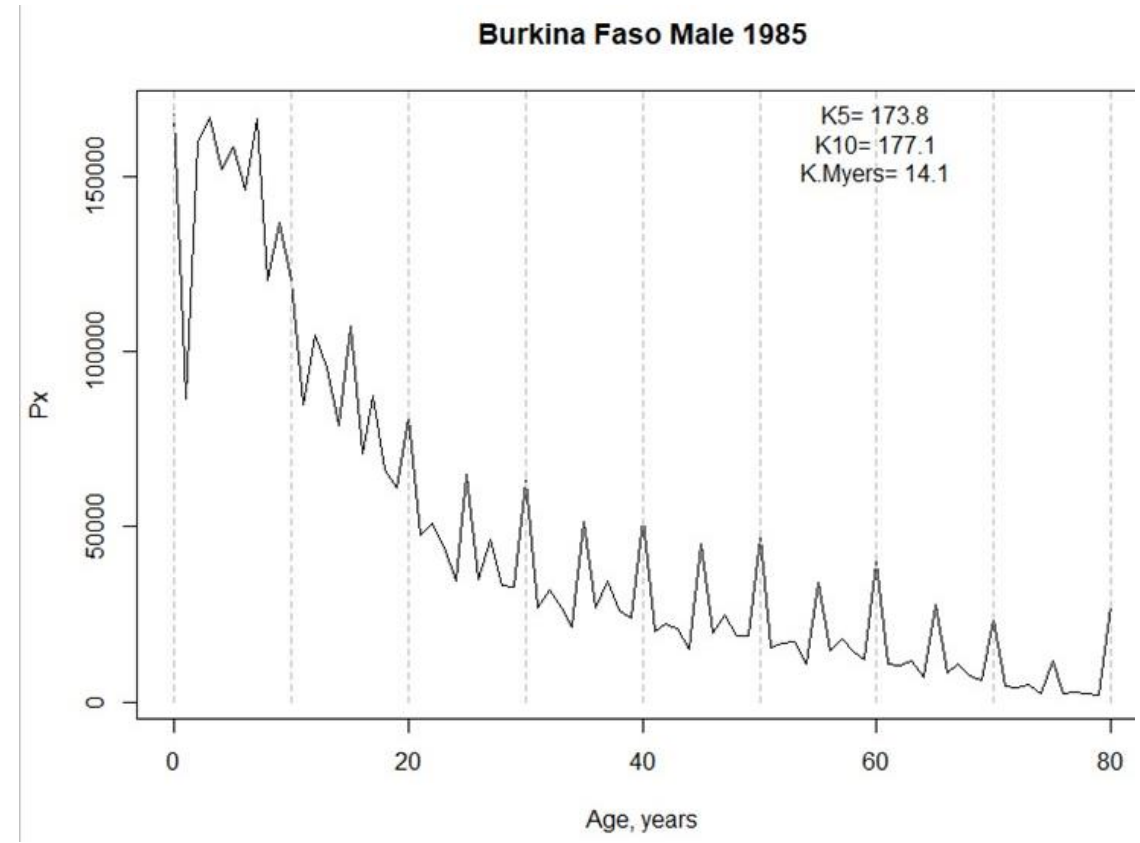
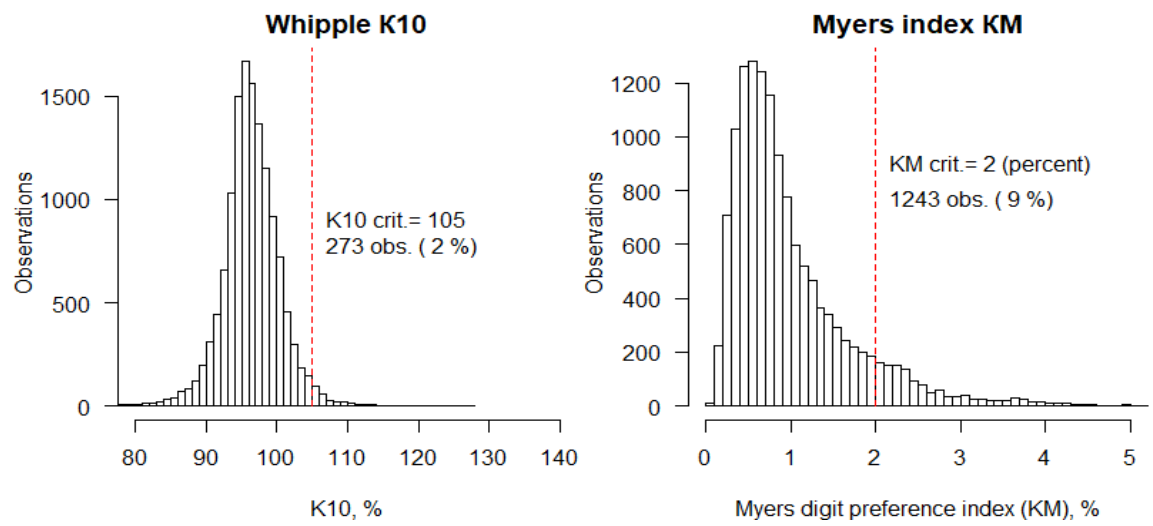
# Need for Interpolation of Demographic Data

- **Correcting data:** cleaning the age heaping / smoothing data

Census data for African countries  
(UN Statistics Division 2019)



(Human Mortality  
Database 2018)



# Need for Interpolation of Demographic Data

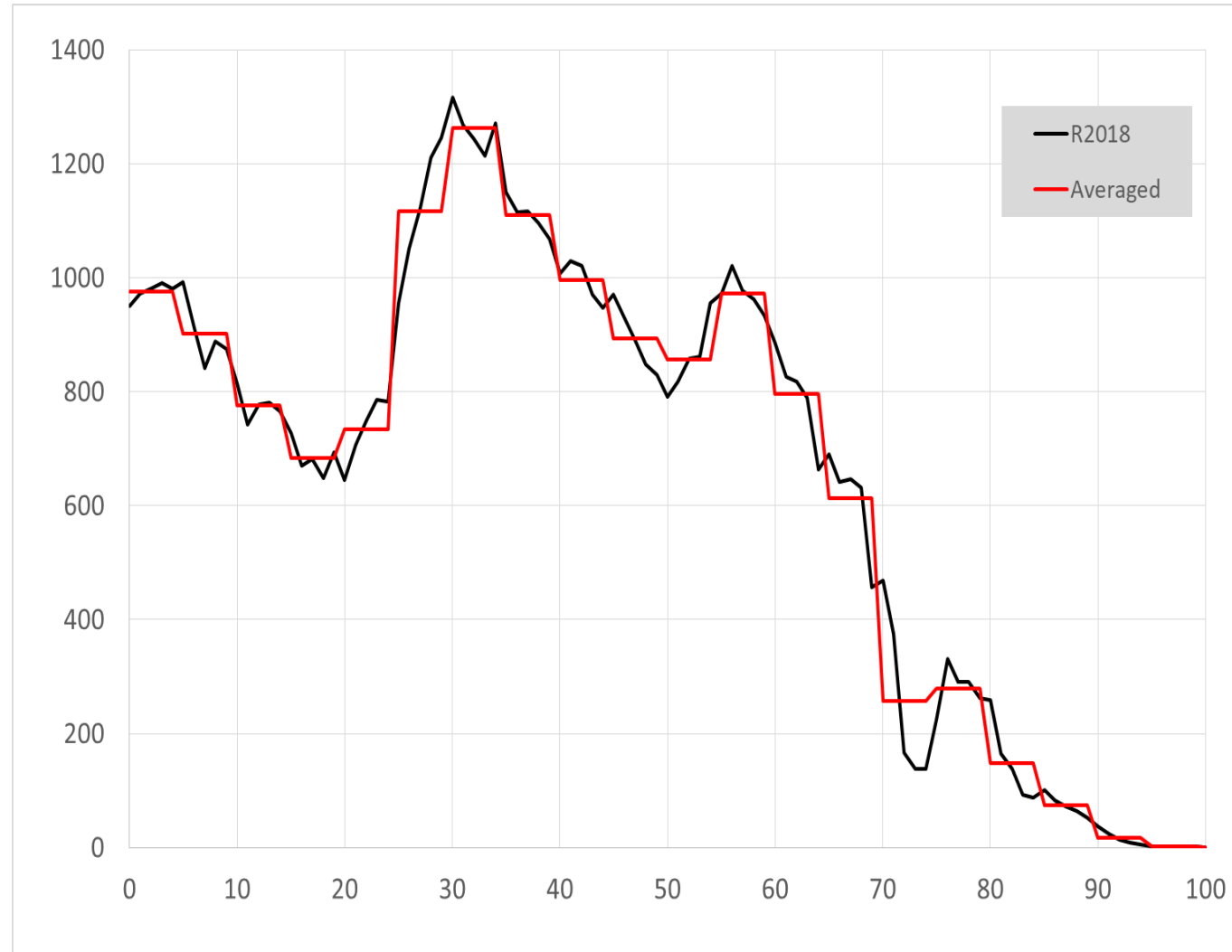
- **Data preparation:** disaggregating population numbers, mortality, fertility, migration into single years of age or Lexis triangles
  - Even advanced data collections (HMD, HFD) partially rely on data splitting
  - In many countries, especially in historical time series, data are typically available in abridged form

# Need for Interpolation of Demographic Data

- **Enhancing population projection's output:** reconstructing the missing age details
  - World population projections (WPP, IIASA) are abridged into 5+ years of age/time [but Eurostat **does it in single years** of age!]
  - Advanced population projections (e.g., probabilistic, multi-regional) might be abridged to save computer time/memory
  - But: certain applications need data in single years of age  
(studying population ageing; pensions; consumption; labor; cohort studies; etc.)

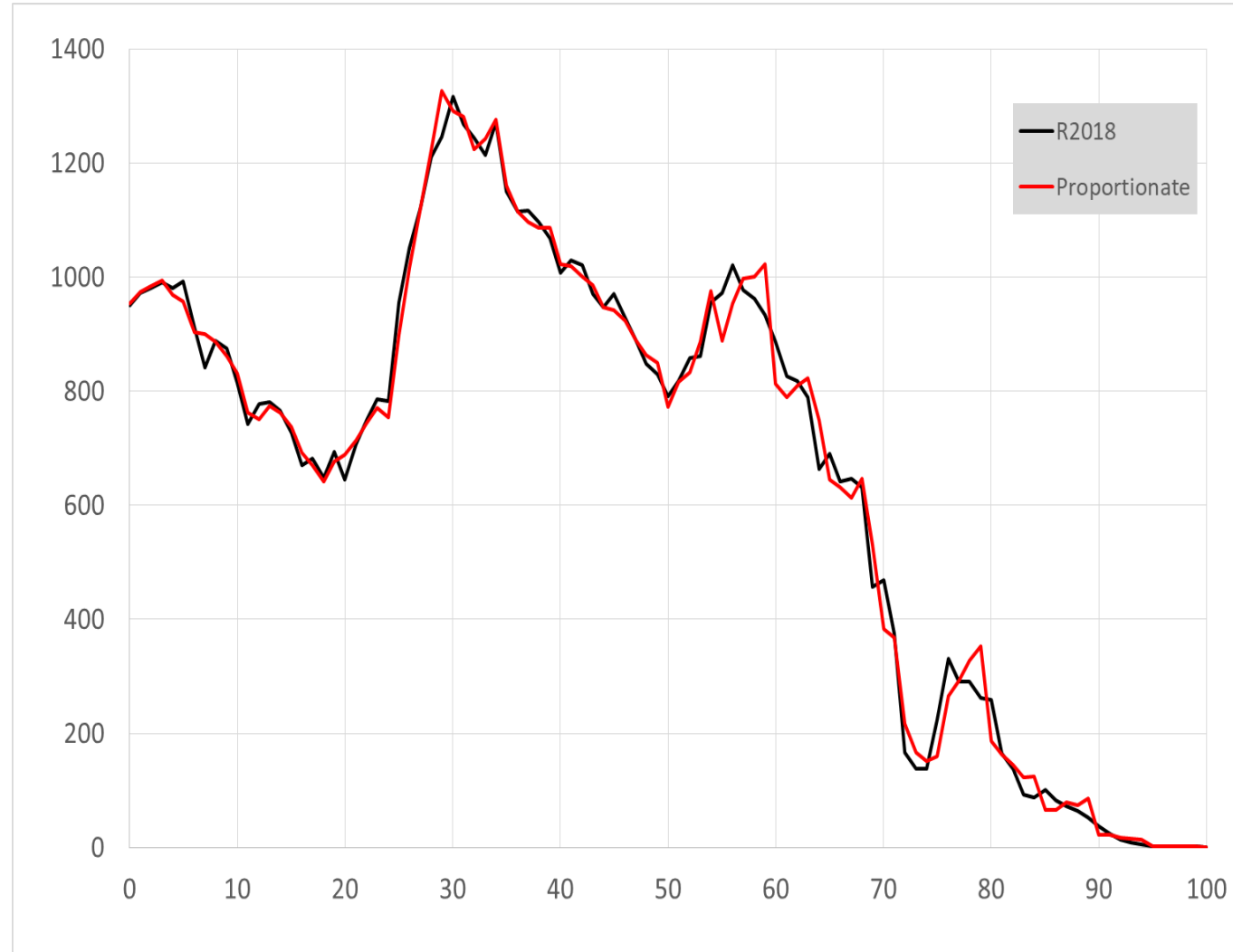
# Traditional approaches

- **Averaging**, Proportionate allocation: discontinuous patterns
- Interpolating polynomials, splines, etc.: results not consistent with the abridged tables; smooth out important details
- TOPALS (Joop de Beer 2012), piece-wise linear ratios to the standard profile of rates: not consistent with the abridged tabulations
- Monotone splines: smooth out important details
- (Rates') Demographic models (Gompertz, relational, model tables, rogers-Castro profiles): not consistent with the original data



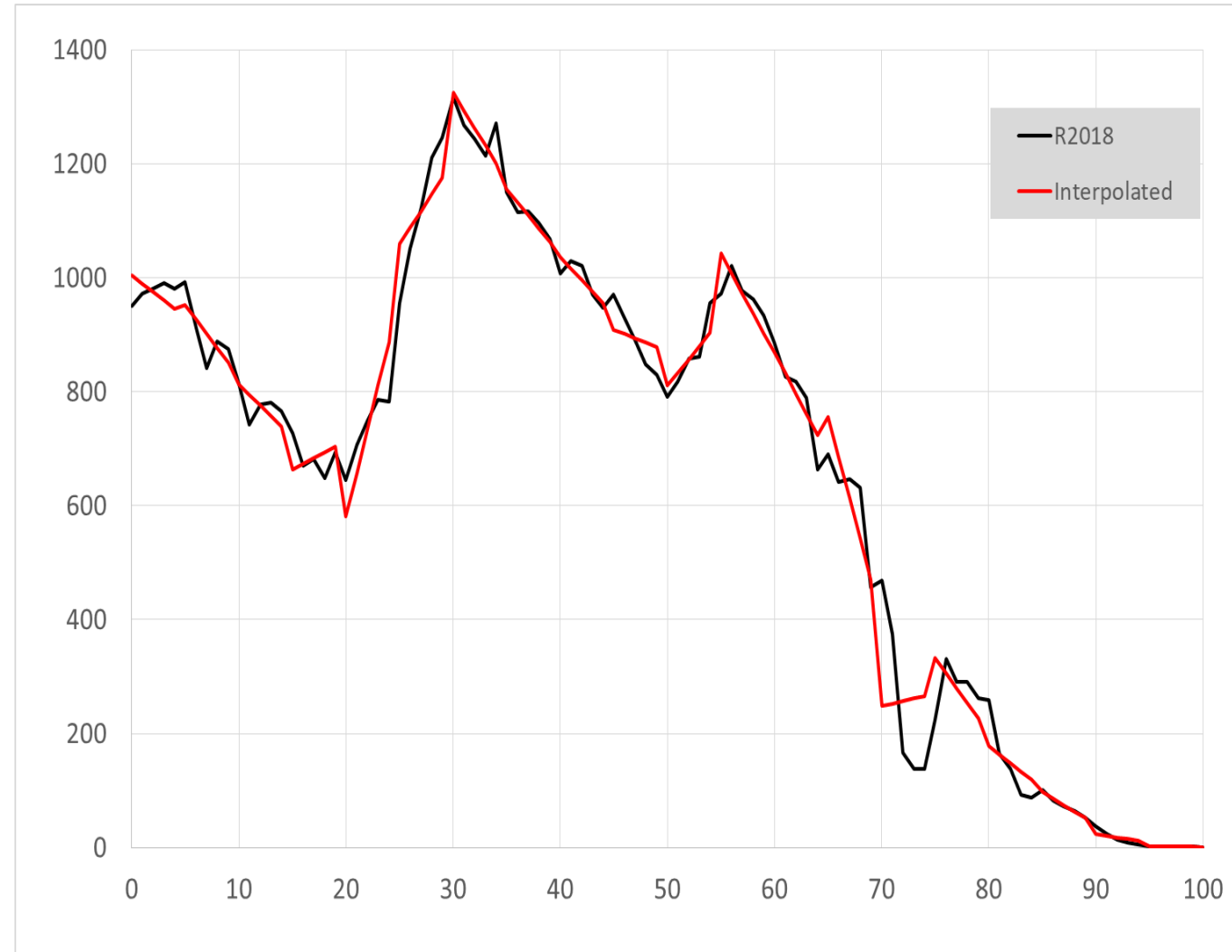
# Traditional approaches

- Averaging, **Proportionate allocation**: discontinuous patterns
- Interpolating polynomials, splines, etc.: results not consistent with the abridged tables; smooth out important details
- **TOPALS** (Joop de Beer 2012), piece-wise linear ratios to the standard profile of rates: not consistent with the abridged tabulations
- Monotone splines: smooth out important details
- (Rates') Demographic models (Gompertz, relational, model tables, rogers-Castro profiles): not consistent with the original data



# Traditional approaches

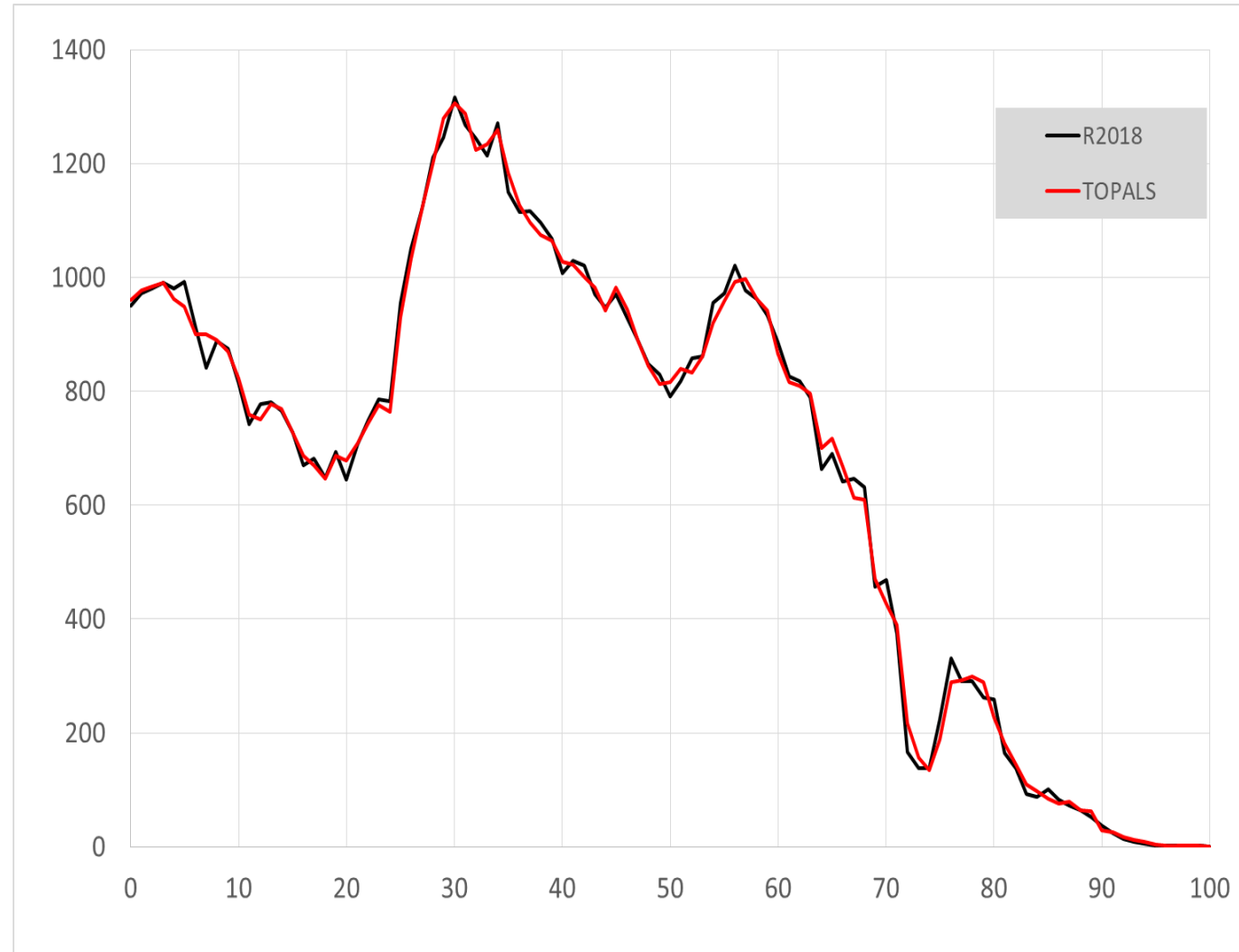
- Proportionate allocation, averaging: discontinuous patterns
- **Interpolating polynomials, splines, etc.:** results not consistent with the abridged tables; smooth out important details
- TOPALS (Joop de Beer 2012), piece-wise linear ratios to the standard profile of rates: not consistent with the abridged tabulations
- Monotone splines: smooth out important details
- (Rates') Demographic models (Gompertz, relational, model tables, rogers-Castro profiles): not consistent with the original data





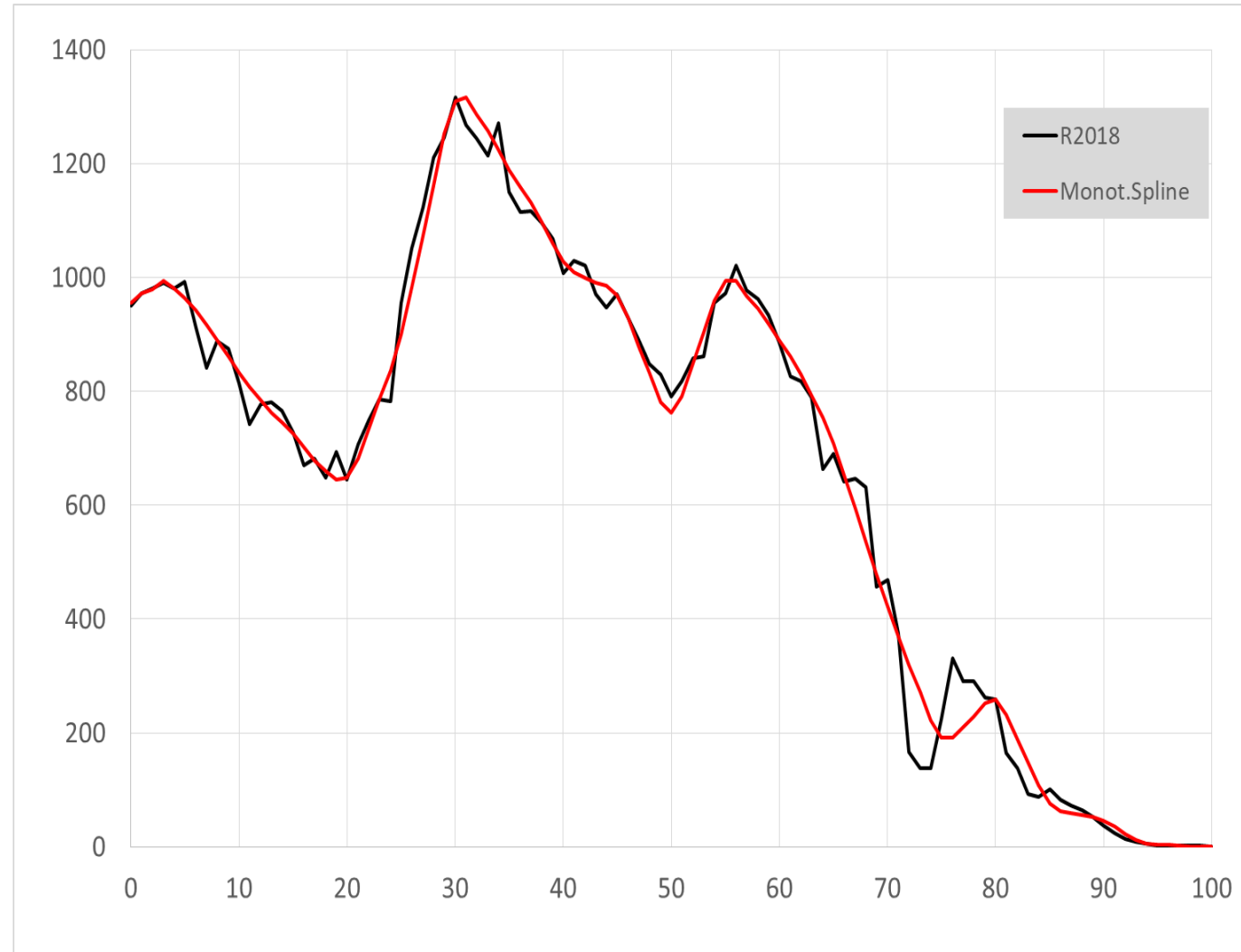
# Traditional approaches

- Proportionate allocation, averaging: discontinuous patterns
- Interpolating polynomials, splines, etc.: results not consistent with the abridged tables; smooth out important details
- **TOPALS** (Joop de Beer 2012), piece-wise linear ratios to the standard profile of rates: not consistent with the abridged tabulations
- Monotone splines: smooth out important details
- (Rates') Demographic models (Gompertz, relational, model tables, rogers-Castro profiles): not consistent with the original data



# Traditional approaches

- Proportionate allocation, averaging: discontinuous patterns
- Interpolating polynomials, splines, etc.: results not consistent with the abridged tables; smooth out important details
- TOPALS (Joop de Beer 2012), piece-wise linear ratios to the standard profile of rates: not consistent with the abridged tabulations
- **Monotone splines:** smooth out important details
- (Rates') Demographic models (Gompertz, relational, model tables, rogers-Castro profiles): not consistent with the original data



# Traditional approaches

- Proportionate allocation, averaging: discontinuous patterns
- Interpolating polynomials, splines, etc.: results not consistent with the abridged tables; smooth out important details
- TOPALS (Joop de Beer 2012), piece-wise linear ratios to the standard profile of rates: not consistent with the abridged tabulations
- Monotone splines: smooth out important details
- (Rates') Demographic models (Gompertz, relational, model tables, Rogers-Castro profiles): not consistent with the original data

# Monotone Spline

1. Cumulate the (abridged) data

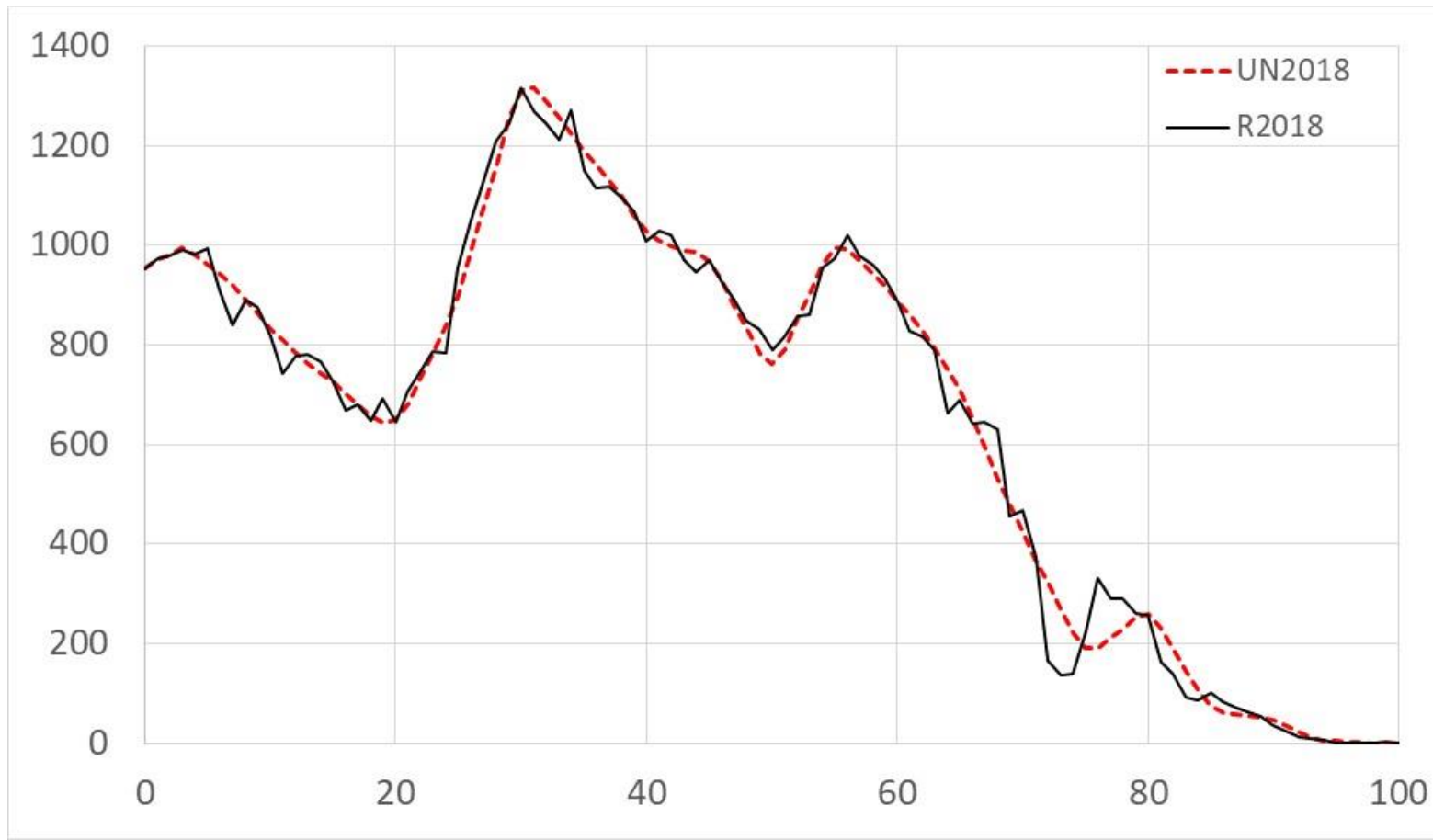
$$\text{cum}P_y = P_{y_1} + P_{y_2} + P_{y_3} + \dots + P_y$$

2. Fit a monotone spline (might be different versions of)

$$S(y) \sim \text{cum}P_y(y)$$

3. Split the original data by differentiating the spline function

$$P_x = S(x + 1) - S(x)$$



# Relational Monotone Spline

1. Cumulate the (abridged) data & the reference

$$cumP_y = P_{y1} + P_{y2} + P_{y3} + \dots + P_y$$

$$cumP_y^* = P_{y1}^* + P_{y2}^* + P_{y3}^* + \dots + P_y^*$$

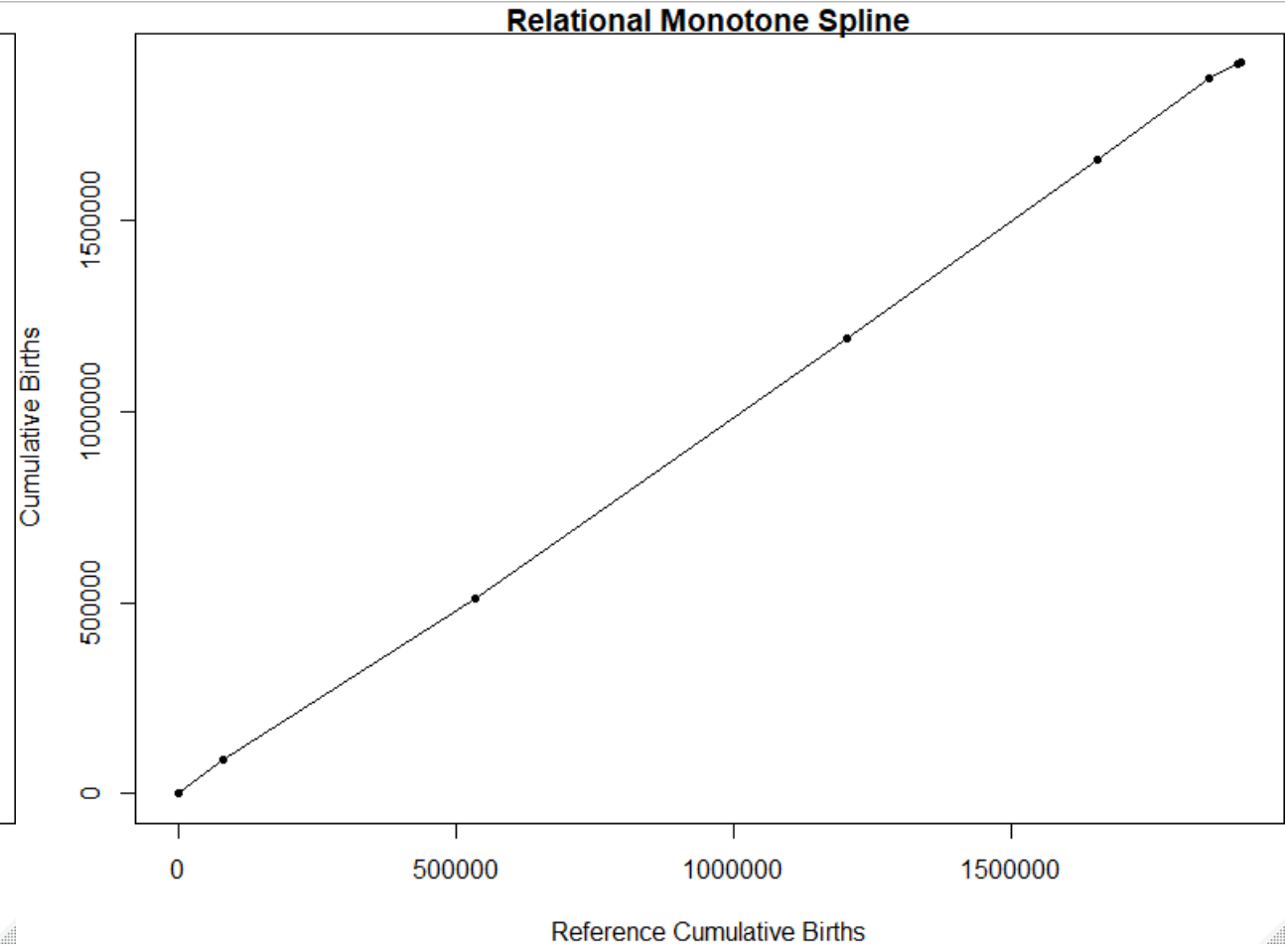
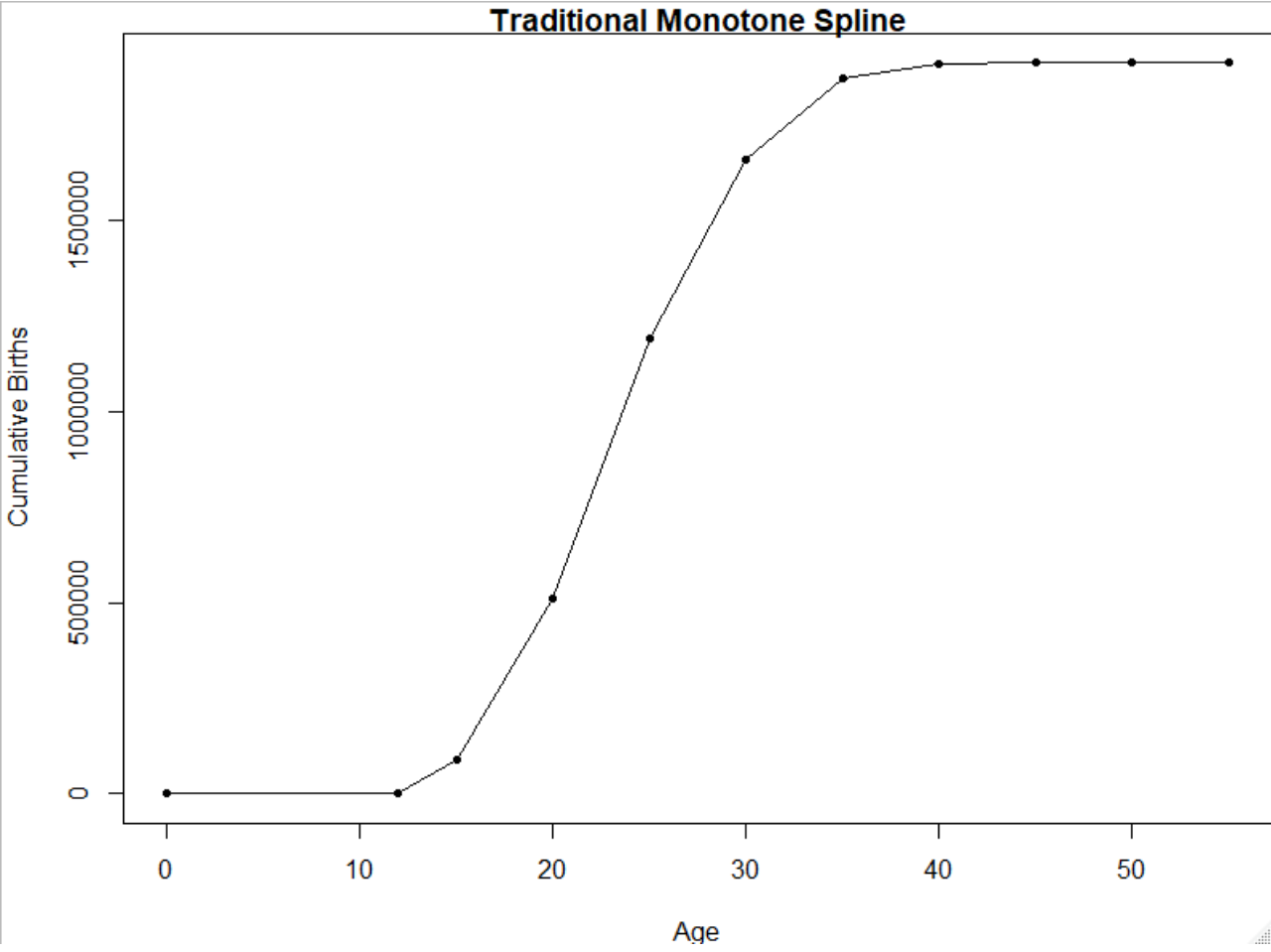
2. Fit a monotone spline relating the observed and reference cumulants:

$$S(p) \sim cumP_y(cumP_y^*)$$

3. Split the original data by differentiating the spline function

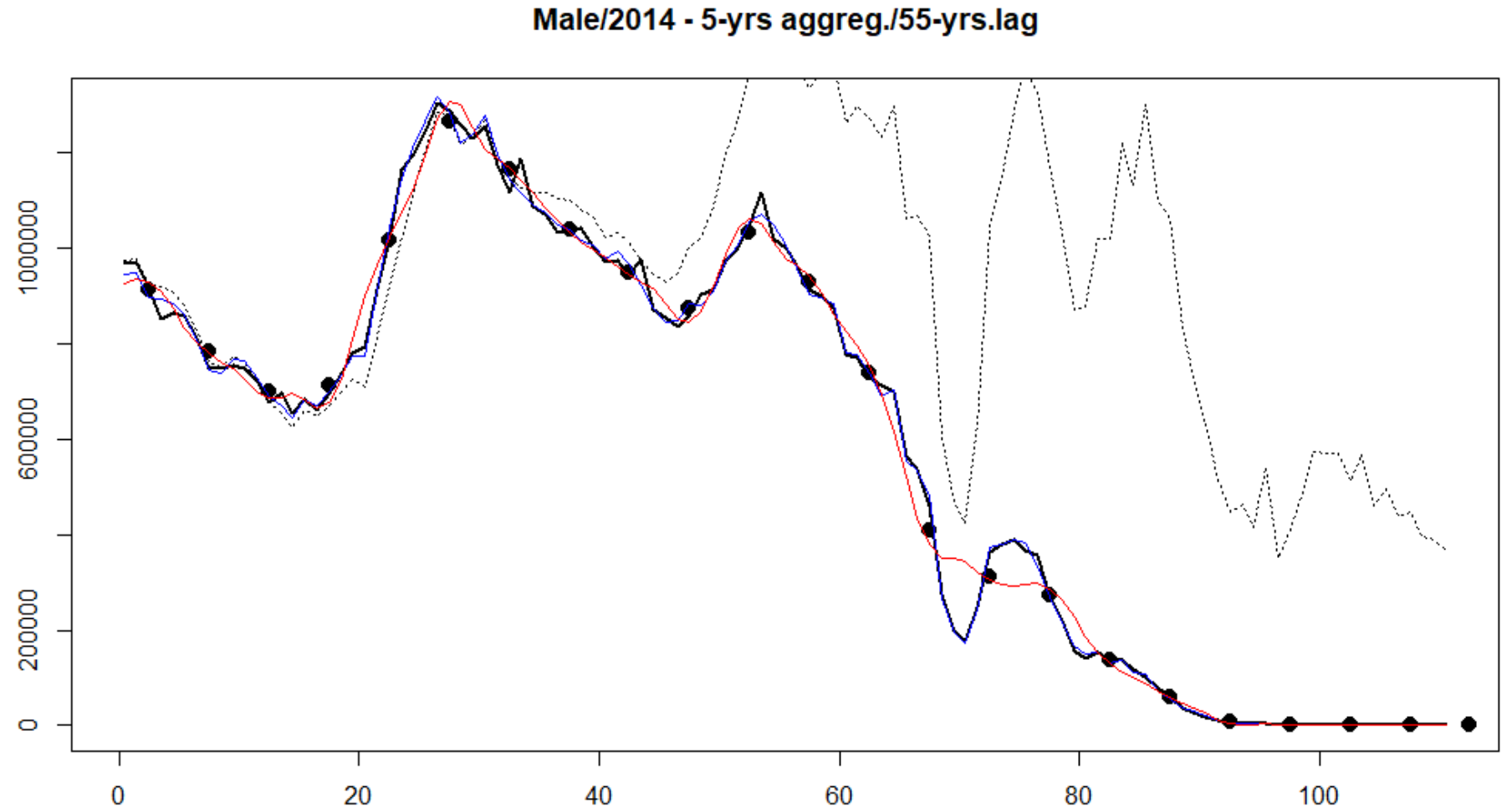
$$P_x = S(cumP_{x+1}^*) - S(cumP_x^*)$$

# Relational Monotone Spline



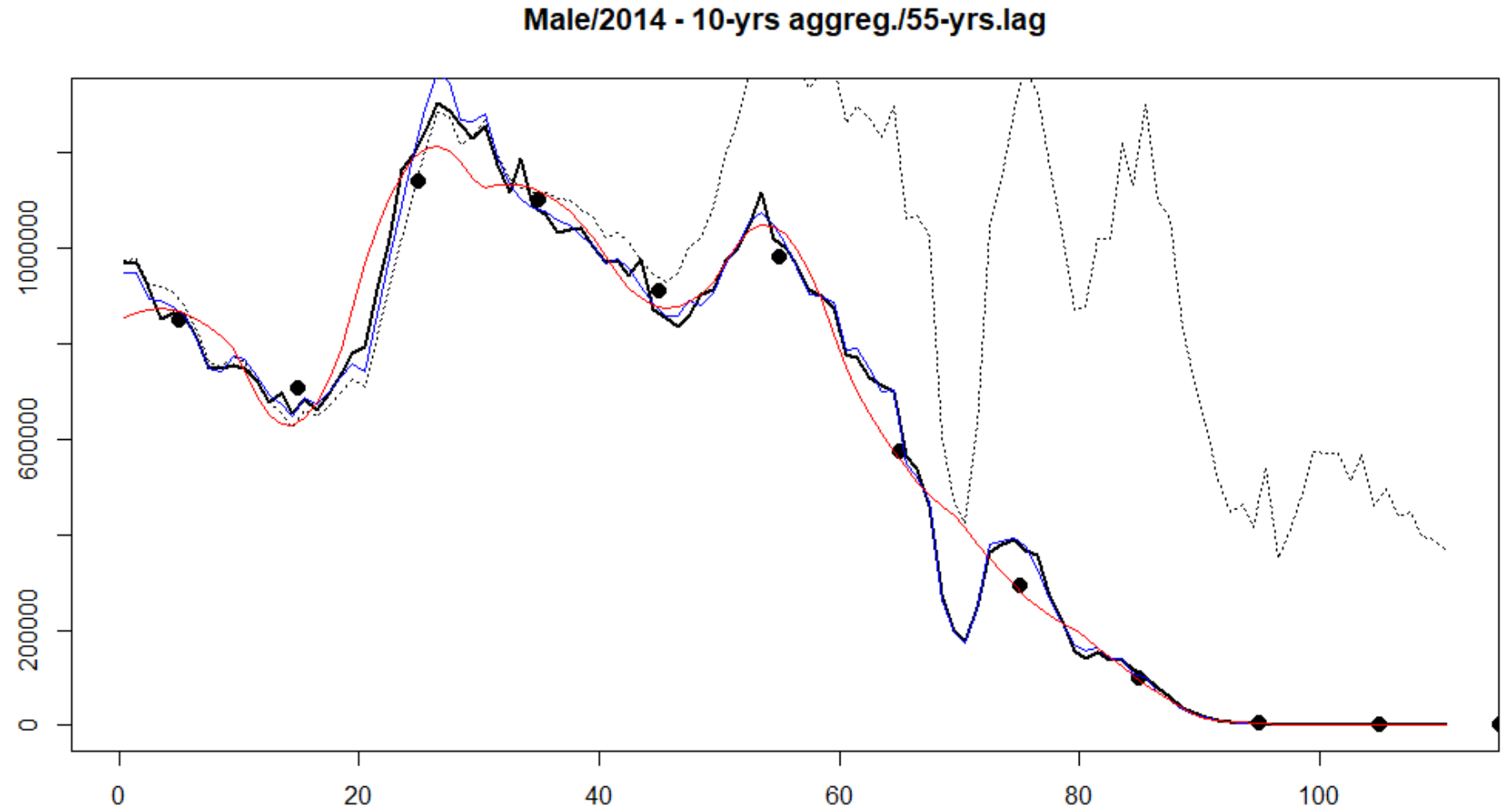
# Relational Monotone Spline

Interpolating 2014 Russian population data (males) with **births  $B(t)+P(1959)$**  as a reference:  
Data aggregated into **5-years age intervals**



# Relational Monotone Spline

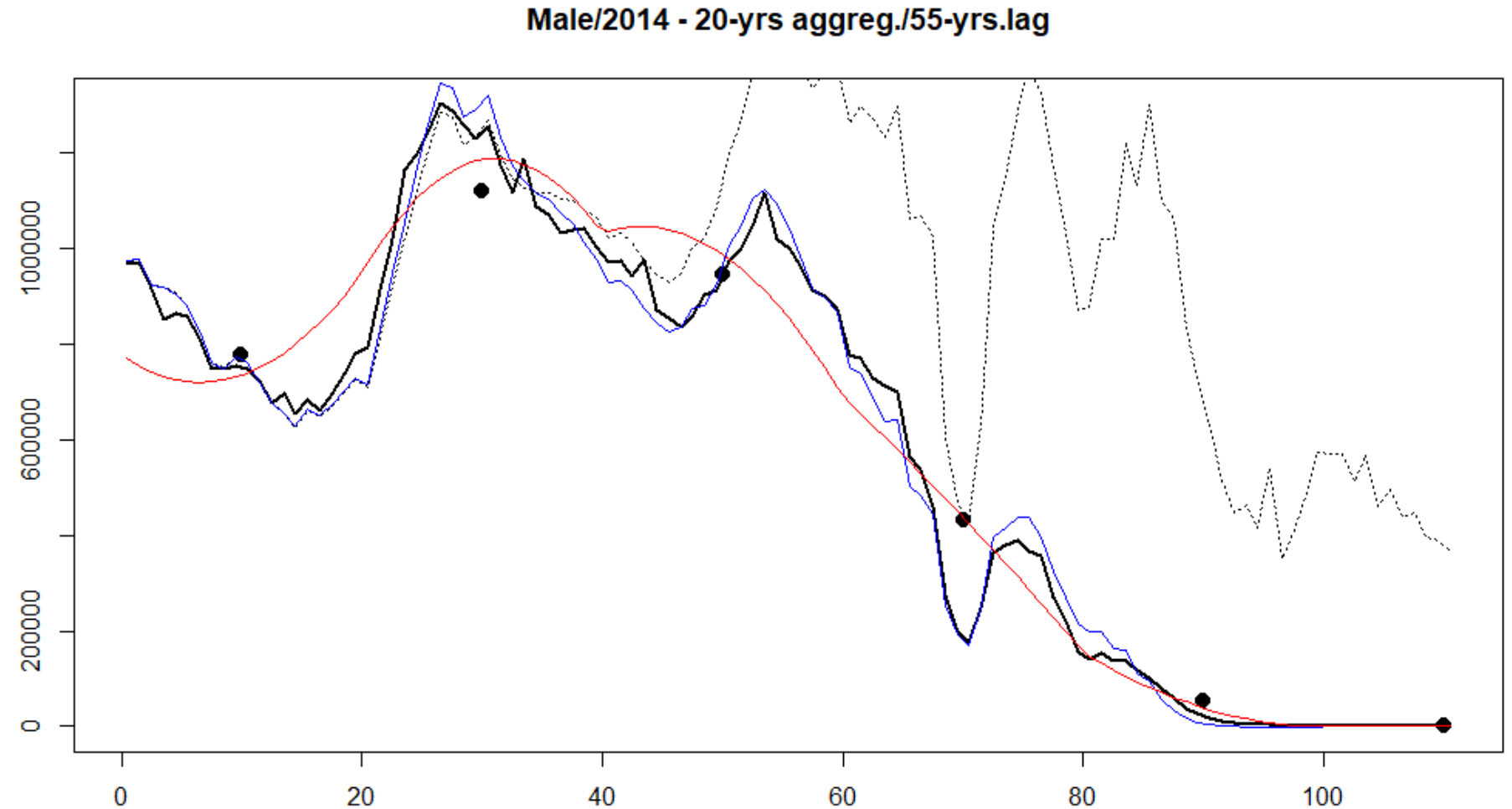
Interpolating 2014 Russian population data (males) with **births  $B(t)+P(1959)$**  as a reference:  
Data aggregated into **10-years age intervals**





# Relational Monotone Spline

Interpolating 2014 Russian population data (males) with **births  $B(t)+P(1959)$**  as a reference:  
Data aggregated into **20-years age intervals**

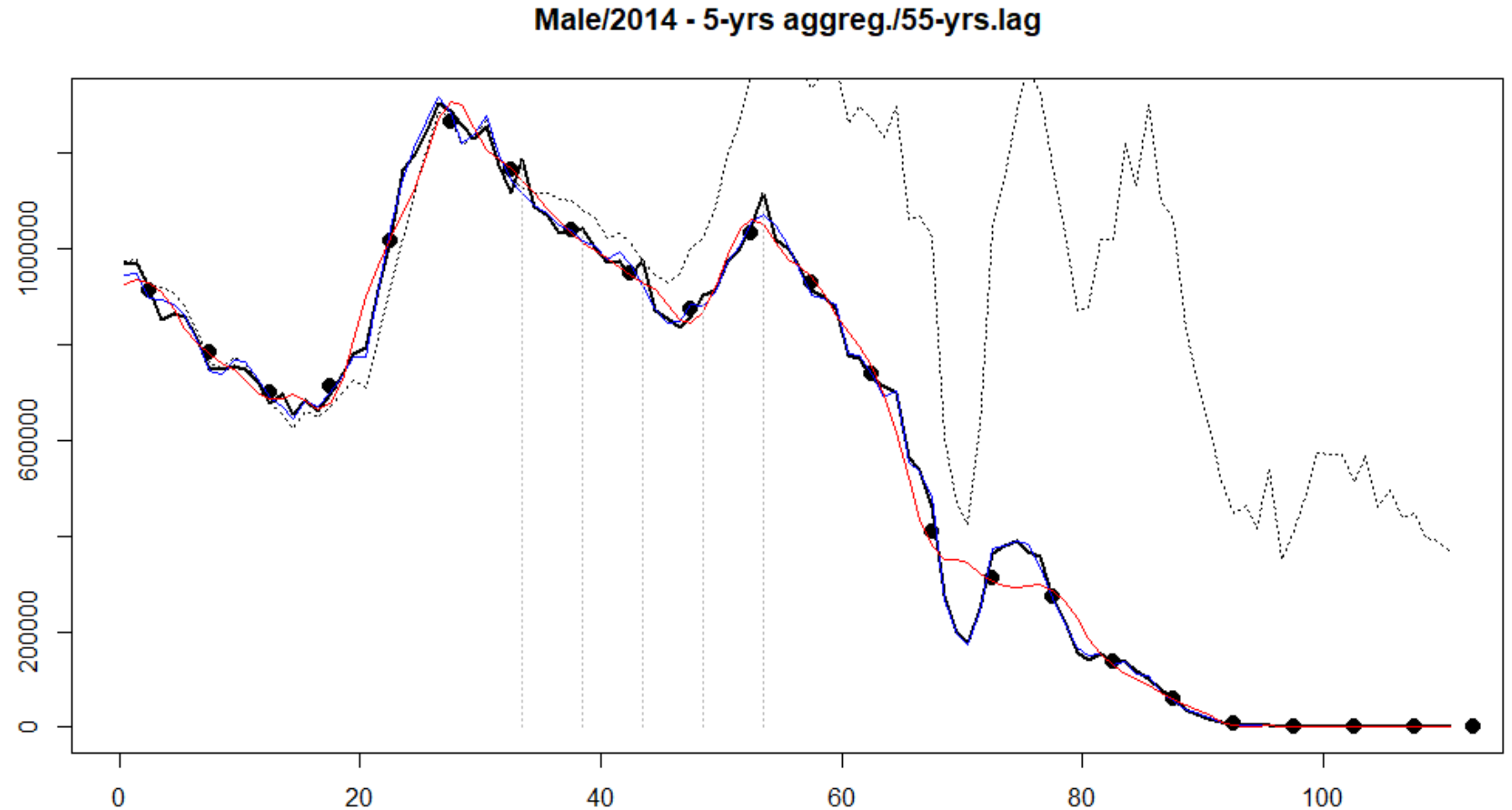


# Relational Monotone Spline

## Correcting data (population)

Interpolating 2014 Russian population data (males) with **births  $B(t)+P(1959)$**  as a reference:

Data aggregated into **5-years age intervals**



# Relational Monotone Spline: rates

Assume:

- detailed population exposure by age;  $P_x$
- abridged rates;  $M_y$
- detailed reference rates/events by age  $M_x^*$   $D_x^*$

1. Abridge the population exposure and calculate abridged events (observed and reference)

$$D_y = P_y M_y, \quad D_y^* = P_y M_y^*$$

2. Apply the relational spline and split the events into single years of age

$$S(d) \sim \text{cum}D_y(\text{cum}D_y^*)$$

$$D_x = S(\text{cum}D_{x+1}^*) - S(\text{cum}D_x^*)$$

3. Produce the detailed rates by dividing the events by detailed population exposure

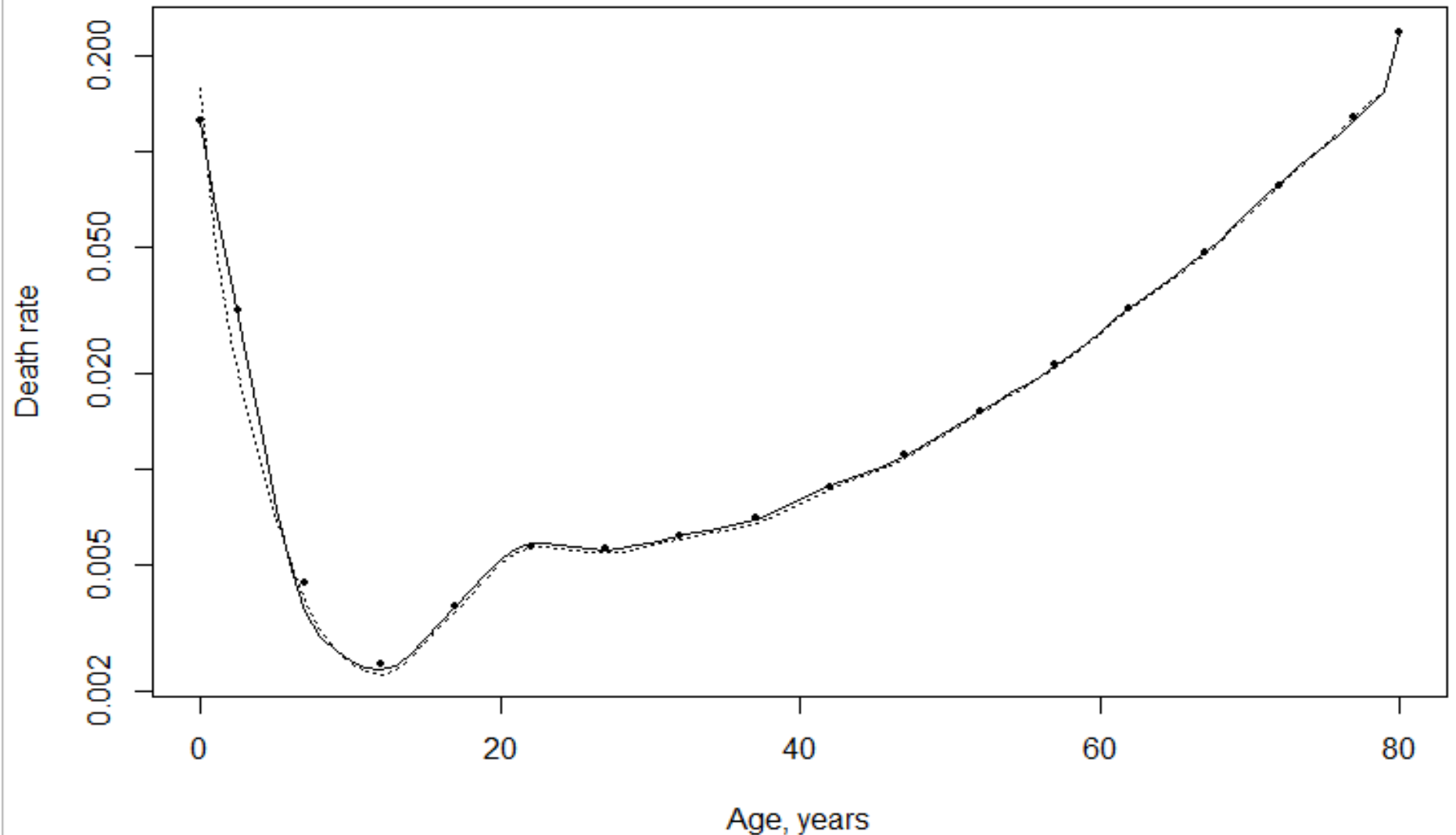
$$M_x = \frac{D_x}{P_x}, \quad x = 0, 1, 2, \dots$$

# Relational Monotone Spline: Preparing data (rates)

Abridged (dots) and interpolated death rates for **Burkina Faso, males, 1985-1990**.

Reference population: accumulation-removed interpolation.

Reference rates: UN model life table (broken line).

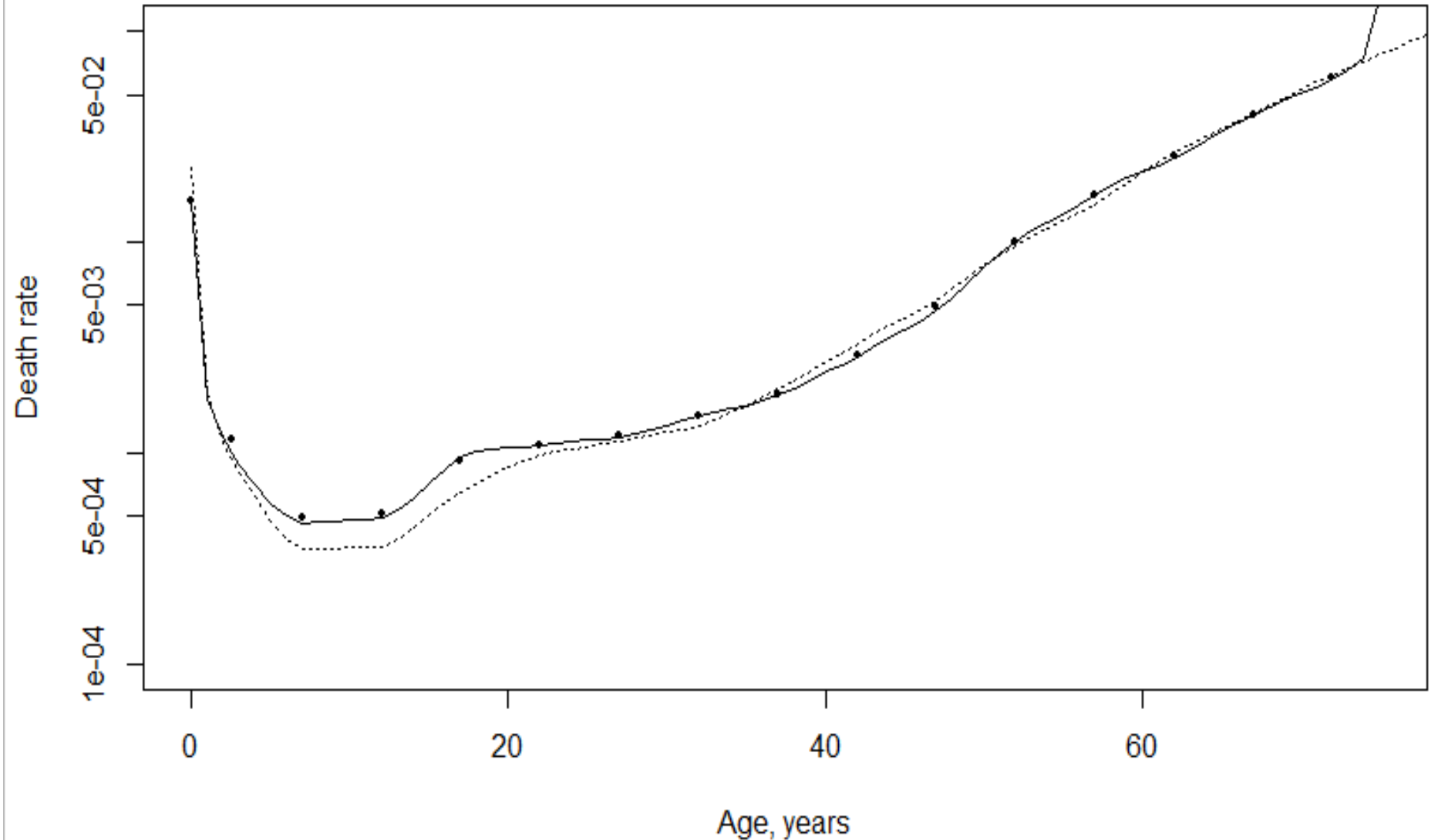


# Relational Monotone Spline: Preparing data (rates)

Abridged (dots) and interpolated death rates for **Egypt, males, 2017**.

Reference population: the reference Life Table

Reference rates: UN model life table (broken line).

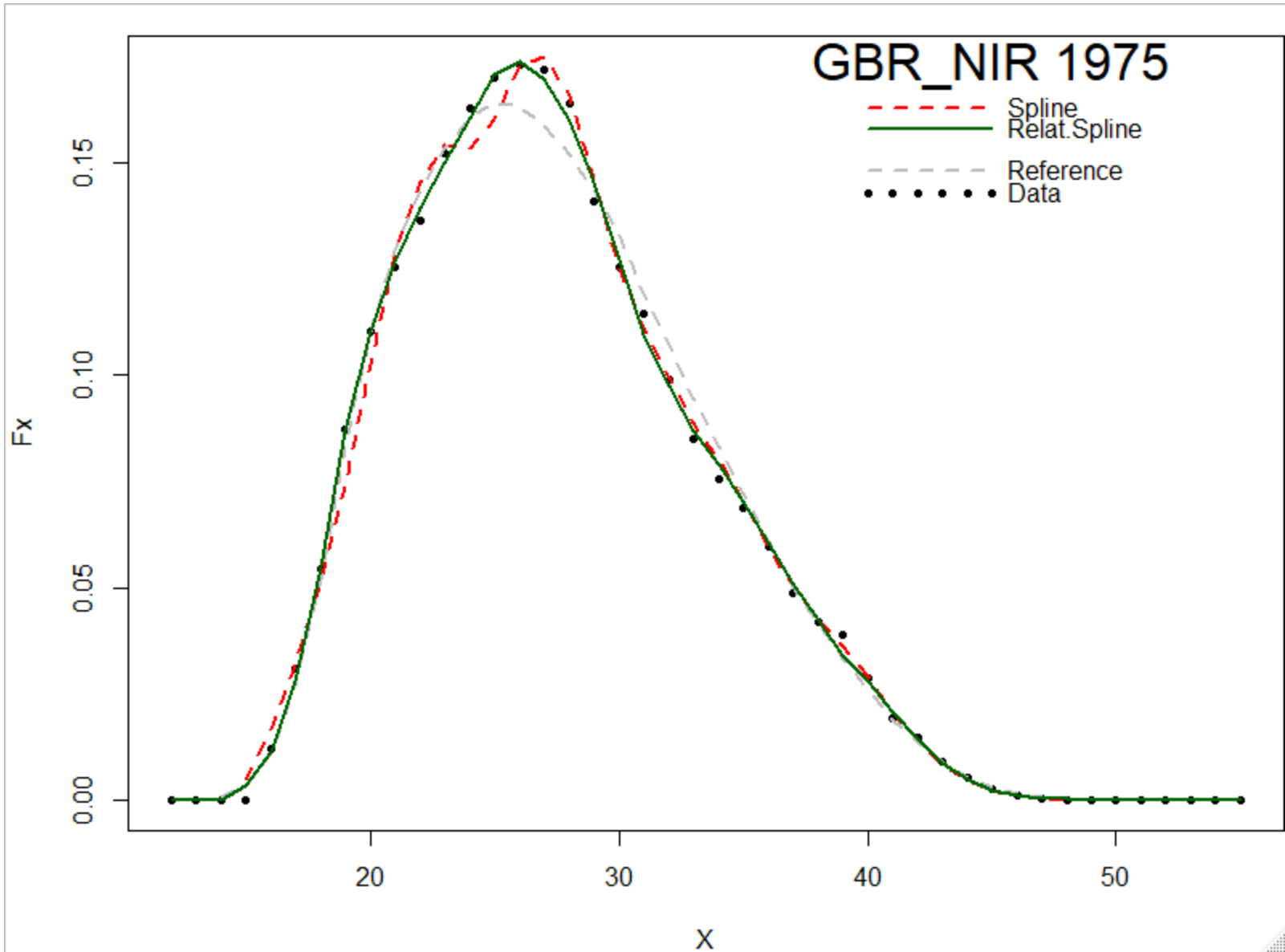


# Relational Monotone Spline: Preparing data (rates)

Abridged (dots) and interpolated fertility rates for GBR\_NIR 1975 (HFD).

Reference rates: the regression model over HFD  $F_x \sim TF + MA$  (Ediev 2013) (grey line)

Traditional spline interpolation: red broken line.

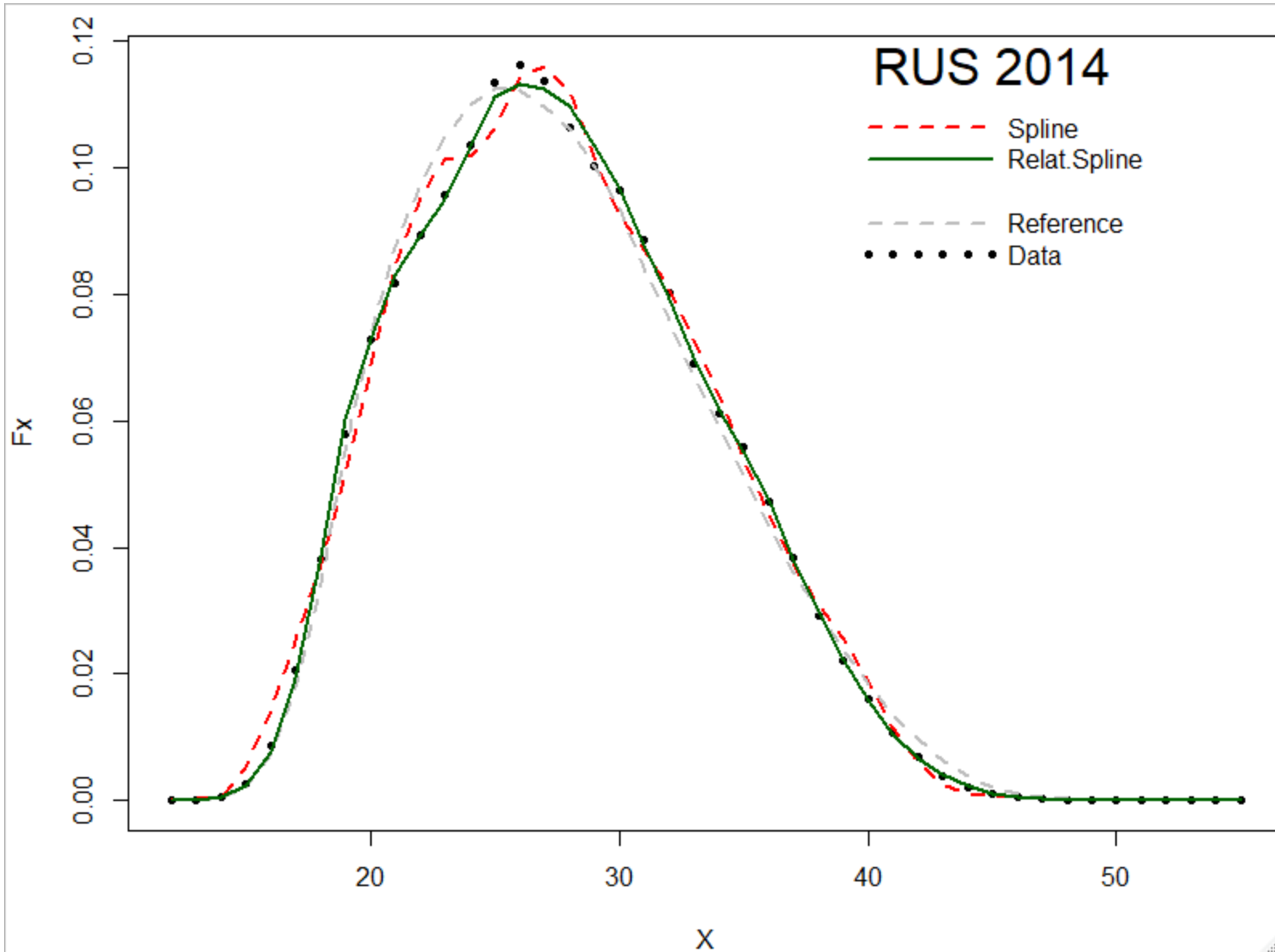


# Relational Monotone Spline: Preparing data (rates)

Abridged (dots) and interpolated fertility rates for RUSSIA 2014 (HFD).

Reference rates: the regression model over HFD  $F_x \sim TF + MA$  (Ediev 2013) (grey line)

Traditional spline interpolation: red broken line.



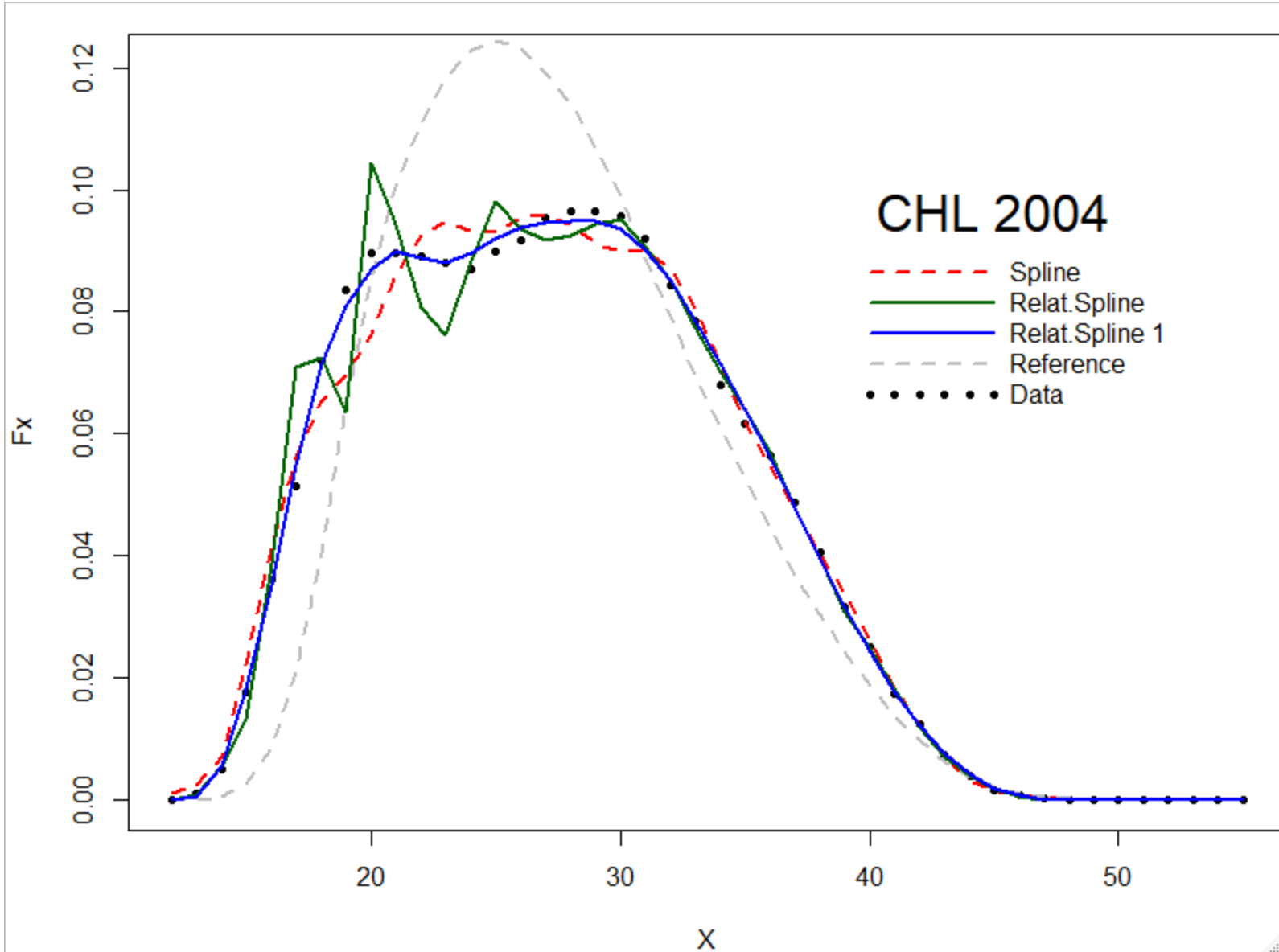
# Relational Monotone Spline: Preparing data (rates)

Abridged (dots) and interpolated fertility rates for GBR\_NIR 1975 (HFD). (Rel.Spline1: reference=smoothed RelSpline)

Reference rates: the regression model over HFD  $F_x \sim TF + MA$  (Ediev 2013) (grey line)

Traditional monotone spline: red broken line.

Bad reference





# Relational Monotone Spline: Enhancing projection outputs

## UN projections:

- Consistent methodology over all countries
- Long range
- Authoritative

## **Yet,**

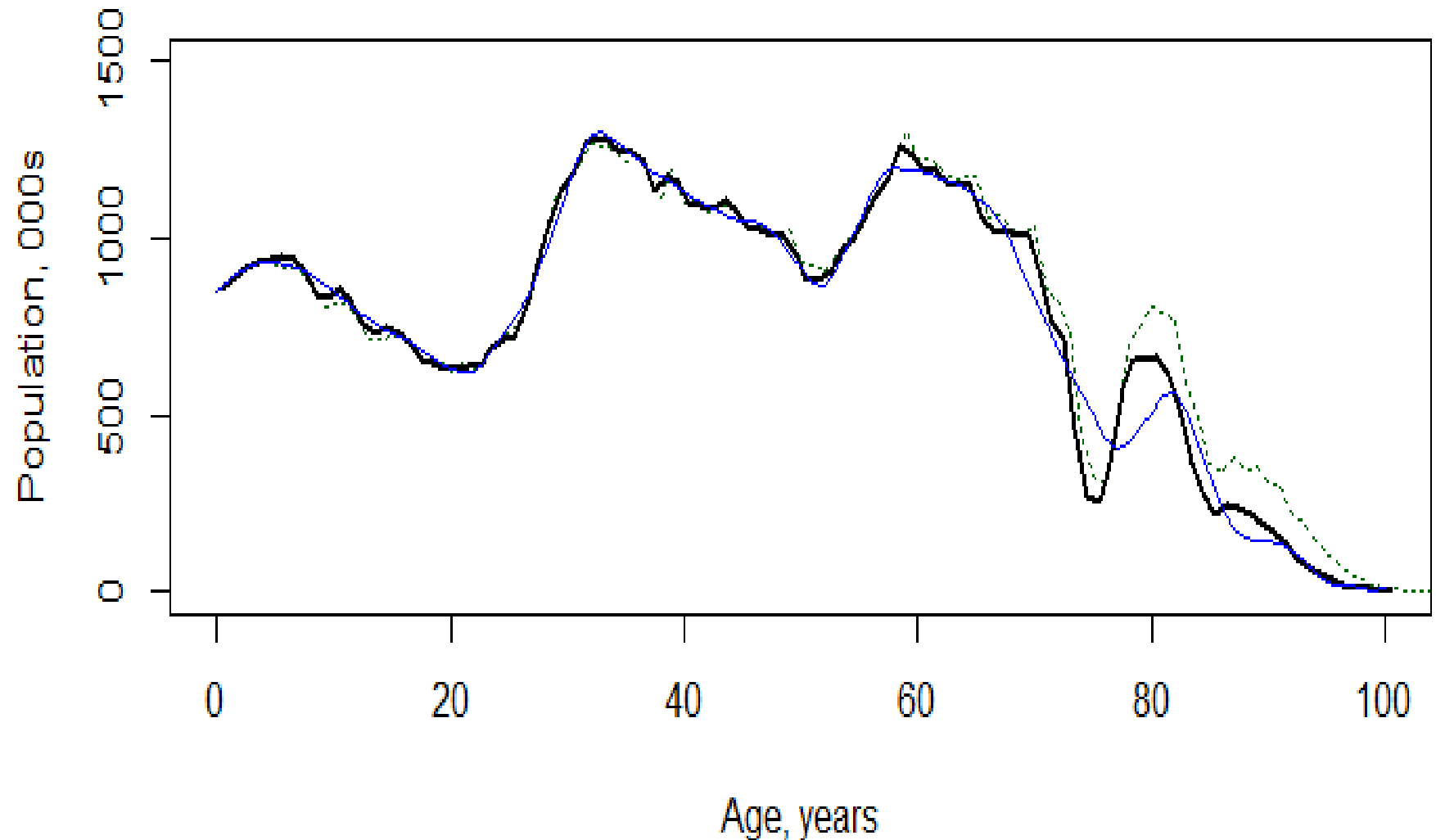
- Abridged tabulations
- (Splines-based) Interpolations too rough a smoothers for countries with fast demographics in the past (e.g. Russia, cohorts 1940s, 1990s)
- No individual cohorts traceable (e.g., in pension analysis)

# Relational Monotone Spline: Enhancing projection outputs

Interpolated UN projection for **Russian women in 2020** (the thick solid line);

the UN's own interpolation (the thin **blue line**);

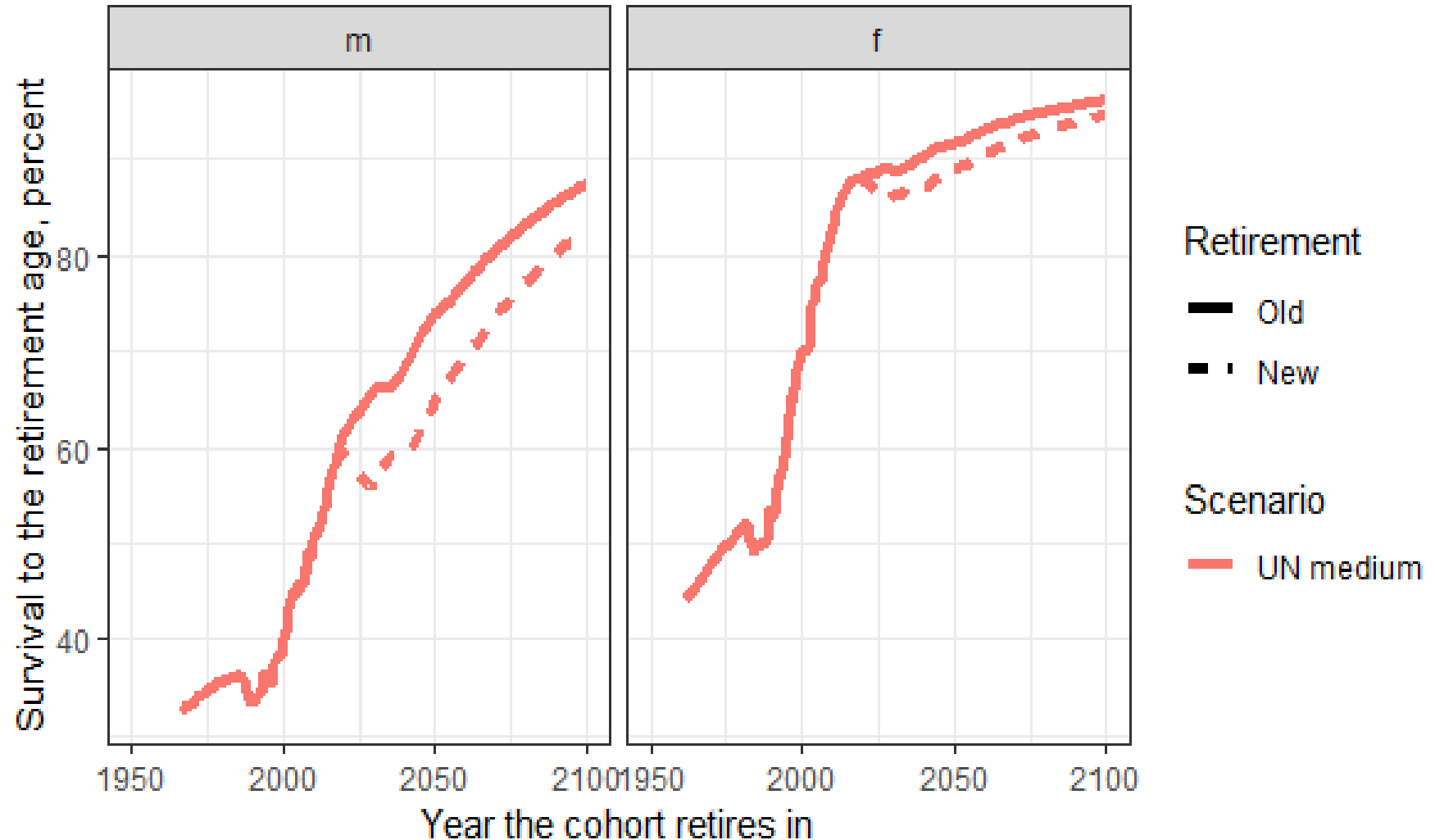
and the 2015 age structure (HMD) shifted by five years (the broken **green line**).



# Relational Monotone Spline: Enhancing projection outputs

**Survival (percent)** of Russian cohorts of men (left) and women (right) **to the old (the solid lines) and the new (the broken lines) ages at retirement\***.

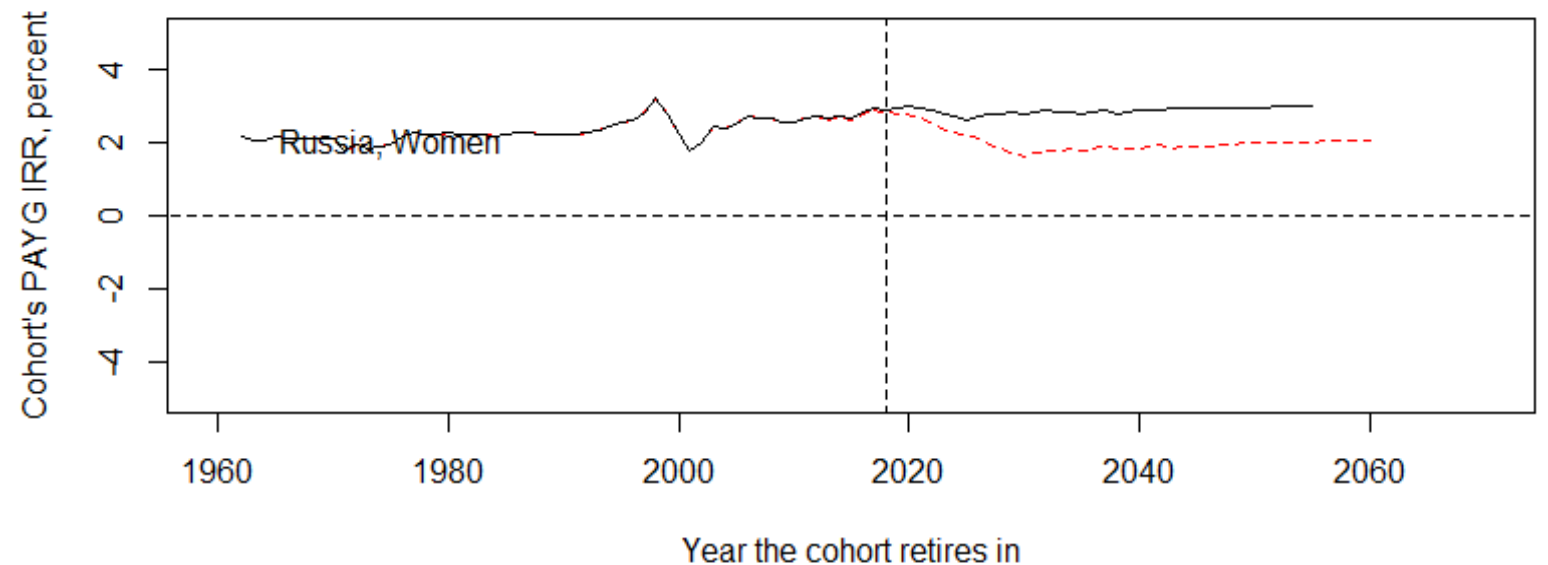
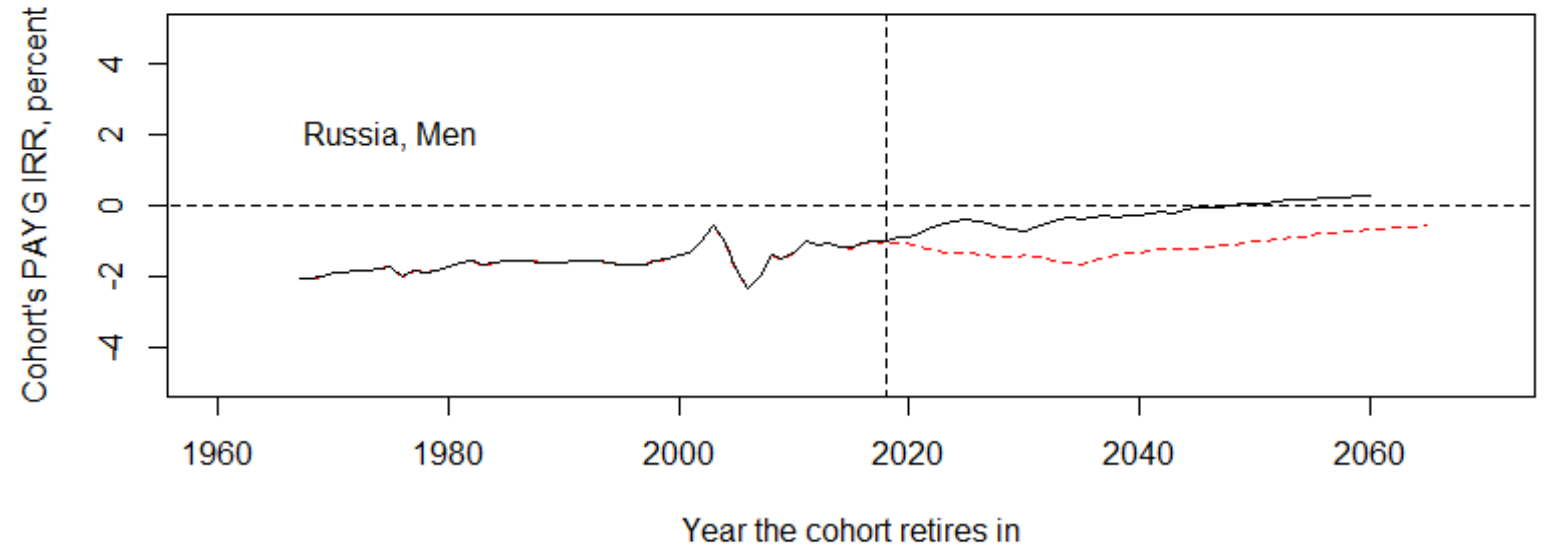
\* 55/60 to 60/65 by 2028



# Relational Monotone Spline: Enhancing projection outputs

**PAYG IRRs' for cohorts of Russian men (up) and women (down) under the old (the solid lines) and the new (the broken lines) ages at retirement\*.**

\* 55/60 to 60/65 by 2028



# Relational Monotone Spline: further steps

- Interpolating various demographic indicators (fertility, headship, marriage and divorce, migrations, etc.)
- 2D, 3D, multi-D extensions:
  - interpolating by age- parity-; by age-time-; by age-cohort, etc.
  - by age- region-?
- Non-local smoothness
- Bridging the (possibly non-representative) BIG data & surveys to traditional (representative) statistics
- Speeding-up (probabilistic) population projections

Thank you!

[Dalkhat@hotmail.com](mailto:Dalkhat@hotmail.com)

[Ediev@iiasa.ac.at](mailto:Ediev@iiasa.ac.at)