HFD: History, challenges and future plans in a nutshell (HFD Round Table Discussion)

Tomáš Sobotka

Vienna Institute of Demography (VID)

First Human fertility Database Symposium
MPIDR Rostock, 4 November 2011

The HFD (hi)story

The dawn of HFD

Initial motivation

- Lack of freely available, high-quality, standardised & detailed fertility data

- Scattered datasets not fully available to other researchers (Willy Bosveld, ODE (JP Sardon & G Calot)), not properly documented (Eurostat New Cronos), or focused on individual countries

- Major efforts to collect individual cross-national data

- The decline of aggregate-level research in fertility oddly contrasting with the prominence of debates about low fertility

- The sad dominance of the period TFR

- Facilitating the use of cohort & order-specific data

- Observing the success of the HMD

The prehistory

(*2002 HMD launched*)

2004   1st HMD Symposium: J Wilmoth suggests fertility DB

2006   First discussions about a possibility of setting up HFD

2007   *Things get moving*….

- Josh Goldstein decides to set up an HFD as a MPIDR funded project
- Wolfgang Lutz supports HFD-related work at VID
- Dimiter Philipov & T Sobotka (VID) and Laurent Toulemon (NIDI) include HFD initiation as one of the goals of EU-funded project REPRO

    → The HFD born as a collaboration of MPIDR & HFD; Josh Goldstein becomes HFD director

(Initial) Challenges

- People: *Methods setup, website creation & maintenance, data collection, checks, and documentation, programming*

- Funding: *Researchers and data*

- Data: *There is nothing like a perfect dataset…*

- Focus: *quality vs. quantity*

- Methods: *laissez-faire or a 'control freak' approach*?

- Website design & data format

- Planning

→ *All types of tensions arising: first crash between ideals and reality…compromises needed from the start*

The history (1)

2007   Initial meetings & plans

2008   Initial work & First outcomes

- HFD outline discussed at 2nd HMD symposium in June
- Research teams & agenda established at MPI & VID
- First Methods Protocol
- Division between HFD and Human Fertility Collection (HFC)
- JP Sardon joins with ODE data
- Eurostat willing to collaborate

The first Methods Protocol in June 2008 (Laurent Toulemon's comments)

Figure 4. Relative and absolute difference between proportion of upper triangle in data and combined model. Austria, 1984-2

12
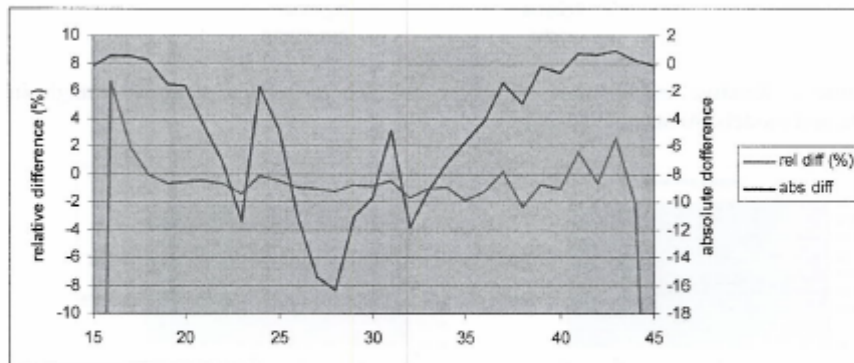
...d on these two different approaches are not comparable.) **?Why?**

...ion, $w_i^r(x)$, initially based on census data is subsequently

...$w_r(x)$ and thus the two estimates converge over time and

We consider using parity distribution from both sources if available and calculating two versions of period fertility table. For countries that do not have census data (e.g., Canada), large sample surveys could be used. Regarding the census-based parity distribution, we would probably use only one initial (starting) census or other estimate of $w_i^r(x)$ and these data would not be continually revised with each subsequent census. This option can be introduced later in the future and sensitivity analysis of such regular updates of $w_i^r(x)$ can be conducted as well.
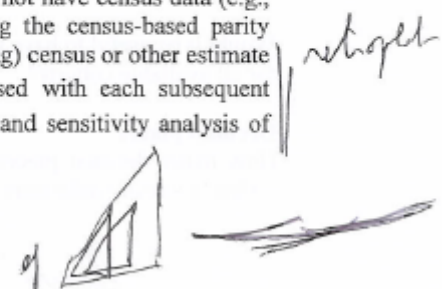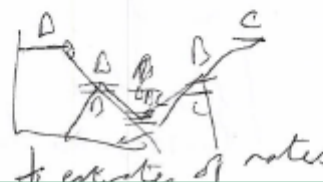
26

The logos...
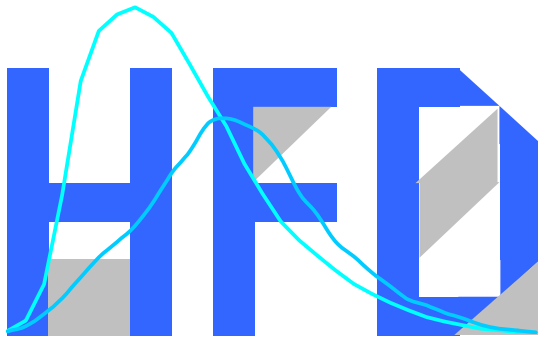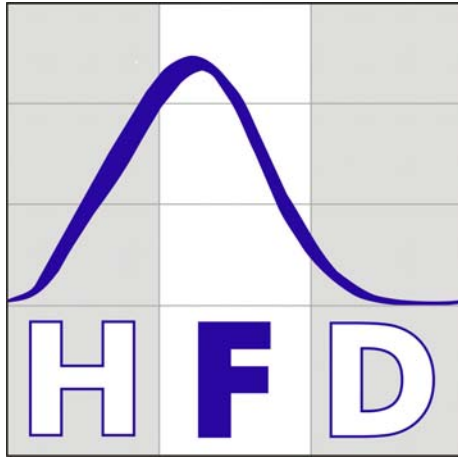
The history (2)

2009   May: HFD prototype ("version 0") presented at the
        PAA meeting in Detroit

        September: Fully functioning website with data for 6
        countries presented at the IUSSP
        meeting in Marrakech

2010   Rapid expansion

2011   New website design (V1), expanded set of data &
        revised methodology

ILLUSTRATION:
Political regime change and cohort first birth rates

BACKGROUND: Has political regime change in the Czech Republic in November 1989 affected first birth trajectories of different cohorts?
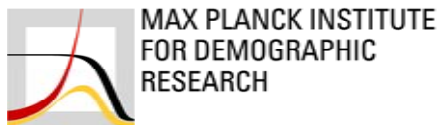

First births rates, m1(x)
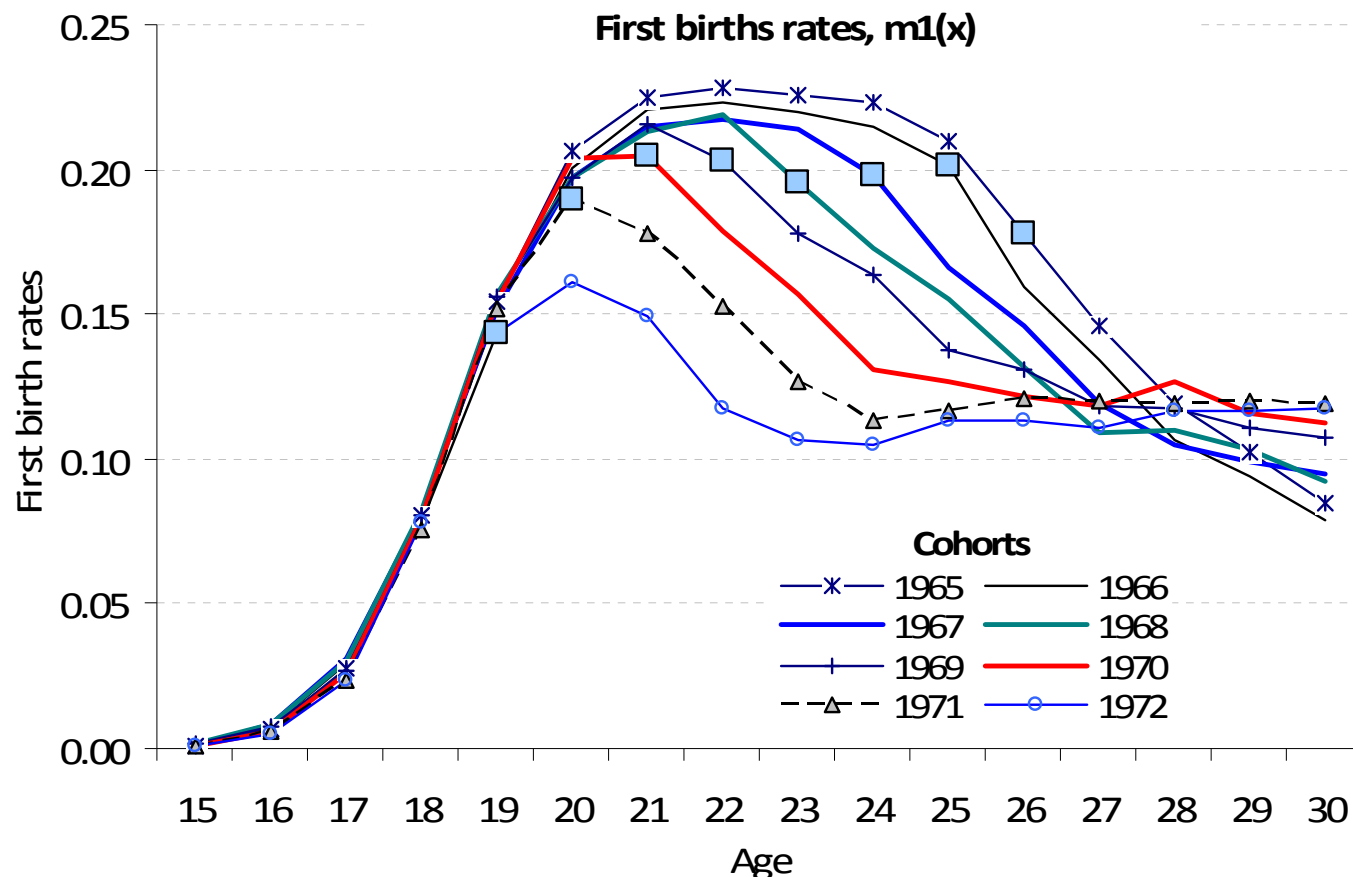
Comparing cohort conditional rates, *m1(x):*
If regime change matters, the trajectories should differentiate since 1991
➢ Blue squares allocated at the age reached in 1991

Source: A. Jasilioniene, T Sobotka, E. M. Andreev, D.A. Jdanov, K. Zeman, V. M. Shkolnikov, and J. R. Goldstein (2010). Based on HFD data.

Current status

Research teams & Advisory Board

HFD Team:

MPIDR:
Joshua R. Goldstein (Director)
Vladimir M. Shkolnikov (Co-Director)
Aiva Jasilioniene (coordinator)
Dmitri A. Jdanov
Dora Kostova
Evgeny M. Andreev
Sigrid Gellers-Barkmann
Karolin Kubisch
Edward J Nash

VID:
Tomáš Sobotka (Co-Director)
Kryštof Zeman

Advisory Board:
John Wilmoth (Berkeley)
Wolfgang Lutz (VID)
Laurent Toulemon (INED)
Jean-Paul Sardon (INED, ODE)

Other researchers involved:
Dimiter Philipov (VID)
Germán Rodríguez (Princeton University)
Evgueni L. Soroko (HSE, Moscow)
….and many others

Countries & areas covered

Current status: 19 countries, 5 regions (UK, Germany)
Population ca. 790 million. Ca 1000 registered users.

Austria	Finland	Netherlands	Sweden
Bulgaria	France	Portugal	Switzerland
Canada	⊞ Germany	Russia	⊞ U.K.
Czech Republic	Hungary	Slovakia	U.S.A.
Estonia	Lithuania	Slovenia	

In the pipeline: coming soon
Belarus, Ireland, Japan, Ukraine (pop. 183 million)

In the pipeline: coming later (some progress or contacts to data providers)
Australia, Belgium, Denmark, Greece, Romania, Iceland, Israel, Italy, Latvia, Luxembourg, New Zealand, Norway, Poland, Spain, Taiwan (pop. 254 million)

Considered in the future & discussion
Chile, Costa Rica, Cuba, Hong Kong, Singapore, (South) Korea, Turkey, Croatia, Macedonia, Serbia, Moldova, Kazakhstan, Armenia, Azerbaijan, Georgia

Future plans
+ challenges

Future plans

Expansion:

- Adding countries
- Adding indicators (???)
- Adding data dimensions: duration & parity (???), cohort parity indicators from censuses (???)
- HFC launched (2012)

Other plans

Population exposure: Going beyond HMD

Discussion on data manipulations & smoothing

Communication with users: News section on the web, news email alerts in three regimes: *minor, major, no*

Early data release (?)

…we need your feedback!

Tensions, challenges, discussions, tradeoffs

Data Quality vs. Quantity: Being perfect vs. featuring more data

Number of dimensions & indicators

Data Standardisation, smoothing & manipulation: How far do we want to go?

Population exposure often problematic (HMD responsibility? Using other exposure data? Different estimates of parity distribution)

Website & data: Ease of use, format & access.

Challenges (2): Data quality issues

Data quality issues

Birth & population data appear fairly straightforward, but….

…missing data on age of mother and/or birth order

…missing data on the 'extremes' of age distribution (<15, 50+)

…data in age groups instead of single ages

…year of registration different from the year of birth occurrence (UK)

…missing distinction of live and stillbirths (BG, CZ)

…gaps and errors in birth order registrations (ESP, F, RU)

…unknowns in the census-based parity distributions

…births to native residents living abroad included in the data (BG): double accounting, wrong exposure population

…mismatch between birth-population and exposure population (CH before 2000)

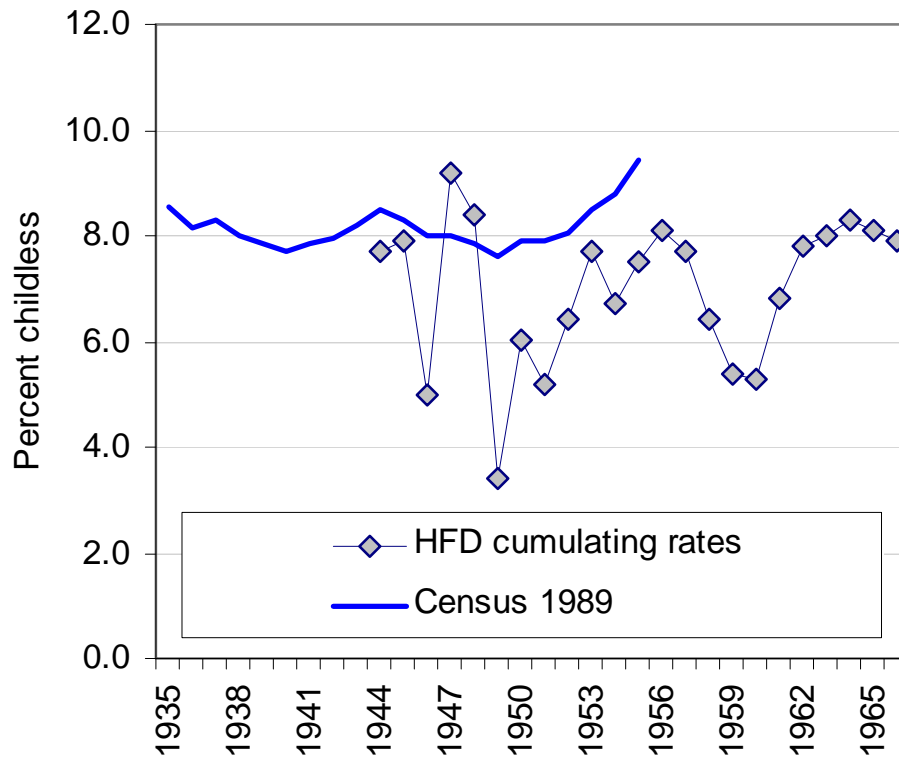…missing info on how the missing data were distributed (CZ)

Data quality issues: Illustration (1)
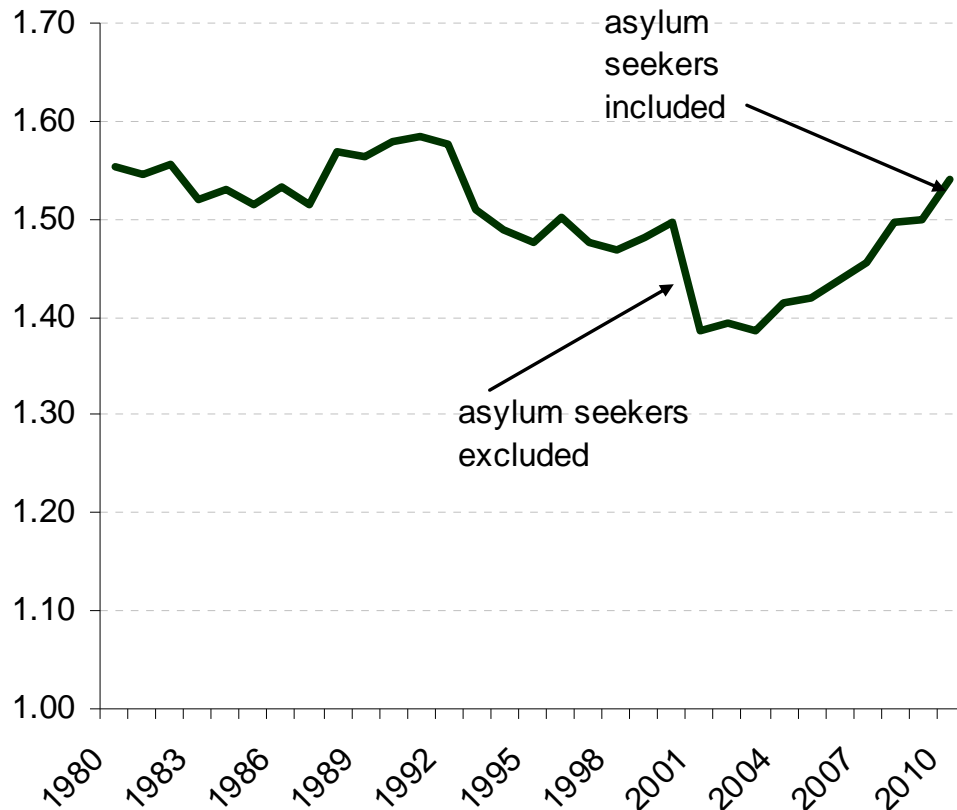
Trend in the mean age at first birth, Lithuania 1980s
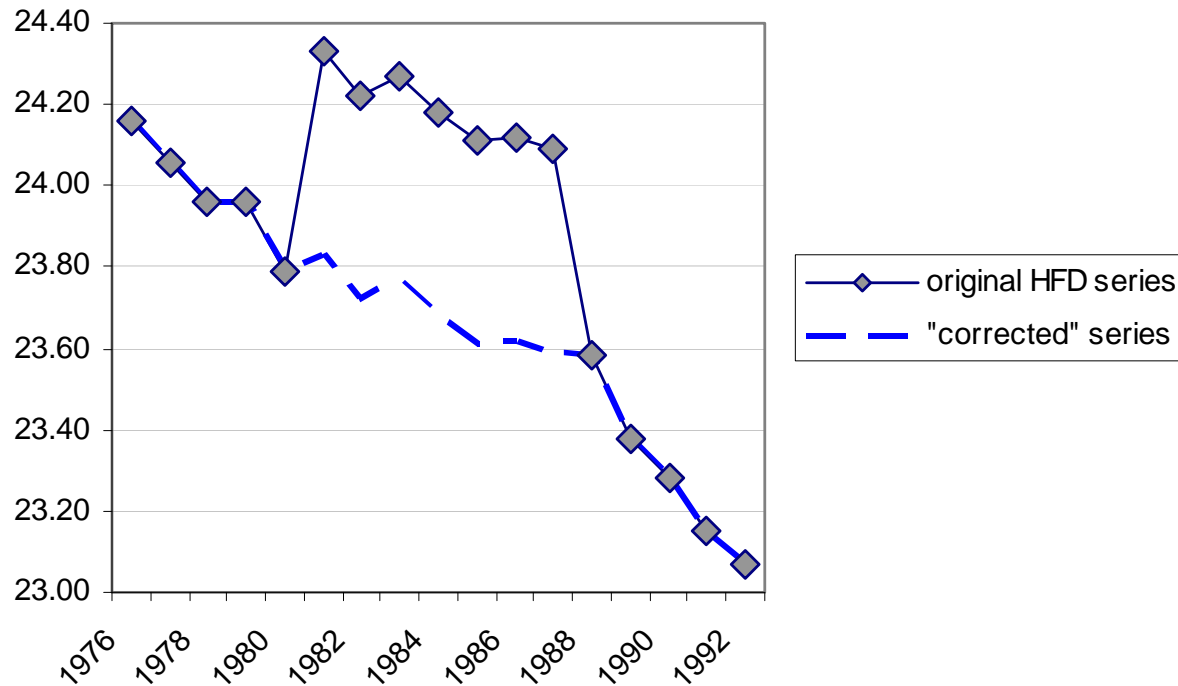
Data quality issues: Illustration (2)

Cohort childlessness in Russia

Data quality issues: Illustration (3)

Period TFR, Switzerland,1980-2010

See country documentation file for Switzerland (Cotter and Zeman 2011, Burkimsher 2011)