COS-R403. Special Research Methods. Forecasting I: Introduction

Lecturer: Christina Bohk-Ewald

Day 1 of intensive 5-day course

University of Helsinki, Finland
04.05.2020–08.05.2020

Brief round of introduction

Who are you? What is your study background? What are you most interested in with respect to demographic forecasting?

Course outline

- Course timeline
- Course content: general purpose
- Course content: intended learning outcomes
- Course content by day
- Organization of (lecture and lab) sessions
- What is expected of you
- Feedback, assessment, and grading
- Course learning material on GitHub

Course timeline by day

We go digital with this course using Zoom and meet twice a day:

Day 1	04.05.2020	10:15–11:45 & 13:15–14:45	Zoom
Day 2	05.05.2020	10:15–11:45 & 13:15–14:45	Zoom
Day 3	06.05.2020	10:15–11:45 & 13:15–14:45	Zoom
Day 4	07.05.2020	10:15–11:45 & 13:15–14:45	Zoom
Day 5	08.05.2020	10:15–11:45 & 13:15–14:45	Zoom

→ Meeting-URLs and passwords have been sent to you by email

Course content in general

This course gives you an **overview of demographic forecasts**, focusing on what they are about, how relevant they are in societies, and first and foremost how they are generated.

You will be introduced to **basic approaches in demographic forecasting** and to a **novel approach for nowcasting COVID-19 infections**.

Analyzing some of these approaches in hands-on exercises with real-world data in the statistical software R will enable you to actively create theoretical and practical knowledge about some crucial issues in demographic forecasting and how to cope with them.

Participating in this course will be very useful for you in terms of **developing skills in demographic forecasting, quantitative data analysis, model implementation, and data visualization**.

Course content in general

Application-oriented basic course on demographic forecasting using R.

You need no prior knowledge in demographic forecasting.

Course content: intended learning outcomes

By the end of the course you will be able

- to list producers and to describe major results of some official demographic forecasts and to reflect upon their societal relevance.
- to know about and download demographic data from open data platforms via the Internet and to load, use, and analyze them in statistical software R.
- to explain basic procedure of demographic forecasting.
- to explain demographic forecast approaches & to implement and apply them (also with real-world data) in statistical software R.
- to present and discuss demographic forecast approaches and their forecast results.

Course content by day

- Days 1 & 2: Introduction to demographic forecasting based on the real-world example *UN World Population Prospects 2019*.
- Days 3 & 4: Introduction to the golden standard in mortality forecasting: the method of *Lee & Carter, 1992*. New directions in mortality forecasting.
- Day 4 & 5: COVID-19 pandemic: exceptional situation, unsolved questions, new data. Analyze trends of COVID-19 pandemic and think about how to make use of forecast tools to tackle pressing issues.

Organization of sessions each day

Sequence of alternating activities including, for example

- Recap of previous material in brief Q & A sessions
- Mini lectures to introduce new topics / content
- Hands-on exercises (also with real-world data) in R
- Discuss emerging issues and solve them together in class
- Present and interactively discuss findings of hands-on exercises

$\rightarrow$ Up to 3 hours per day allow to flexibly organize course sessions and provide sufficient time for self-study

What is expected of you

Active participation in class to deeply understand demographic forecasting.

Feedback, assessment, and grading

...are in alignment with key learning contents and learning activities:

- Formative feedback during course to discover and close learning gaps
- Summative assessment to generate course grade
  - Present and interactively discuss findings of your and other course participants' hands-on exercises. (50%)
  - Report that covers all three core topics: population forecasting, mortality forecasting, and COVID-19 pandemic nowcasting. Report should briefly summarize hands-on exercises, and be 1500-3000 words long without figures, tables, and R-code. You are supposed to write your report during the course, however, it will be due on May 18, 2020. (50%).

Course learning materials

Course learning materials on GitHub:

`https://github.com/christina-bohk-ewald/2020-COS-R403-forecasting-I-introduction`

Contact

Email: christina.bohk-ewald@helsinki.fi

Office: Unioninkatu 35, room 202

Appointments: arrangement by email and personal communication

First day's class:

Introduction to demographic forecasting

- What they are about
- How they can look like in real life
- What they are good for
- How they are generated in general
- What might be potential sources of error and how to account for them

First day's class in the lab: Hands-on exercises in demographic forecasting with R

- Explore UNWPP 2019 and load demographic data from website
- Analyze and visualize the UNWPP data, 1950 through 2100:
  - Global population size
  - Life expectancy at birth
  - Total fertility rate
- Do this again for an assigned world region.
- Present and discuss your findings in class.

What is: Demography & Forecasting

- **Demography** is the science of populations (*demos*) and their measurement (*graphy*).
- **Forecasting** is the process of making statements about likely future development of variable(s) of interest.

What is: Demographic forecasting

- **Demographic forecasts** predict how populations will develop over time in the future
- Population balance equation:

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

Population forecasts	Fertility forecasts	Mortality forecasts	Migration forecasts

$\rightarrow$ Demographic forecasting comprises all these components

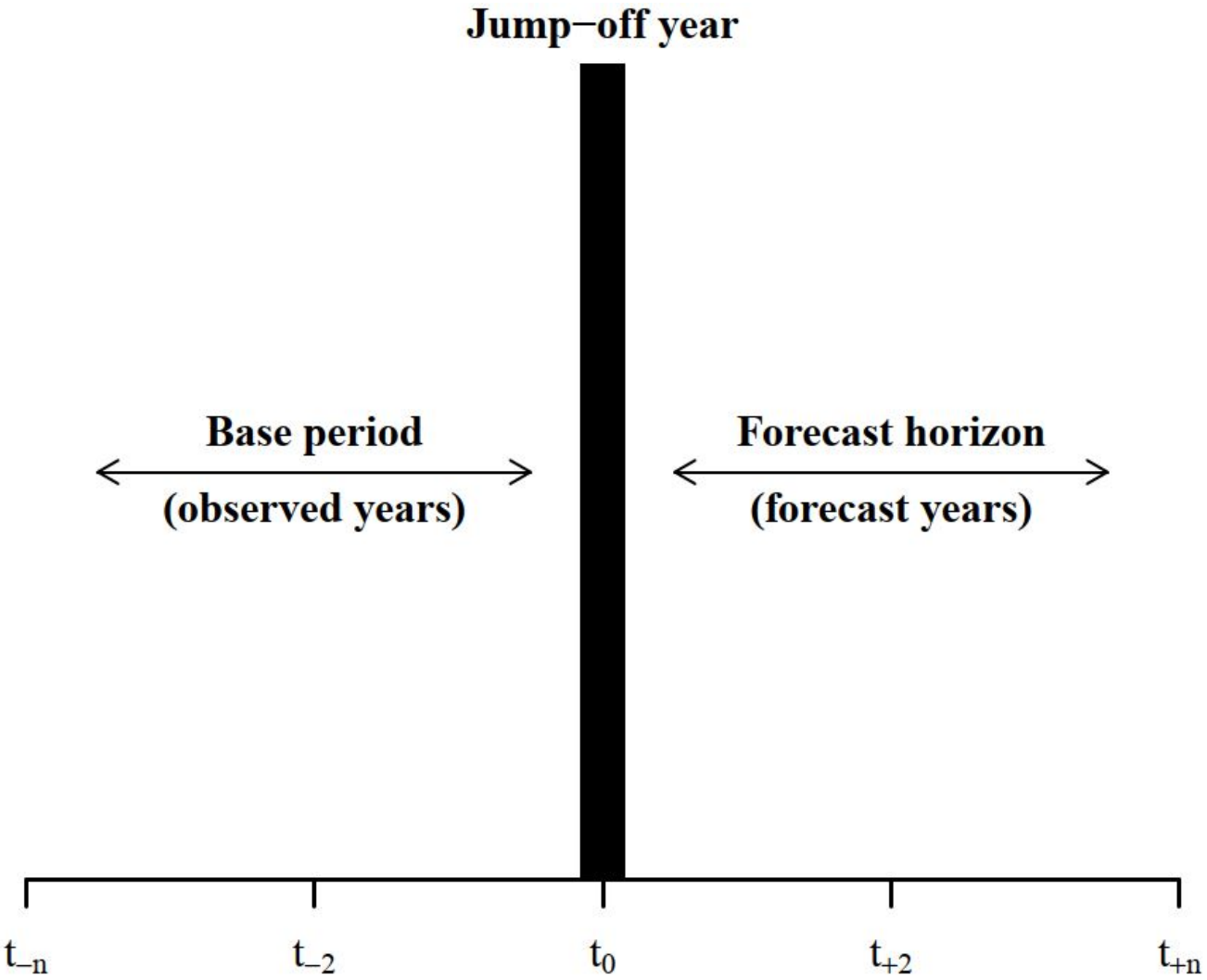
Typical questions

- How long will people live?
- How many of the remaining years of life will people spend in good health, in poor health, or in work and in retirement?
- How many children will people have in 5 and in 50 years from now?
- How many people will live worldwide in upcoming years?
- *What if ...?*

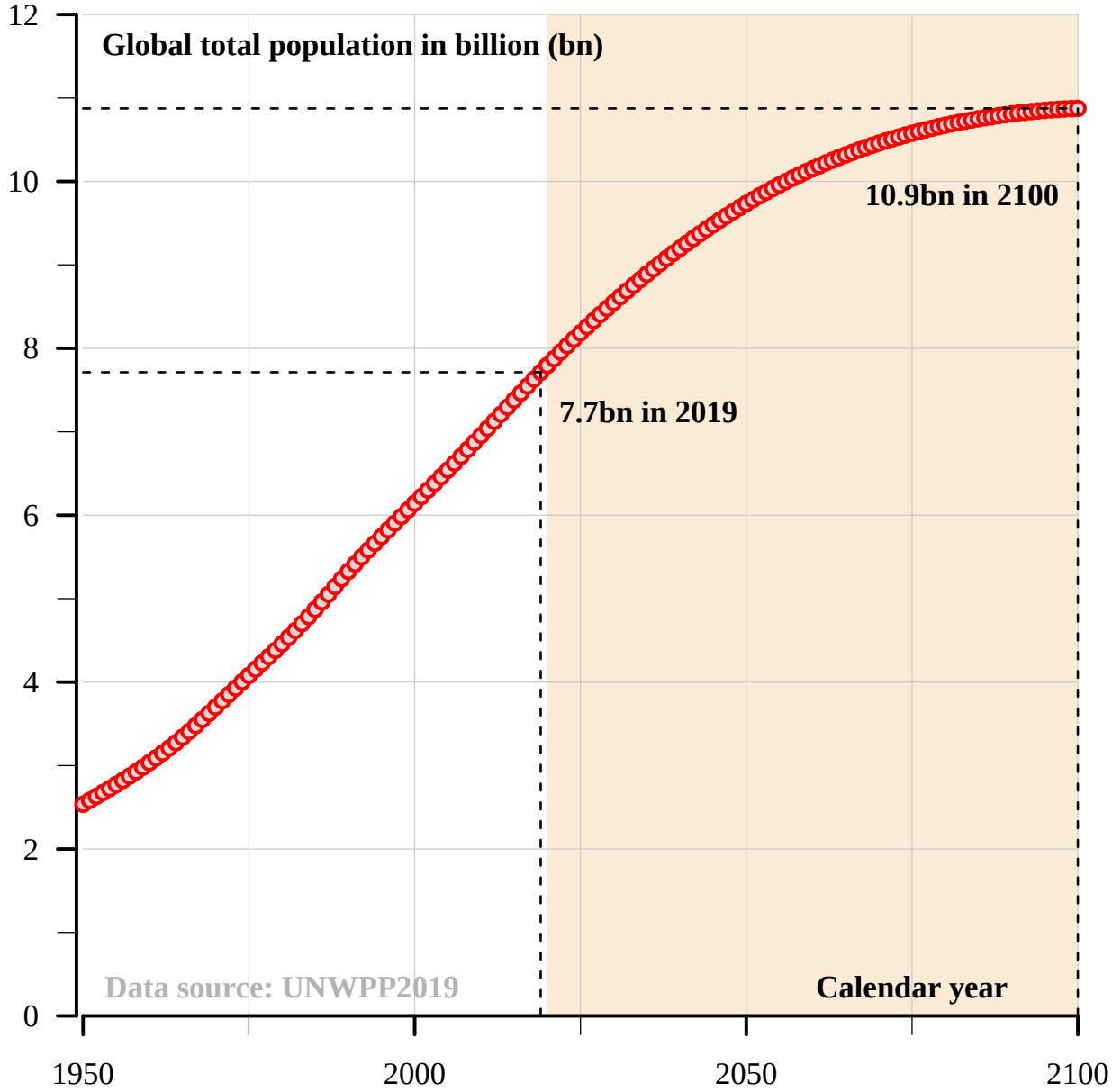
Demographic forecasts in the real world

Example: United Nations World Population Prospects 2019
($\rightarrow$ published only in June last year)

Some terminology first...



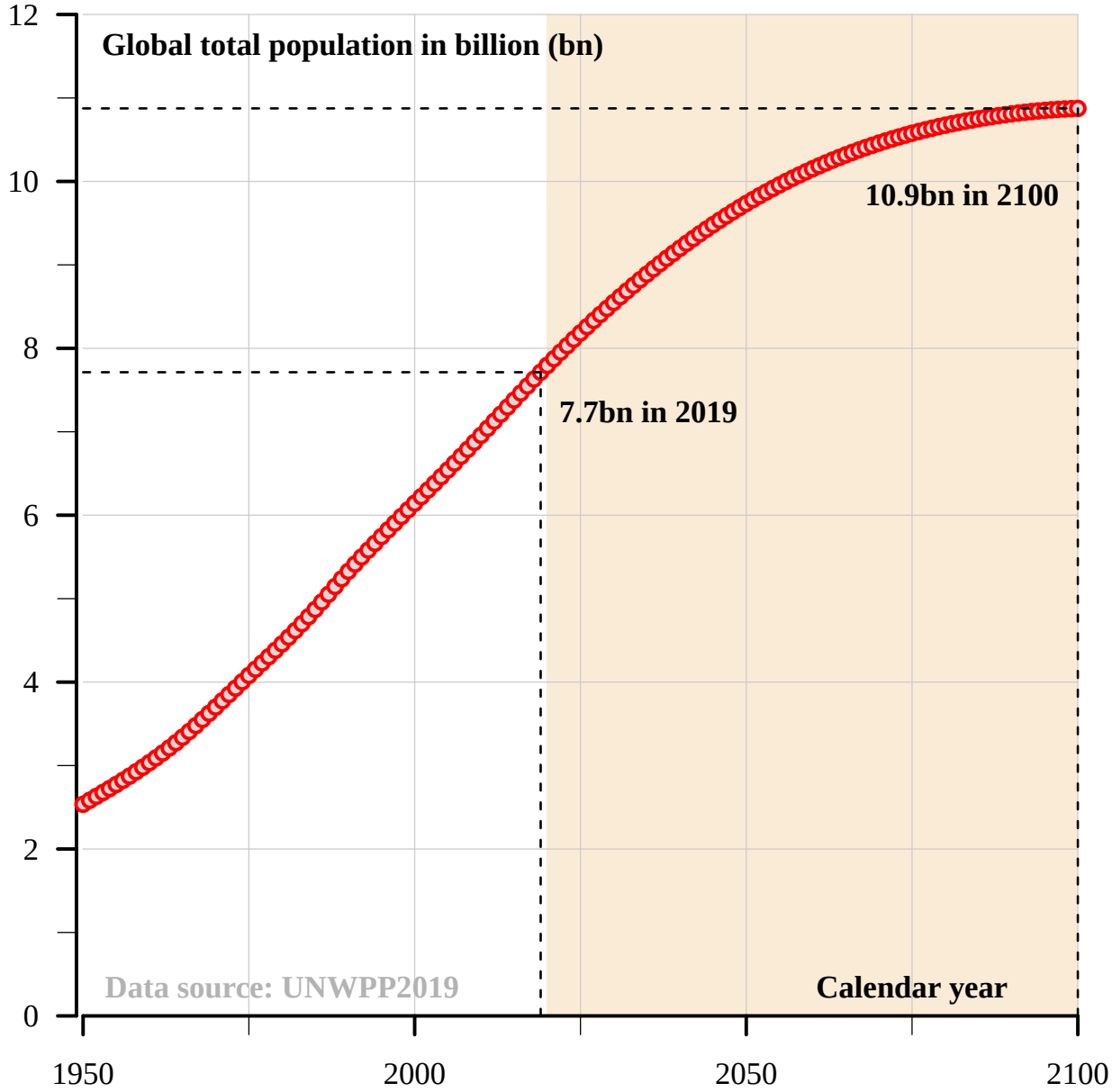
Global population size and growth



[one billion is equal to one thousand million]

1.
2.

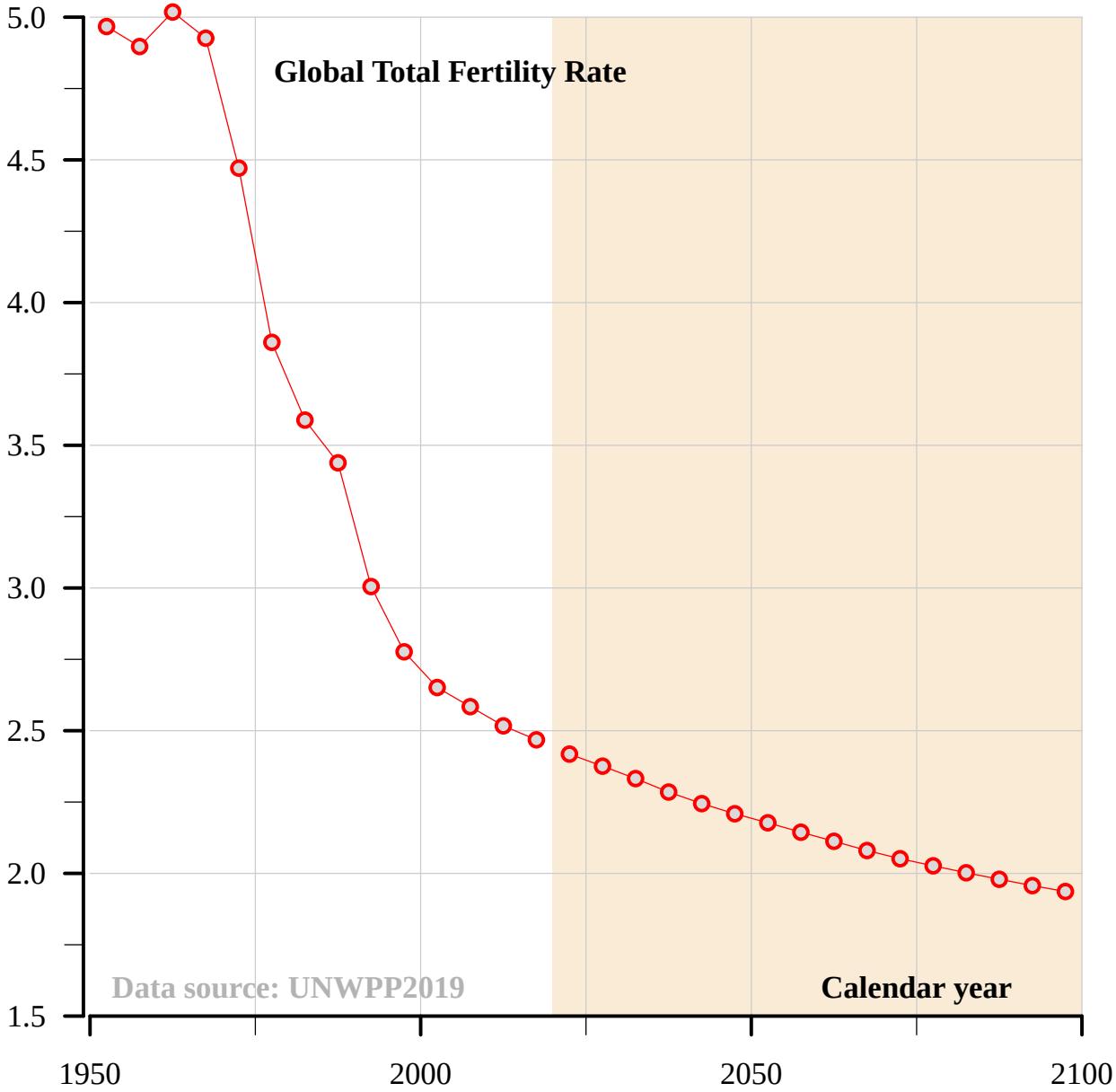
Global population size and growth



[one billion is equal to one thousand million]

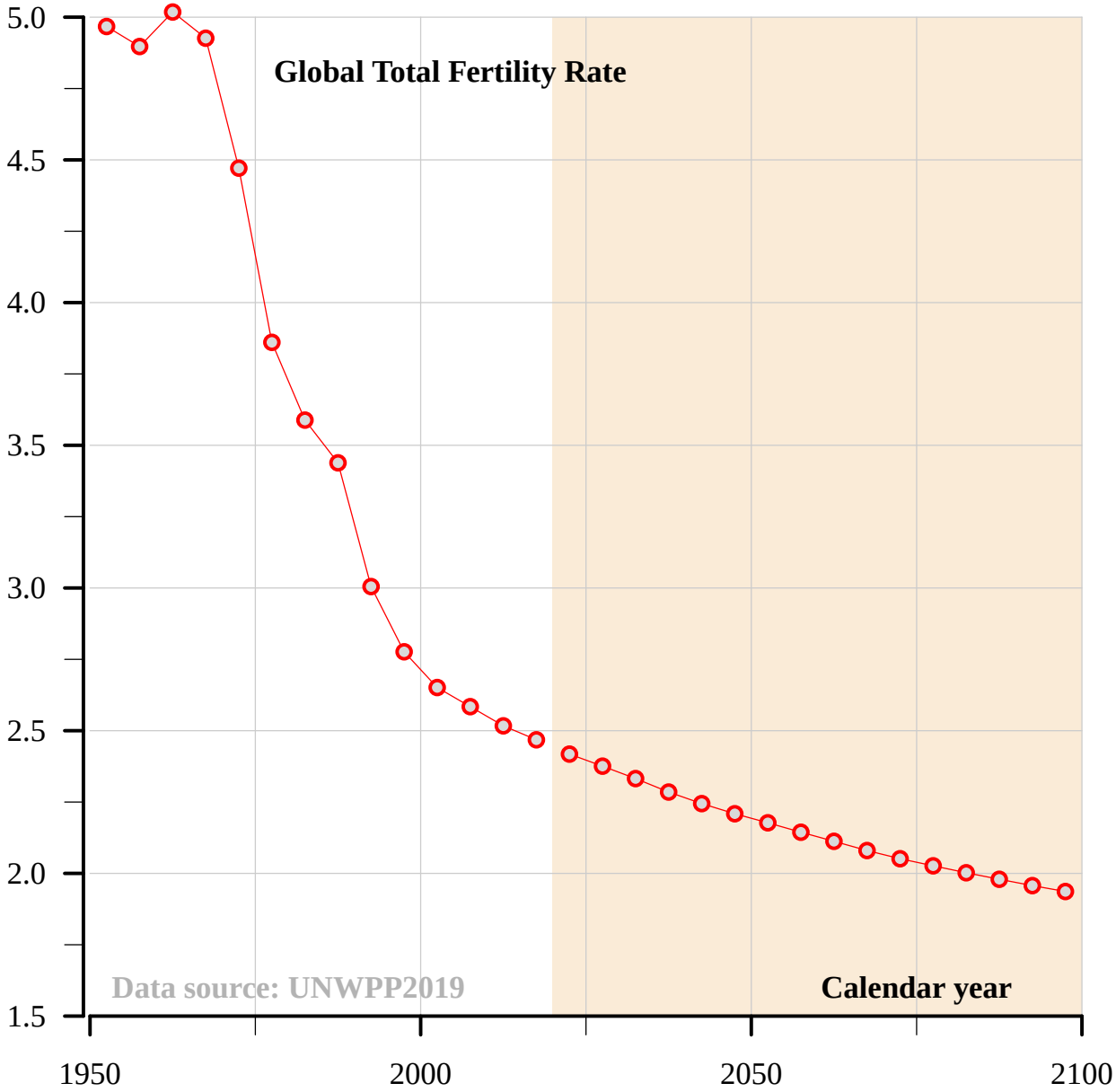
1. Global population size is forecasted to increase
2. ...but less and less with each forecast year

Global Total Fertility Rate



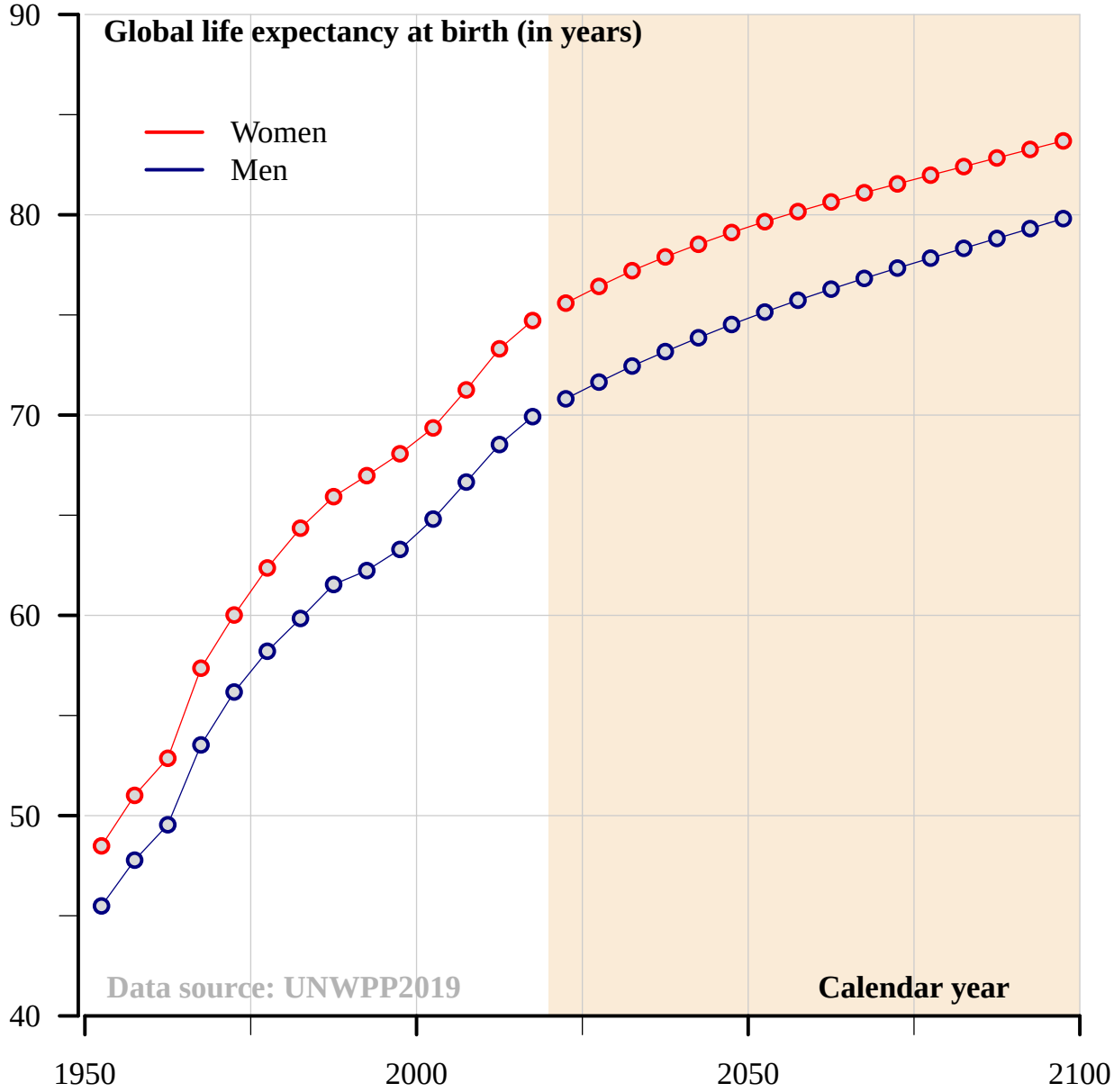
1. How to calculate and interpret TFR?
2.

Global Total Fertility Rate



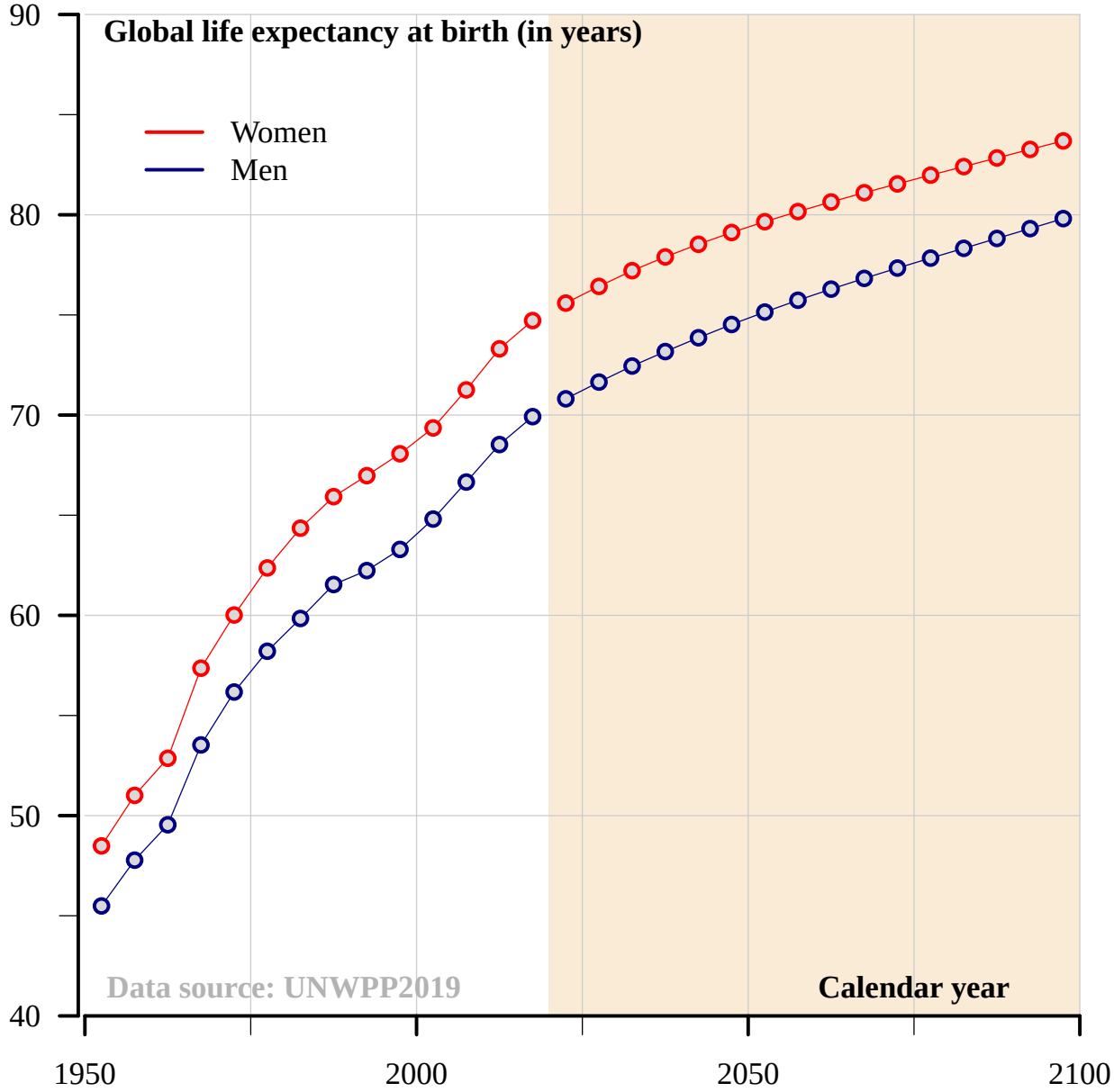
1. How to calculate and interpret TFR?
2. Global TFR is forecasted to decline ...but less and less with each forecast year

Global life expectancy at birth



1. How to calculate and interpret e_0?
2.
3.

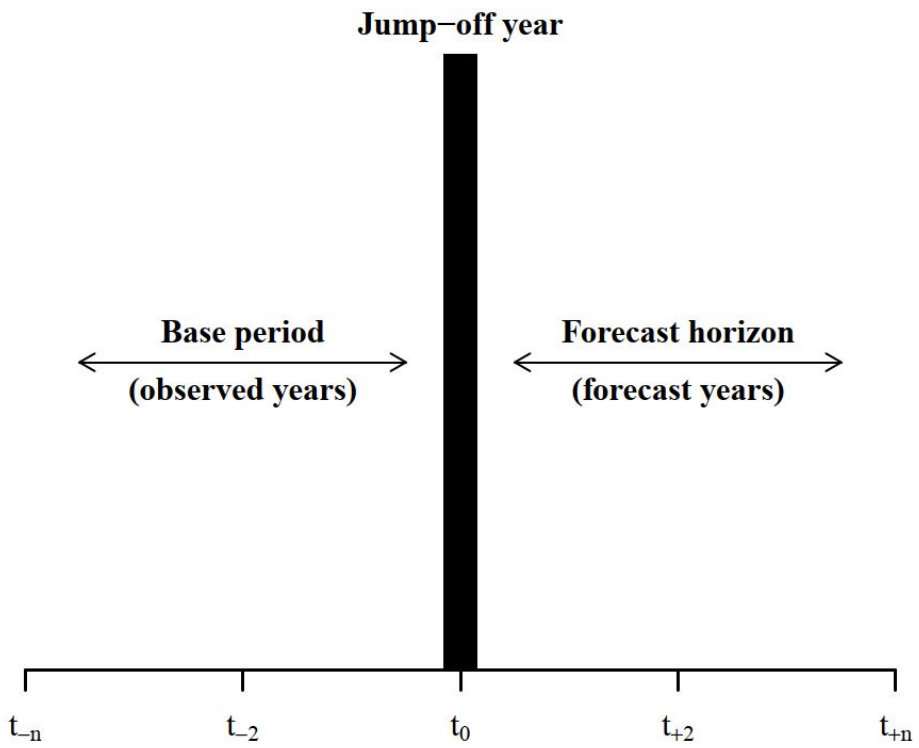
Global life expectancy at birth



1. How to calculate and interpret e_0?
2. Global e_0 is higher for women than for men
3. Global e_0 is forecasted to increase ...but less and less with each forecast year

Some more terminology...

Period versus cohort forecasting



- Lexis diagram

Societal relevance

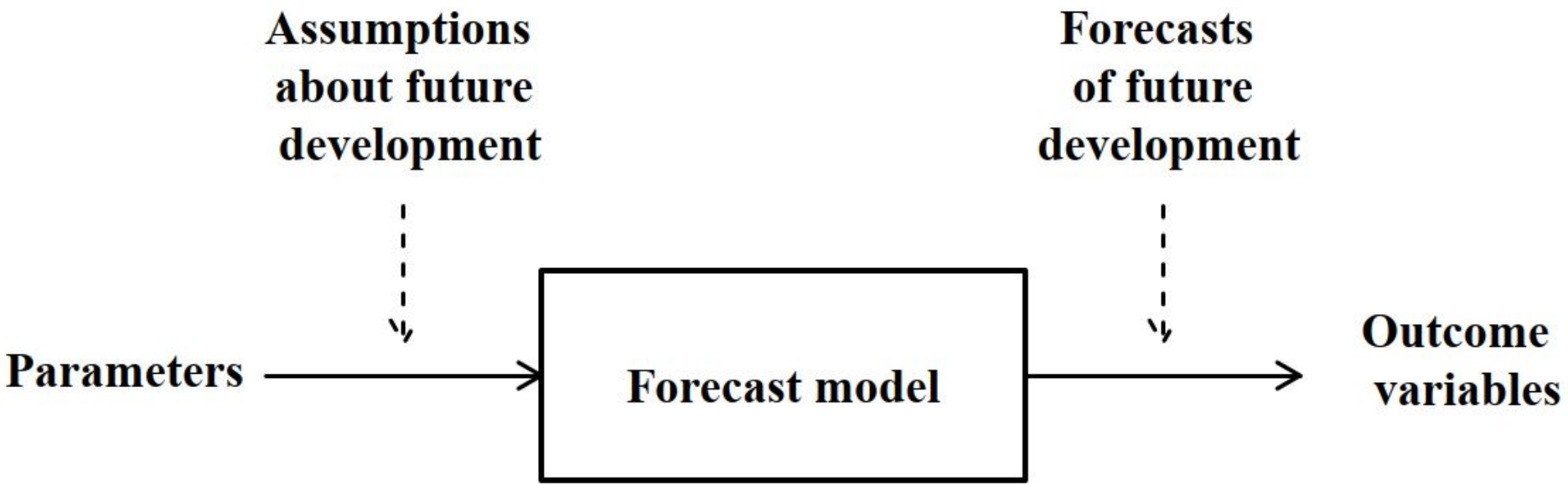
Reliable demographic forecasts lead to informed decisions, policies, and programs affecting e.g. people's social & economic welfare

- They serve as basis for further analysis
  (that predict demand for resources and services in upcoming years)
- Assist in planning, decision making, and creating goals
  in areas such as health care, education, housing,
  energy consumption, transport, retirement planning

$\longrightarrow$ applies to different geographical levels and types of organization

$\longrightarrow$ useful for Finland / Helsinki?

Basic procedure



Basic procedure

Demographic forecasting is a form of prediction
(that you already know from regression analysis / statistics).

Demographic forecast methods often:

- extrapolate past developments into the future
- take time into account
- assume that future is a continuation of the past
- assume that functional relationships between model parameters (expressed in the forecast model) are valid throughout time

$\longrightarrow$ underlying assumptions might be questionable?

Potential sources of error

"Prediction is very difficult, especially about the future."

attributed to Niels Bohr

Forecast errors are deviations between forecasts and their true realizations.

What are potential sources of error?
What could possibly go wrong?

Potential sources of error

- model misspecification (all models are wrong)
- missing / incorrect data (as input for forecast model)
- errors in model implementation
- unexpected events (leading to gross shifts in demographic behavior)
- ...

$\longrightarrow$ contribute to forecast uncertainty

Potential sources of error

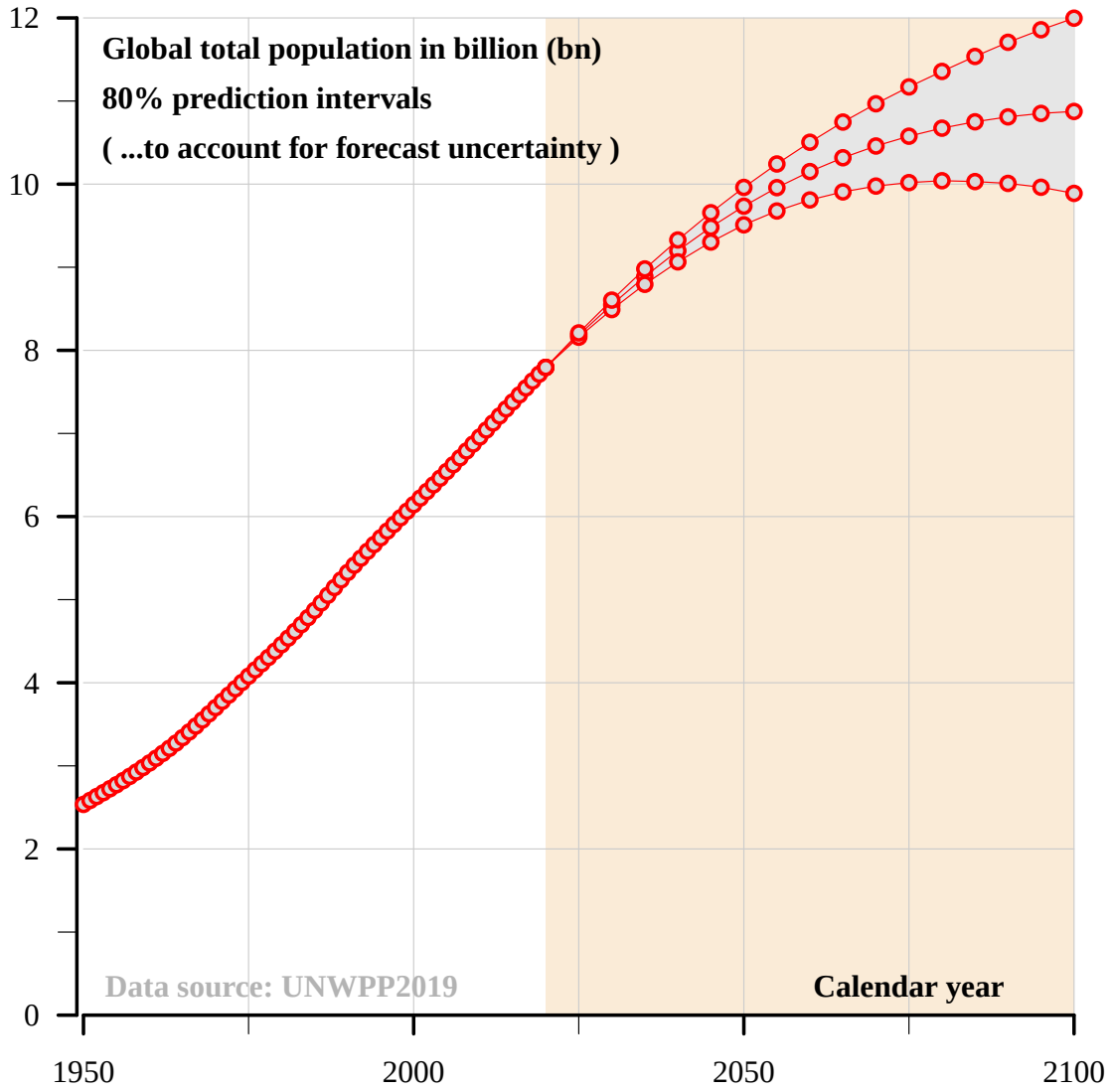
Anecdote about the inductivist chicken:

"The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken."

Bertrand Russell (1912)

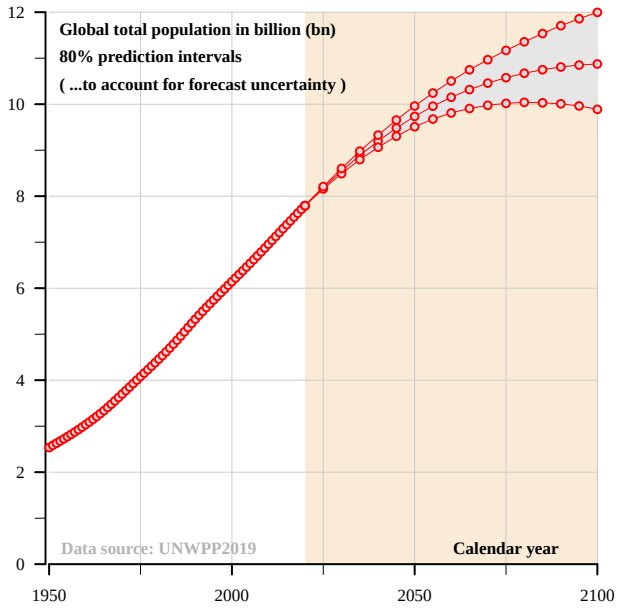
Probabilistic forecasts

...quantify forecast uncertainty

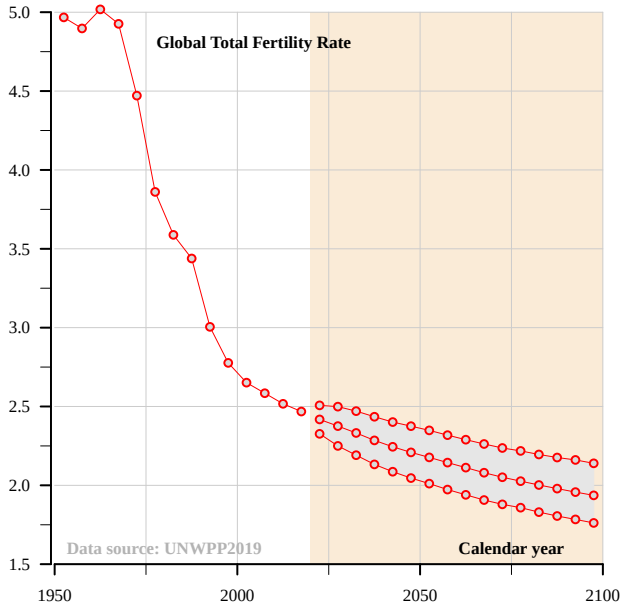


Probabilistic forecasts quantify forecast uncertainty

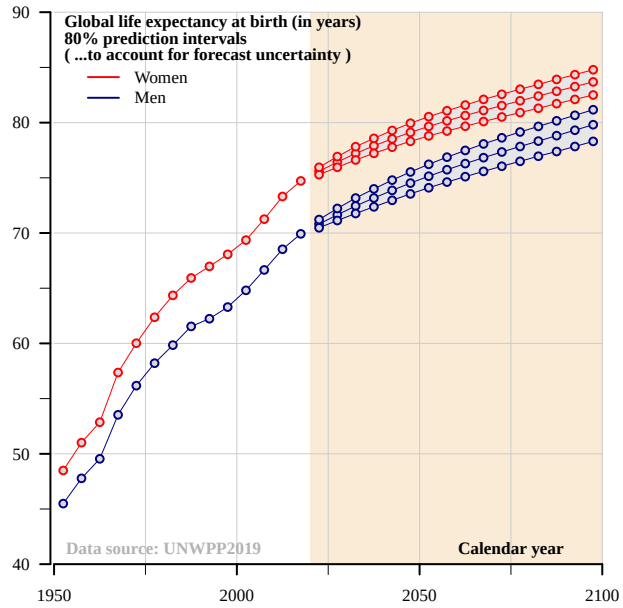
Global population



Total fertility rate



Life expectancy at birth



→ deterministic forecasts versus probabilistic forecasts

What you have learned today about demographic forecasting

- Define demographic forecasting
- List typical questions
- Apply basic terminology of demographic forecasting
- Describe UN forecast of the global population
- Describe societal relevance
- Describe basic procedure how to generate demographic forecasts
- Describe potential sources of error
  and know that they can be accounted for in probabilistic forecasts

First day's class in the lab: Hands-on exercise in demographic forecasting with R

- Explore UNWPP 2019 and load demographic data from website
- Analyze and visualize the UNWPP data, 1950 through 2100:
  - Global population size
  - Life expectancy at birth
  - Total fertility rate
- Do this again for an assigned world region.
- Present and discuss your findings in class.
- What else? Would be a good idea to put this in your report :-)

Course learning materials

Course learning materials on GitHub:

`https://github.com/christina-bohk-ewald/2020-course-COS-R403-forecasting-1-introduction`

R programming

Some functions we will use in hands-on exercises:

- setwd()
- read.xlsx()
- plot(), axis(), legend(), text(), par(fig=, las=, mai=)
- rect(), segments(), lines(), points(), polygon()
- array(), for(){}, as.numeric(), as.character()
- log(, base=exp(1))

→ Get information about what they are and how to use them

Recommended learning material for today's class

- **Christina Bohk (2012)**
  Ein probabilistisches Bevölkerungsprognosemodell. Entwicklung und Anwendung für Deutschland
  Springer VS
  https://link.springer.com/book/10.1007%2F978-3-531-19267-3
- **Rau, R., Bohk-Ewald, C., Muszyńska, M. M. and Vaupel, J. W. (2017)**
  Visualizing mortality dynamics in the Lexis diagram
  The Springer Series on Demographic Methods and Population Analysis, 44. Cham, Springer International Publishing AG.
- **Alho, J. and Spencer, B. (2006)**
  Statistical demography and forecasting
  Springer Science & Business Media.
- **Preston, S., Heuveline, P., and Guillot, M. (2000)**
  Demography. Measuring and modeling population processes
  Blackwell Publishers Ltd.

Recommended learning material for today's class

- **UNWPP2019**: `https://population.un.org/wpp/`
  Publications, Graphs, & Data files.
- **Bohk, C., Ewald, R., and Uhrmacher, A. M. (2009)**
  Probabilistic population projection with JAMES II.
  Proceedings of the 2009 Winter Simulation Conference (WSC), Austin, TX, USA, pp. 2008–2019.
  `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5429715`
- **Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., and Heilig, G. K. (2012)**
  Bayesian probabilistic population projections for all countries.
  Proceedings of the National Academy of Sciences, 109(35), 13915–13921.
- **Alho, J. and Spencer, B. (1997)**
  The practical specification of the expected error of population forecasts. Journal of Official Statistics, 13(3), 203–225.

Thank you for your attention!

christina.bohk-ewald@helsinki.fi

First day's class in the lab: Hands-on exercises in demographic forecasting with R

- Explore UNWPP 2019 and load demographic data from website
- Analyze and visualize the UNWPP data, 1950 through 2100:
  - Global population size
  - Life expectancy at birth
  - Total fertility rate
- Do this again for an assigned world region.
- Present and discuss your findings in class.
- What else? Would be a good idea to put this in your report :-)

Start to prepare your report

- Load demographic data from UNWPP2019 website
- Analyze and visualize them, 1950 through 2100:
  - Population size
  - Life expectancy at birth
  - Total fertility rate
- Briefly describe, visualize, and interpret your findings.

$\rightarrow$ You could use R Markdown (`https://rmarkdown.rstudio.com/`) to document everything—R code, data preparation, method application, and result visualization and interpretation—all in one place.