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

Lecturer: Christina Bohk-Ewald

Day 5 of intensive 5-day course

University of Helsinki, Finland
04.05.2020–08.05.2020

Fifth day's class:

- Recap main concepts of last lecture
- Research in progress: nowcasting COVID-19 infections
- Course summary

Fifth day's class in the lab:
Hands-on exercises on COVID-19 pandemic with R

- Load cases and deaths from COVID-19 from JHU CSSE

  `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`

- Get modal infection fatality rates from Verity et al. (2020, p. 5)
- Get abridged life tables and population counts from UNWPP 2019
- Estimate actual number of infections using approach of Bohk-Ewald, Dudel and Myrskylä (2020) for Italy
  - How many people could have been actually infected with COVID-19?
  - How many more infections could there perhaps be than confirmed cases from COVID-19?
  - How do the daily numbers of infections and of confirmed cases develop over time?
  - ...
- Do the same for Finland. Put this in your report.
- Present and discuss your findings.

Recap main concepts of last lecture

- What are two ways to extend the LC method?
- What is case fatality rate?
- What is infection fatality rate?
- What is a data source for reported deaths and confirmed cases from COVID-19?
- What country has most confirmed cases and repored deaths as of yesterday? $\longrightarrow$ Afternoon lab
- ...

$\Rightarrow$ Questions?

A demographic scaling model for estimating the total number of COVID-19 infections

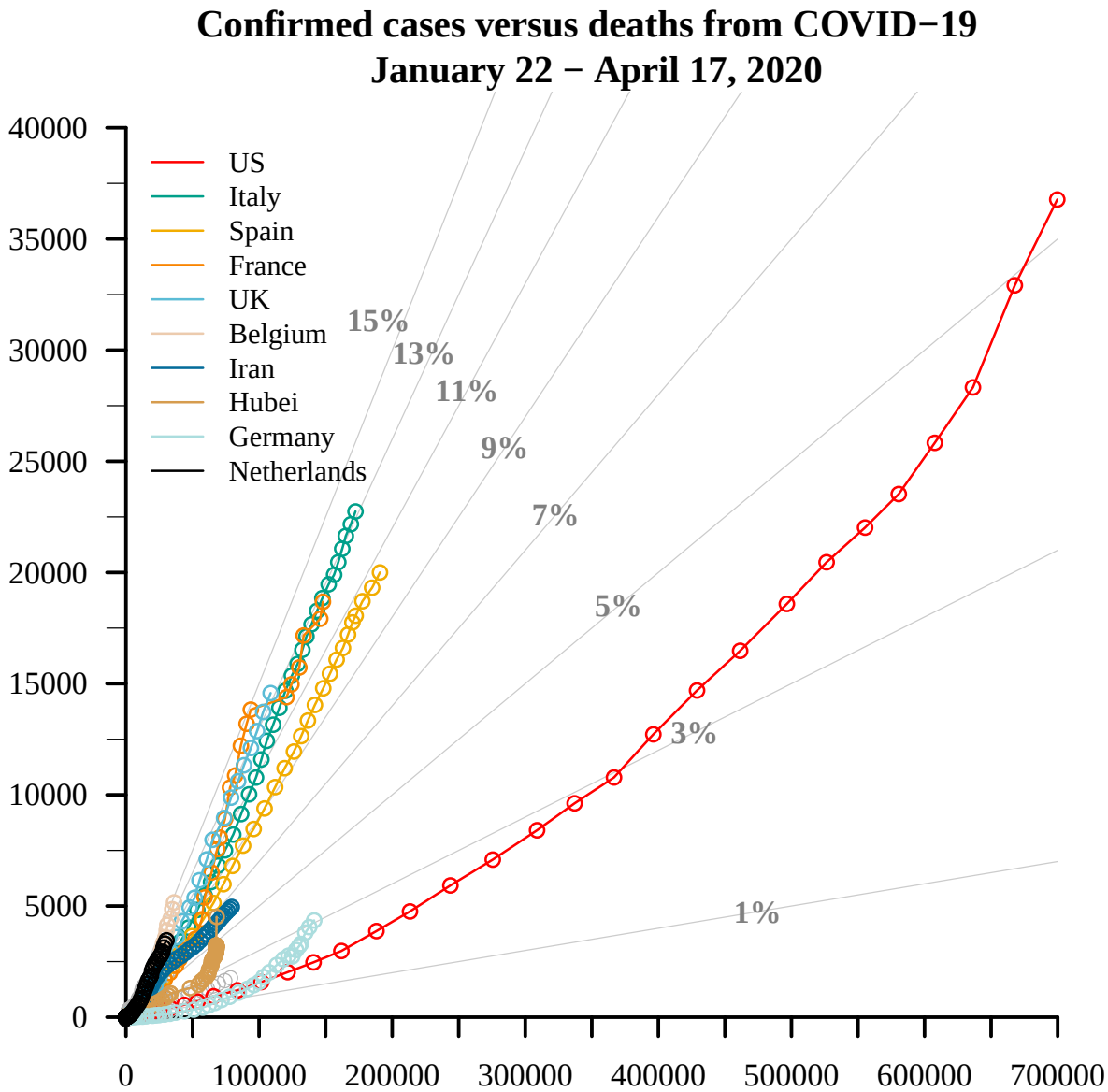
Research in progess by Bohk-Ewald, Dudel, and Myrskylä (2020)

Motivation

- Infections are critical information for decision makers to manage COVID-19 pandemic
- Confirmed cases may underestimate actual numbers of infections
- Develop demographic scaling model to estimate COVID-19 infections in a straightforward way
- This model is designed to (1) be broadly applicable as it uses only few input data and (2) take on a demographic perspective as it accounts for cross-country differences in age structures, health conditions, and health care systems

---

Reported findings are based on data of April 17, 2020, as shown in Bohk-Ewald, Dudel, and Myrskylä (2020)



Source: Bohk-Ewald, Dudel, Myrskylä (2020)

The more confirmed cases, the more deaths from COVID-19?

Empirical trends from Jan 22 to April 17, 2020

- Countries fan out between Belgium and Germany
- The US have largest numbers of confirmed cases and reported deaths as of April 17, 2020
- Case fatality rates (CFR; deaths over cases) vary strongly among countries; between 3% for Germany and 14% for Belgium. Why?

Data source: Johns Hopkin University CSSE

Cross-country variation in CFR may be due to

- Real differences in mortality attributable to COVID-19
- Age structure of cases
- Stage of progress of COVID-19 epidemic
- Classification of COVID-19 deaths
- Testing practices wrt to testing intensity and testing specificty
- Capacity of health-care systems for intensive care
- Effectiveness of control measures to curb severe COVID-19 infections
- ...

$\Rightarrow$ Confirmed cases and reported deaths may be biased

Confirmed cases may underestimate the number of infections

- Cases with mild symptoms or asymptomatic cases might go undetected
- Test coverage
  - Test kits may not be available in large numbers
  - Focus on sub-populations, e.g., cases with proven contact to other COVID-19 cases or hospitalized cases
- False negatives
  - People get tested after the first week of infection, when it is likely that SARS-CoV-2 cannot be detected in pharynx anymore (PCR)
  - Test for antibodies could be be carried out before a body has had a chance to produce them
- ...

Reported deaths may be biased

- Reporting delays may amount to several days.
- Inconsistent practices for classifying COVID-19 deaths within and between countries. For example, only deceased individuals who (1) were hospitalizd or (2) died from COVID-19 as primary and / or secondary cause of death may be counted.
- Test coverage and test specificity may be insufficient. For example, not all deaths are tested for COVID-19. Persons dying at home or in other institutions may not be counted.

$\Rightarrow$ Deaths are likely to be underreported and studies of excess mortality could help to account for this bias

$\Rightarrow$ We argue that deaths are probably more reliable than cases and take them as core empirical input for our model

Scaling model to estimate COVID-19 infections

- $I_x = P_x \cdot \lambda_x$ (1)

  The number of infections I is a fraction λ of the total population P in each age group x. P is known, I and λ are unknown.

- Knowing that $IFR_x = \frac{D_x}{I_x}$ we estimate the unknown population fraction of infections λ with the known number of deaths D, scaled infection fatality rates IFR, and population counts P in each age group x

  $D_x = IFR_x \cdot P_x \cdot \lambda_x$ (2)

  Rearranging equation (2) yields:

- $\lambda_x = \frac{D_x}{IFR_x \cdot P_x}$ (3)

Scaling model to estimate COVID-19 infections, ctnd

- We finally estimate the total number of infections as the sum of population counts P and the population fraction λ over all ages x:

  $$I = \sum_x P_x \cdot \lambda_x \quad (4)$$

  Inserting the definition of λ of equation 3 yields:

- $$I = \sum_x \frac{D_x}{IFR_x} \quad (5)$$

Scaling model to estimate COVID-19 infections, ctnd

How to get credible estimates of IFR_x.

- We borrow IFRs from a reference country and scale them to control for differences in the age structure, health conditions, and medical care between a country of interest and a reference country
- Scaling is based on life years left (thanatological age):

$$IFR_{e_x}^{COI} = IFR_{e_x}^{RC}$$

COI denotes country of interest, RC denotes reference country, and e_x is remaining lifetime.

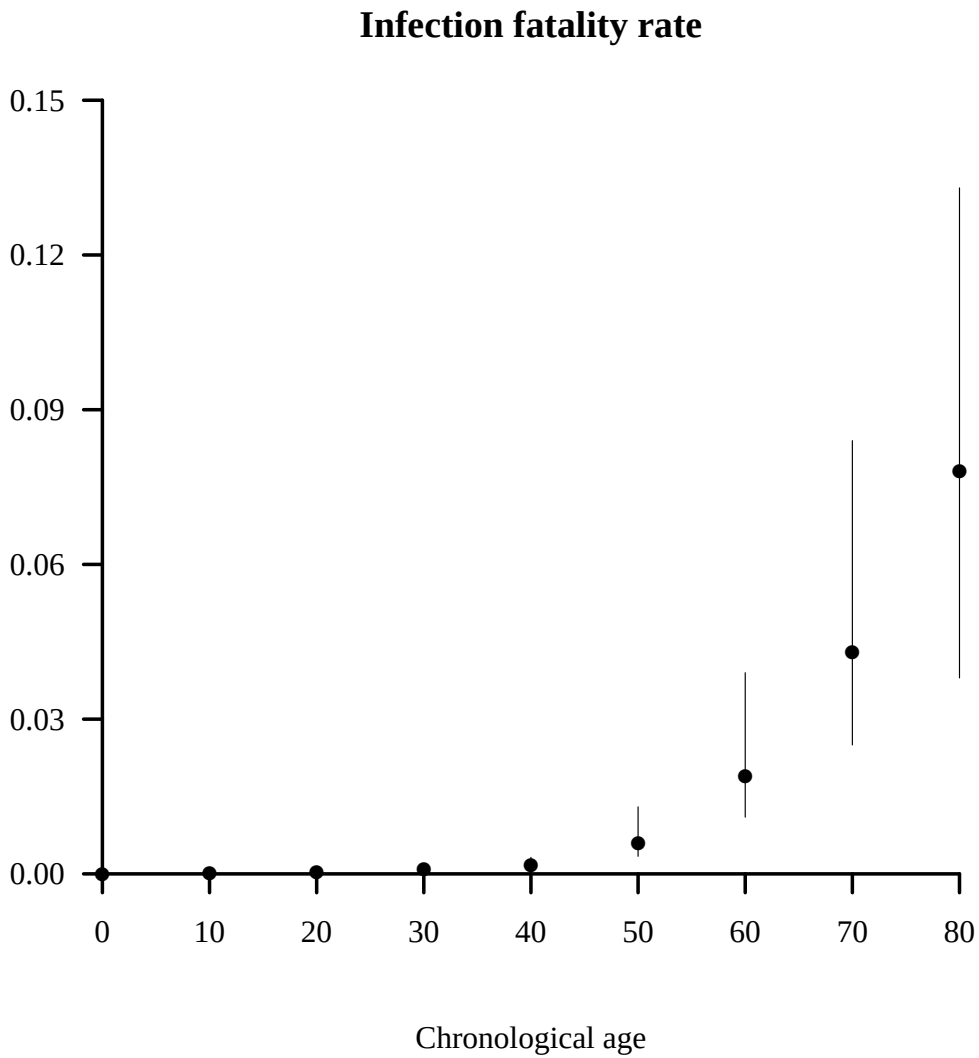
- e_x can be regarded as a marker for health conditions and health care system's capability to treat them
- e_x is a life table parameter and life tables are available for most countries

Apply scaling model to estimate COVID-19 infections

How dos this look like in practice? We take

- IFRs for Hubei, China, as reported in Verity et al. (2020)
- Deaths from COVID-19 of JHU CSSE (2020)
- Population counts and life tables from UNWPP (2019)

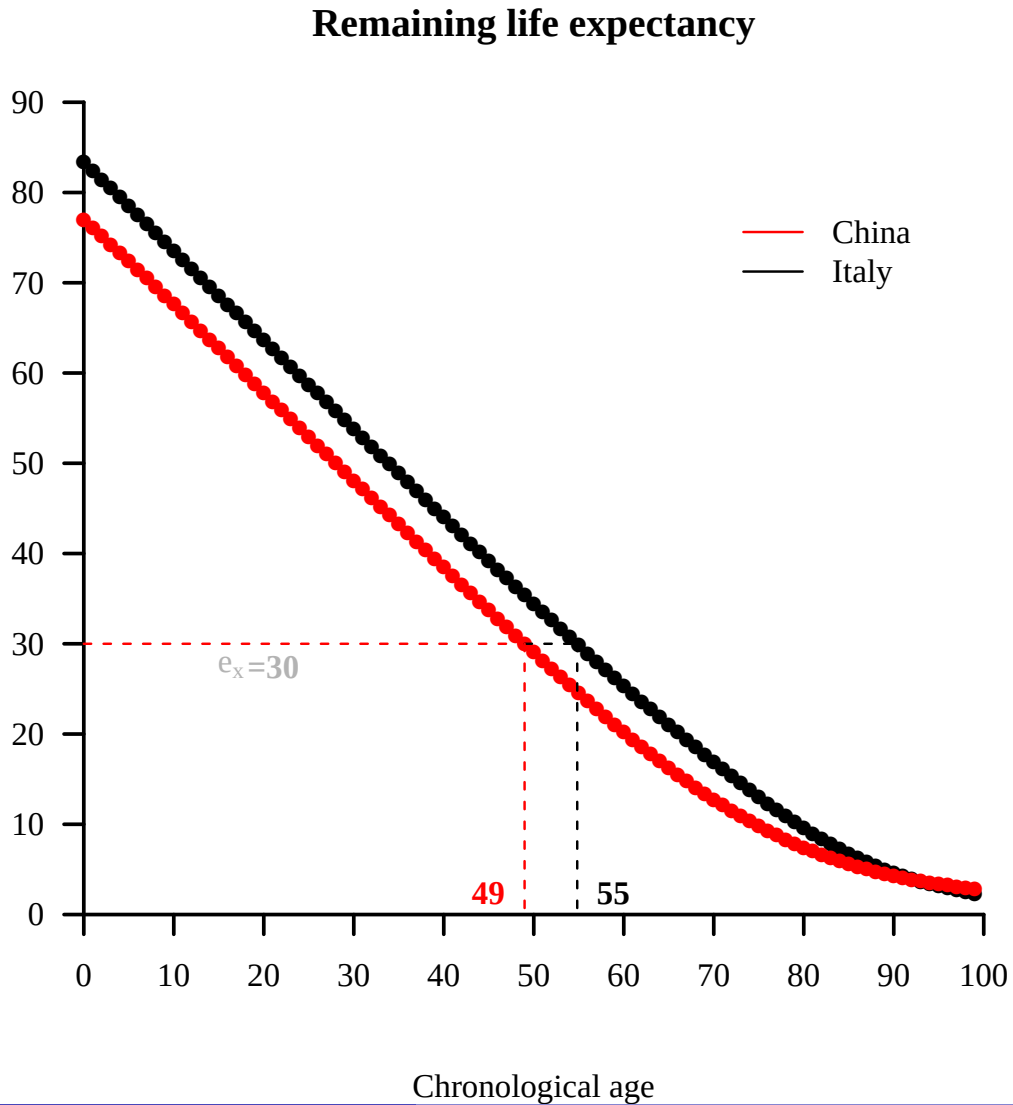
Scale infection fatality rates



Data source: Verity et al. (2020)

Scale infection fatality rates

We assign the same infection fatality rate (IFR) to people of two countries who have, on average, the same number of life years left e_x

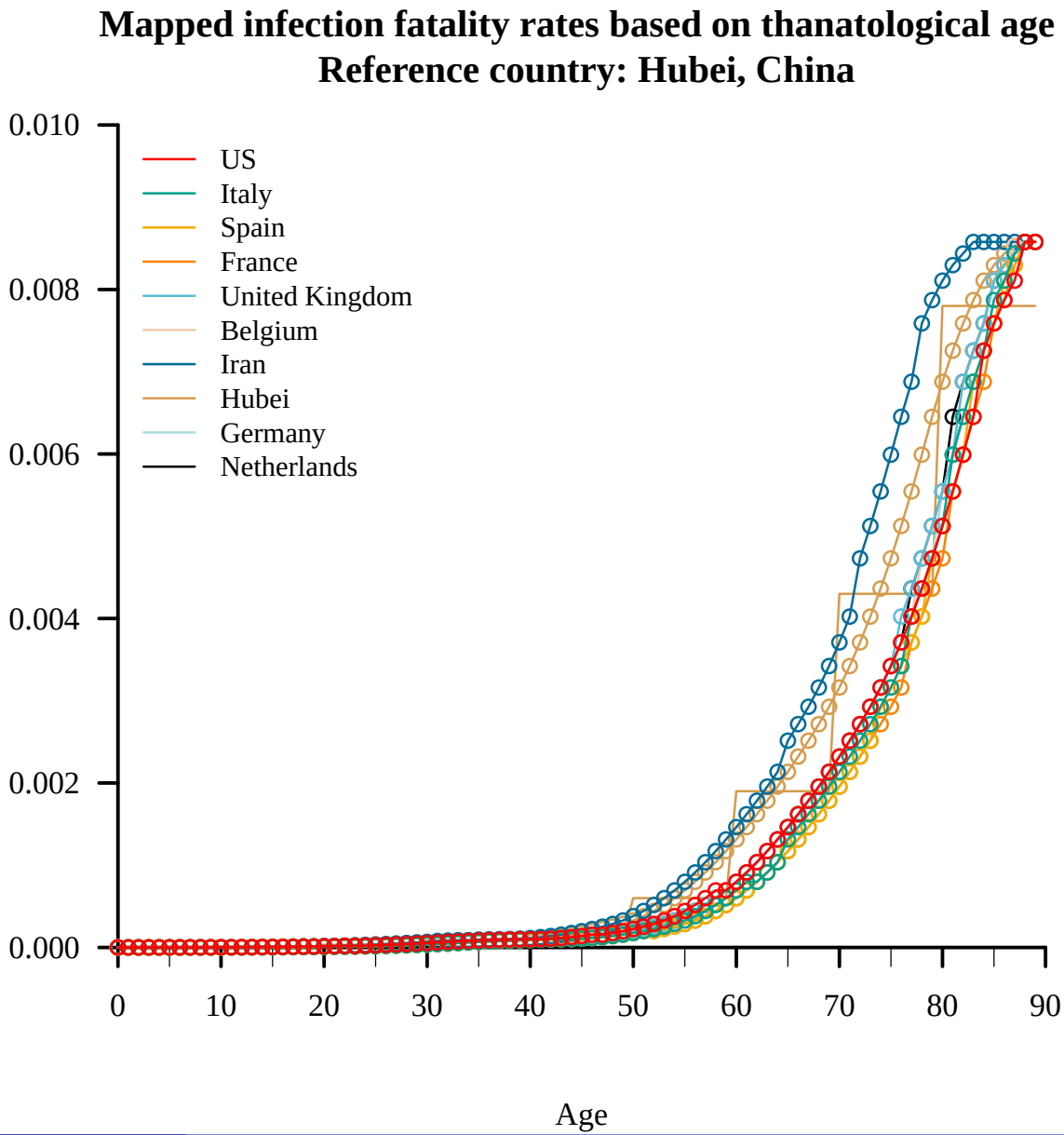


- $IFR^{China}_{e_x=30} = IFR^{Italy}_{e_x=30}$

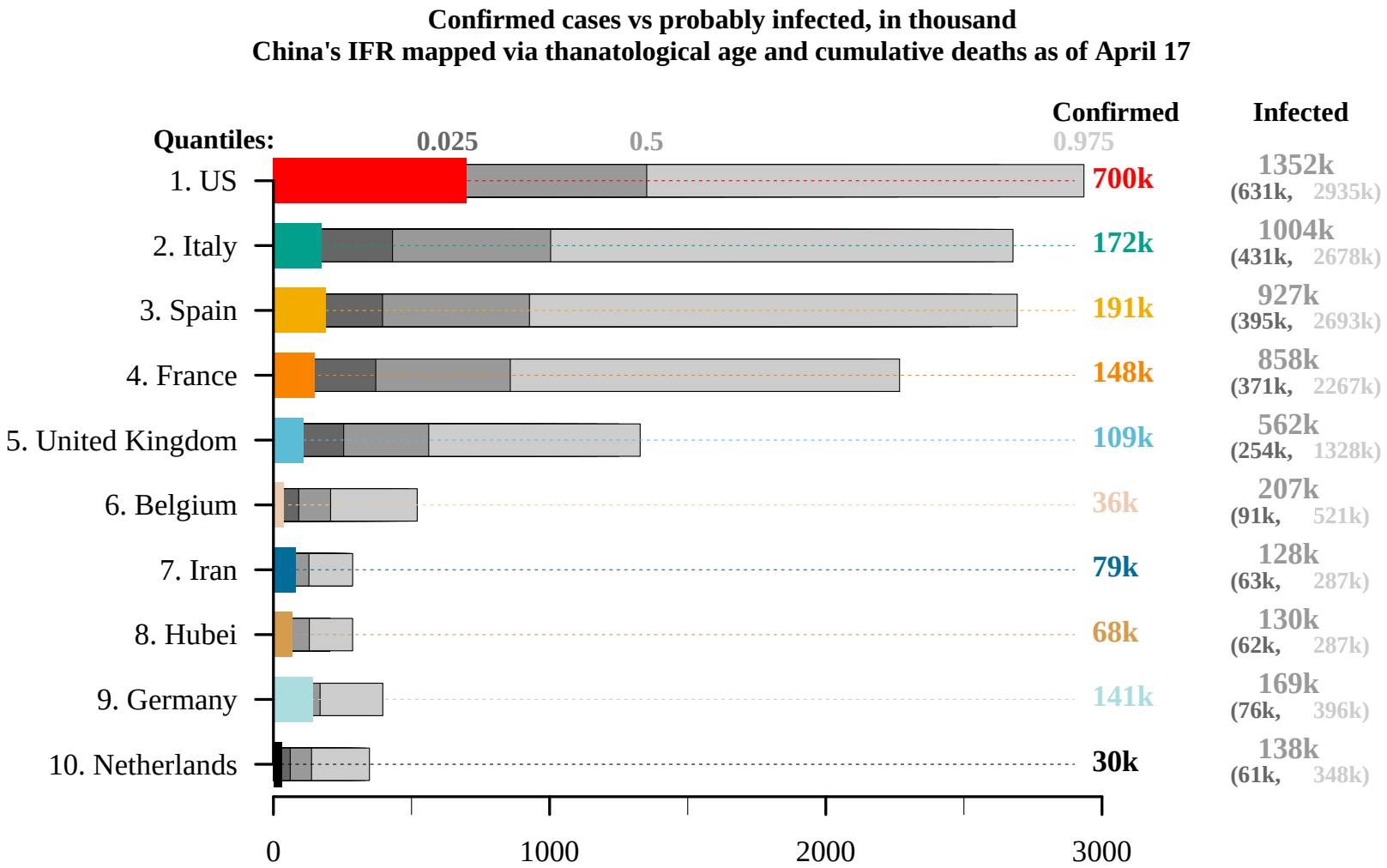
$\Downarrow$

- $IFR^{China}_{x=49} = IFR^{Italy}_{x=55}$
- Data source: UNWPP (2019)

Scaled infection fatality rates



Results: total number of COVID-19 infections



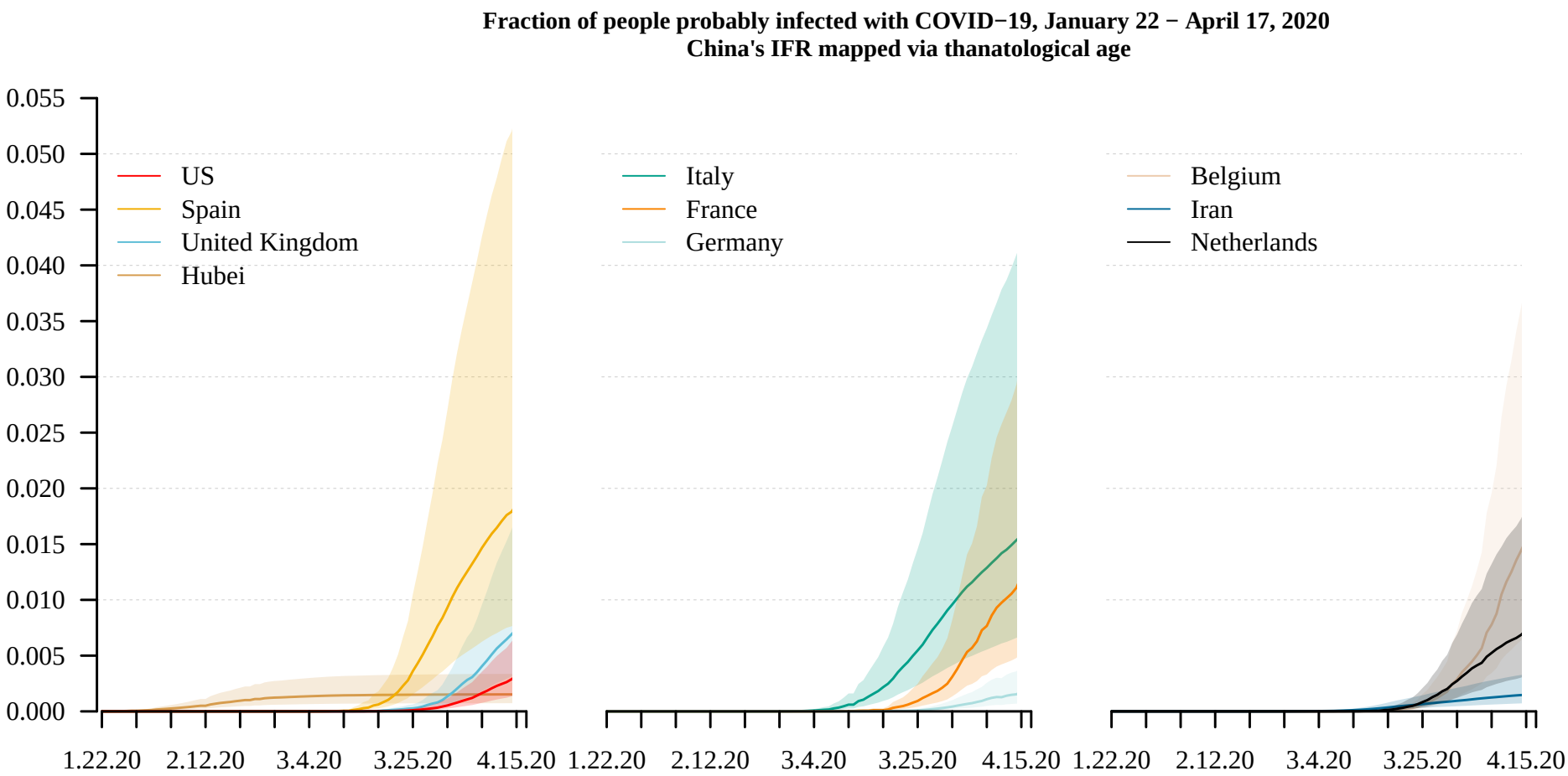
Source: Bohk-Ewald, Dudel, Myrskylä (2020)

Results: total number of COVID-19 infections

Across the 10 countries with most COVID-19 deaths as of April 17, 2020

- The total number of infected individuals is approximately four times higher than the number of confirmed cases
- Uncertainty of findings is high. 95% prediction interval: 2 to 10 times as many infections as confirmed cases
- Variation across countries is also high.

Results: population fraction of COVID-19 infections



Source: Bohk-Ewald, Dudel, Myrskylä (2020)

Results: population fraction of COVID-19 infections

Across the 10 countries with most COVID-19 deaths as of April 17, 2020

- Population fraction of COVID-19 infections increases with time
- Variation across countries is high
- Central estimates range between 2% in Spain and less than 0.4% in Germany
- Uncertainty of findings is large

Model assumptions

Demographic scaling model is built on two assumptions that at best only partially hold

1. Deaths from COVID-19 are fairly accurately recorded
   - Deaths might be biased due to e.g. reporting delays, inconsistent classification of COVID-19 deaths, and poor test coverage
   - If the numbers of total deaths were too small, the estimated numbers of infections would be biased downward
   - If amount of bias due to misreporting was known, this model could easily take this information into account

2. IFR borrowed from a reference country reflects the true IFR of the target country after appropriate scaling
   - Control measures are similarly in RC and COI, after controlling for differences in age structure, health conditions, and health care systems
   - $\rightarrow$ this can only partially reflect country-specific trends

Concluding remarks

- The total number of infections is unknown in countries worldwide
- Demographic scaling model is a simple and fast approach to estimate the total number of infections
- This is particularly important in situations in which detailed data needed for precise estimation are lacking
- We find that numbers of infections are on average four times larger (upper bound: 10 times larger) than confirmed cases
- Model outcome can be used in decision-making and as input in more complex (epidemiological) models

$\rightarrow$ Modeling an estimating mortality (COVID-19 related deaths) and population (fraction that is infected) are core topics in demographic forecasting

What you have learned today

- Describe demographic scaling model for estimating COVID-19 infections (research in progress)
- Use it as an example on how to use and adjust demographic forecast tools to address urgent questions of the COVID-19 pandemic

Fifth day's class in the lab:
Hands-on exercises on COVID-19 pandemic with R

- Load cases and deaths from COVID-19 from JHU CSSE

  `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`

- Get modal infection fatality rates from Verity et al. (2020, p. 5)
- Get abridged life tables and population counts from UNWPP 2019
- Estimate actual number of infections using approach of Bohk-Ewald, Dudel and Myrskylä (2020) for Italy
  - How many people could have been actually infected with COVID-19?
  - How many more infections could there perhaps be than confirmed cases from COVID-19?
  - How do the daily numbers of infections and of confirmed cases develop over time?
  - ...
- Do the same for Finland. Put this in your report.
- Present and discuss your findings.

What you have learned in this course, ctnd

Gained insight into core topics in demographic forecasting

1. Population forecasting using the real-world example of the *UN World Population Prospects 2019*.
2. Mortality forecasting using the method of *Lee & Carter, 1992*, and looking at new directions in this field.
3. Reflecting upon and making use of forecast tools to analyze trends of the COVID-19 pandemic and to model and nowcast COVID-19 infections with research in progress.

What you have learned in this course, ctnd

- 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.

Your final report

Report should cover all three core topics: population forecasting, mortality forecasting, and COVID-19 pandemic nowcasting.

It should briefly summarize hands-on exercises, and be 1500-3000 words long, without figures, tables, references, and R-code.

It will be due on May 18, 2020.

You can send it to me at: `christina.bohk-ewald@helsinki.fi`

Course learning materials

Course learning materials on GitHub:

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

Recommended learning material for today's class

- Johns Hopkins University CSSE. Novel Coronavirus (COVID-19) Cases Data. Published online: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases. Download on April 17, 2020.
- United Nations, Department of Economic and Social Affairs, Population Division. (2019). Population Prospects 2019. Published online: https://population.un.org/wpp/. Download on April 17, 2020.
- Roser M, Ritchie H, Ortiz-Ospina E (2020). Coronavirus Disease (COVID-19) – Statistics and Research. Our World in Data 2020; published online. https://ourworldindata.org/coronavirus.

Recommended learning material for today's class

- Bohk-Ewald C, Dudel C, Myrskylä M. A demographic scaling model for estimating the total number of COVID-19 infections. medRxiv.org 2020; published online April 29. DOI:https://doi.org/10.1101/2020.04.23.20077719.
- Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis 2020; published online March 30. DOI:10.1016/S1473-3099(20)30243-7.
- Epidemiology Group of the New Coronavirus Pneumonia Emergency Response Mechanism of the Chinese Center for Disease Control and Prevention. Epidemiological characteristics of the new coronavirus pneumonia. Chinese Journal of Epidemiology 2020; 41: published online February 17. DOI: 10.3760 / cma.j.issn.0254-6450.2020.02.003.

Thank you for your attention!

christina.bohk-ewald@helsinki.fi

Fifth day's class in the lab:
Hands-on exercises on COVID-19 pandemic with R

- Load cases and deaths from COVID-19 from JHU CSSE

  `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`

- Get modal infection fatality rates from Verity et al. (2020, p. 5)
- Get abridged life tables and population counts from UNWPP 2019
- Estimate actual number of infections using approach of Bohk-Ewald, Dudel and Myrskylä (2020) for Italy
  - How many people could have been actually infected with COVID-19?
  - How many more infections could there perhaps be than confirmed cases from COVID-19?
  - How do the daily numbers of infections and of confirmed cases develop over time?
  - ...
- Do the same for Finland. Put this in your report.
- Present and discuss your findings.