COS-D407. Scientific Modeling and Model Validation

Lecturer: Christina Bohk-Ewald

Week 5

University of Helsinki, Finland
26.10.2020–09.12.2020

Fifth week's class:

Scientific modeling & model validation in practice

- Q&A: recap of material of previous session
- Present your findings of previous lab session
- Validity & robustness of the demographic scaling model's COVID-19 infection estimates, continued

Fifth week's class in the lab: Sensitivity of demographic scaling model's results.

- Test the sensitivity of the demographic scaling model's results for Finland with respect to death counts from two different sources:
  - Total death counts caused by COVID-19 from the Johns Hopkins University CSSE & global age pattern.
  - Age-specific death counts caused by COVID-19 from the COVerAGE-DB database.
  - Critically think about the plausibility of these D_x estimates and their implications for the results of the demographic scaling model.
- Systematically analyze the combined impact of various IFR_x and D_x estimates on the results of the demographic scaling model.

$\longrightarrow$ Present and discuss your findings in class at the beginning of the next session on Monday.

Brief Q&A: recap material of previous session

- What different sources of IFR estimates do you know of?
- How does the age profile of COVID-19-related infection fatality rates look like?
- How do these various IFR estimates differ?
- How sensitive are the COVID-19 infection estimates in Finland with respect to IFR estimates?

$\longrightarrow$ Open questions?

Brief Q&A: recap material of previous session

Central questions for validating a model:

- How do findings come to be and what factors might influence them in what way?
- In what situations does a model work, in what situations does a model not work?
- How sensitive are a model's findings with respect to (changes in) input parameters?

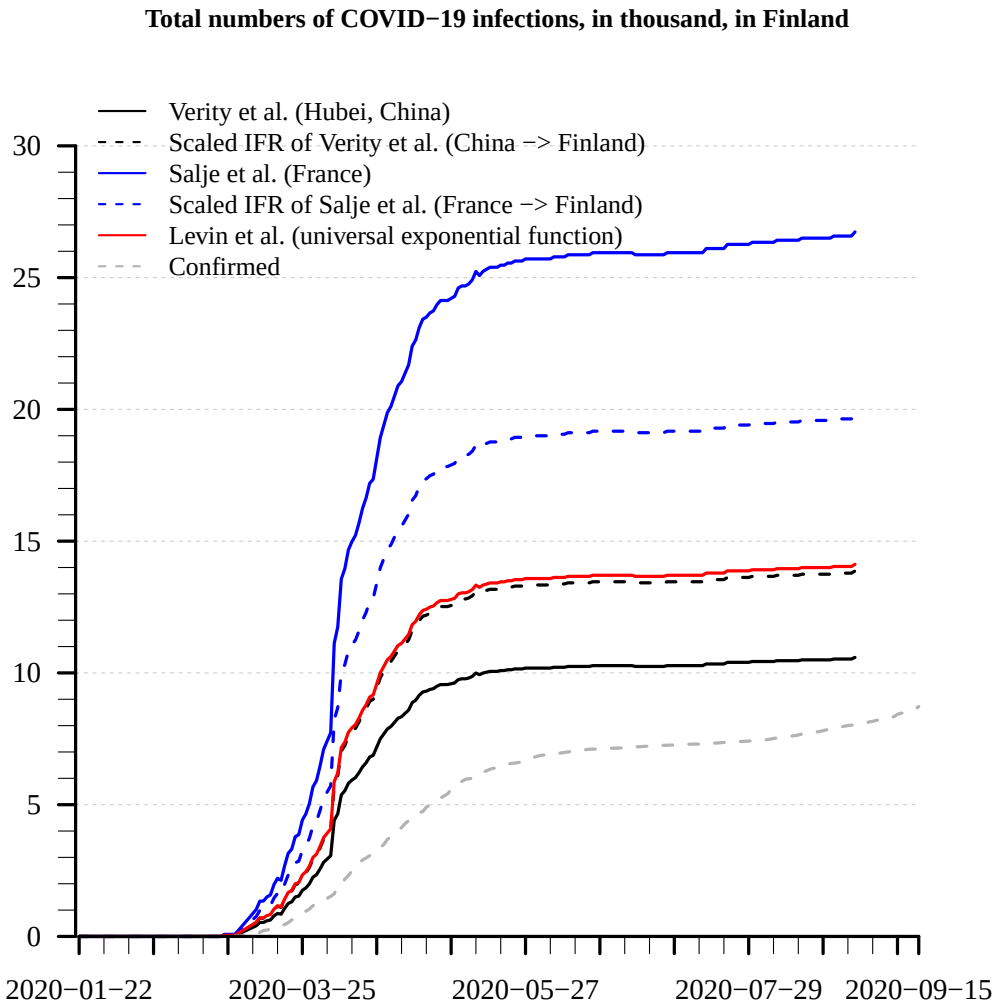
$\rightarrow$ You need to know the foundation of the model you use and how it behaves in the real world in order to assess its findings

Brief Q&A: recap material of previous session

How we have evaluated the demographic scaling model so far:

- Look at model's key assumptions and reflect on to what extent they may hold in the real world and how they might (potentally) impair the model's results.
- Look at how sensitive the model's results are with respect to (changes in) the input parameter: IFR_x.
  - IFR_x are rarely available yet
  - Do the IFR_x impact the model's results at all?
  - If model's results change substantially with respect to IFR_x, it means that they are sensitive to IFR_x and that it does matter what IFR_x you use, and vice versa.

Brief Q&A: recap material of previous session



→ Estimates of total numbers of COVID-19 infections for Finland appear to be sensitive to age-specific infection fatality rates as input.

Present your findings of previous lab session:

- How large are the COVID-19 infection estimates for Finland based on different sources for the IFR_x until mid November 2020?
- Have you done the same analysis for another country? Do the validation results differ?

$\longrightarrow$ Open issues?

Present your findings of previous lab session:

Some more questions for validating a model:

- Are the results specific for a country of interest and point in time? Or are these rather general findings that apply to multiple countries and points in time?
- I send you through the scientific process for developing the demographic scaling model. These are general steps that could be applicable / transferrable to your scientific discipline and research. Please start to reflect on this as we will talk about it in the lab of week 7.

Topic today

Validity & sensitivity

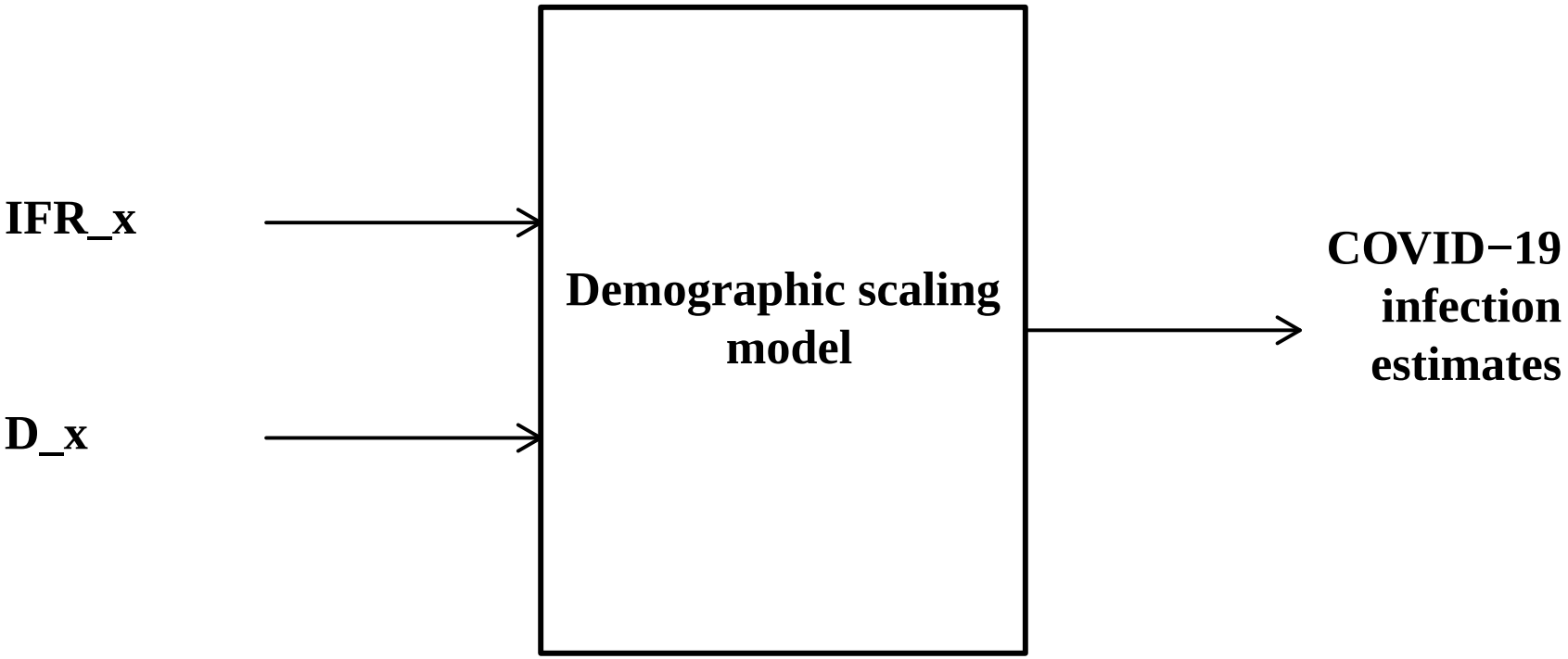
of demographic scaling model's

COVID-19 infection estimates.

$\longrightarrow$ Particularity: no *true* values have been observed

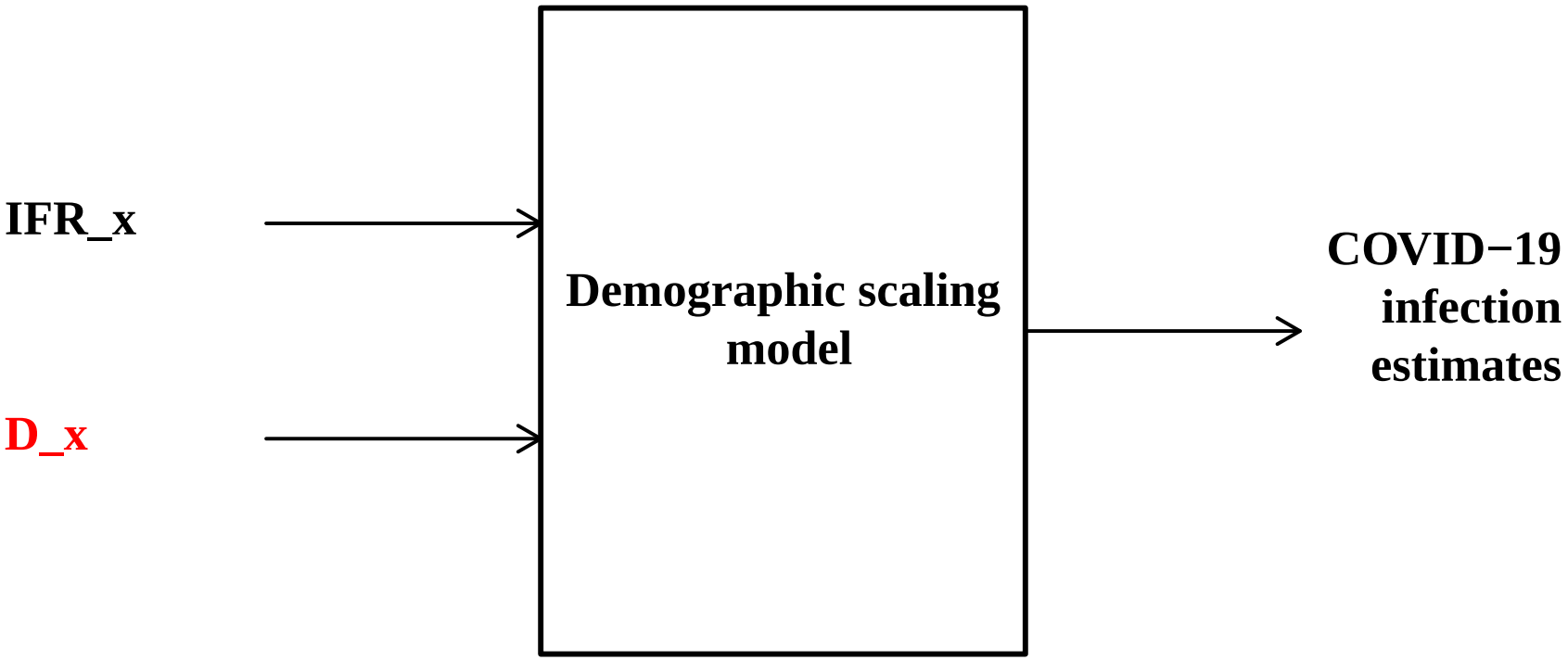
that we could compare with our model infection estimates

The demographic scaling model that needs to be evaluated



→ We have already looked at the plausibility of the model's two key assumptions & its results' sensitivity with respect to infection fatality rates

How sensitive are model infection estimates wrt D_x?



$\longrightarrow$ Think creatively and critically about the possible impact of COVID-19-related death counts, D_x, on model's infection estimates.

Topic today

Sensitivity of demographic scaling model's COVID-19 infection estimates **with respect to** D_x

Sensitivity wrt D_x

- Aim: test by how much the COVID-19 infection estimates change if we take COVID-19-related death counts from different sources.
- New sources for COVID-19 deaths become available over time:
  - Total death counts caused by COVID-19 from the Johns Hopkins University CSSE, disaggregated by age using a global age pattern based on data of Dudel et al. (2020).
  - Age-specific death counts caused by COVID-19 from the COVerAGE-DB database.
  - Excess mortality based on data of the short-term mortality fluctuations (STMF) data series of the Human Mortality Database.
  - ...

Sensitivity wrt D_x

What do you think?

What to possibly think creatively and critically about here?

Sensitivity wrt D_x

- Aim: test by how much the COVID-19 infection estimates change if we take COVID-19-related death counts from different sources.
- New sources for COVID-19 deaths become available over time.
- Think creatively and critically about the plausibility of these D_x estimates and their potential implications for the results of the demographic scaling model.
  - How much do these D_x estimates differ in the total number of COVID-19-related deaths?
  - How much do these D_x estimates differ in their age profile?
  - In what situations does the global age pattern of COVID-19-related deaths works well and when does it not work well?
  - How robust are the COVID-19 infection estimates with respect to these D_x estimates in general and how much does this robustness vary across countries and over time?
  - ...

Sensitivity wrt D_x

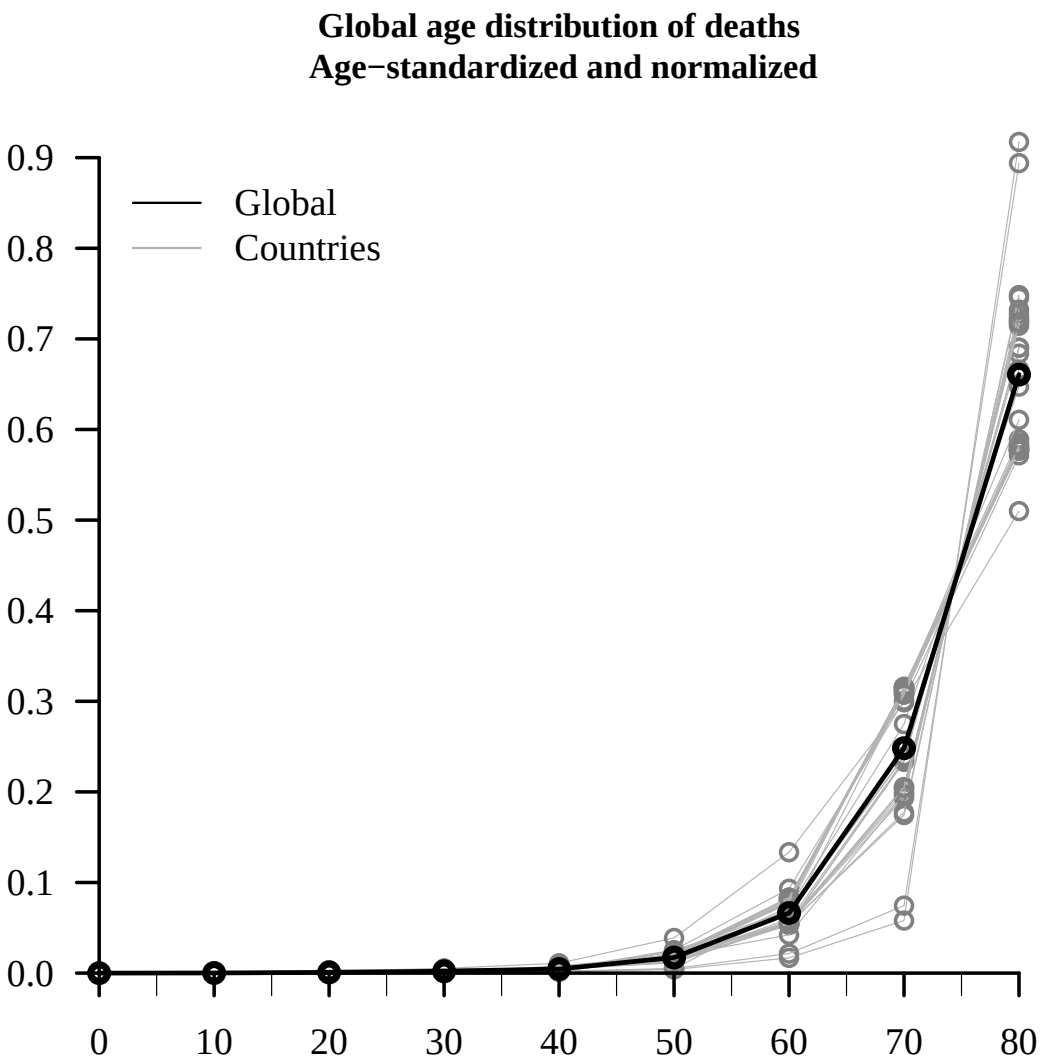
Let us start

and have a look at these different D_x estimates

Total deaths of JHU CSSE & global age pattern

- The Johns Hopkins University CSSE provides total death counts caused by COVID-19 on a daily basis since Jan 22, 2020
- `https://github.com/CSSEGISandData/COVID-19`
- We disaggregate these total death counts into 10-year age groups with a global age pattern based on data of Dudel et al. (2020):
  - Based on age-specific death counts for sample of countries
  - $\frac{\frac{D_x}{P_x}}{\sum_x \frac{D_x}{P_x}}$
  - Age -standardized and normalized to sum to one

Total deaths of JHU CSSE & global age pattern



Source: Bohk-Ewald et al. (2020)

Total deaths of JHU CSSE & global age pattern

- JHU CSSE provides total death counts caused by COVID-19 on a daily basis since Jan 22, 2020
- `https://github.com/CSSEGISandData/COVID-19`
- We disaggregate these total death counts into 10-year age groups with a global age pattern based on data of Dudel et al. (2020)
- Benefit: demographic scaling model is broadly applicable to the many countries that are covered in the JHU CSSE data since Jan 22, 2020
- Drawback: the country-specific age patterns of COVID-19-related death counts are missing and are likely to impact the COVID-19 infection estimates of the demographic scaling model

$\rightarrow$ However, simplification has been necessary due to the lack of data in the early stage of the coronavirus pandemic

D_x from the COVerAGE-DB

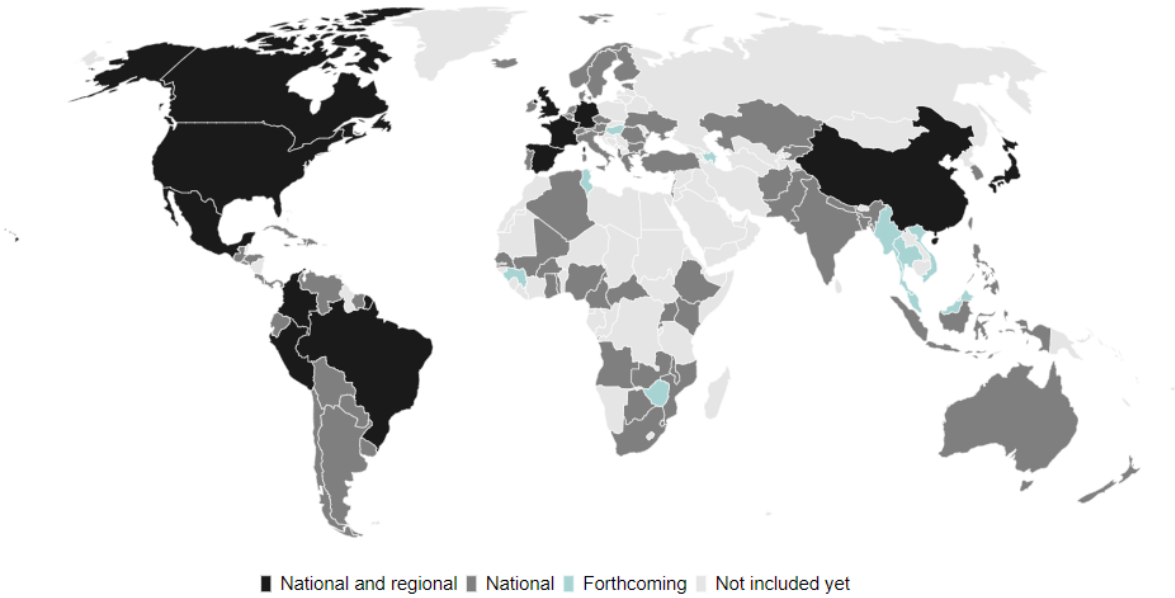
- COVerAGE-DB database provides age-specific death counts caused by COVID-19
- Project of the Max Planck Institute for Demographic Research, lead by Tim Riffe and Enrique Acosta, since late March 2020
- COVerAGE-DB collects and harmonizes cases and deaths related to COVID-19 by age & sex & time for a growing number of countries
- Tim Riffe and Enrique Acosta (2020)
  COVeAGE-DB: A database of age-structured COVID-19 cases and deaths. medRxiv 2020; published online September 23.
  `https://doi.org/10.1101/2020.09.18.20197228`
- GitHub repository at `https://github.com/timriffe/covid_age`
- Most recent data available at `https://osf.io/mpwjq/`

D_x from the COVerAGE-DB

Data availability

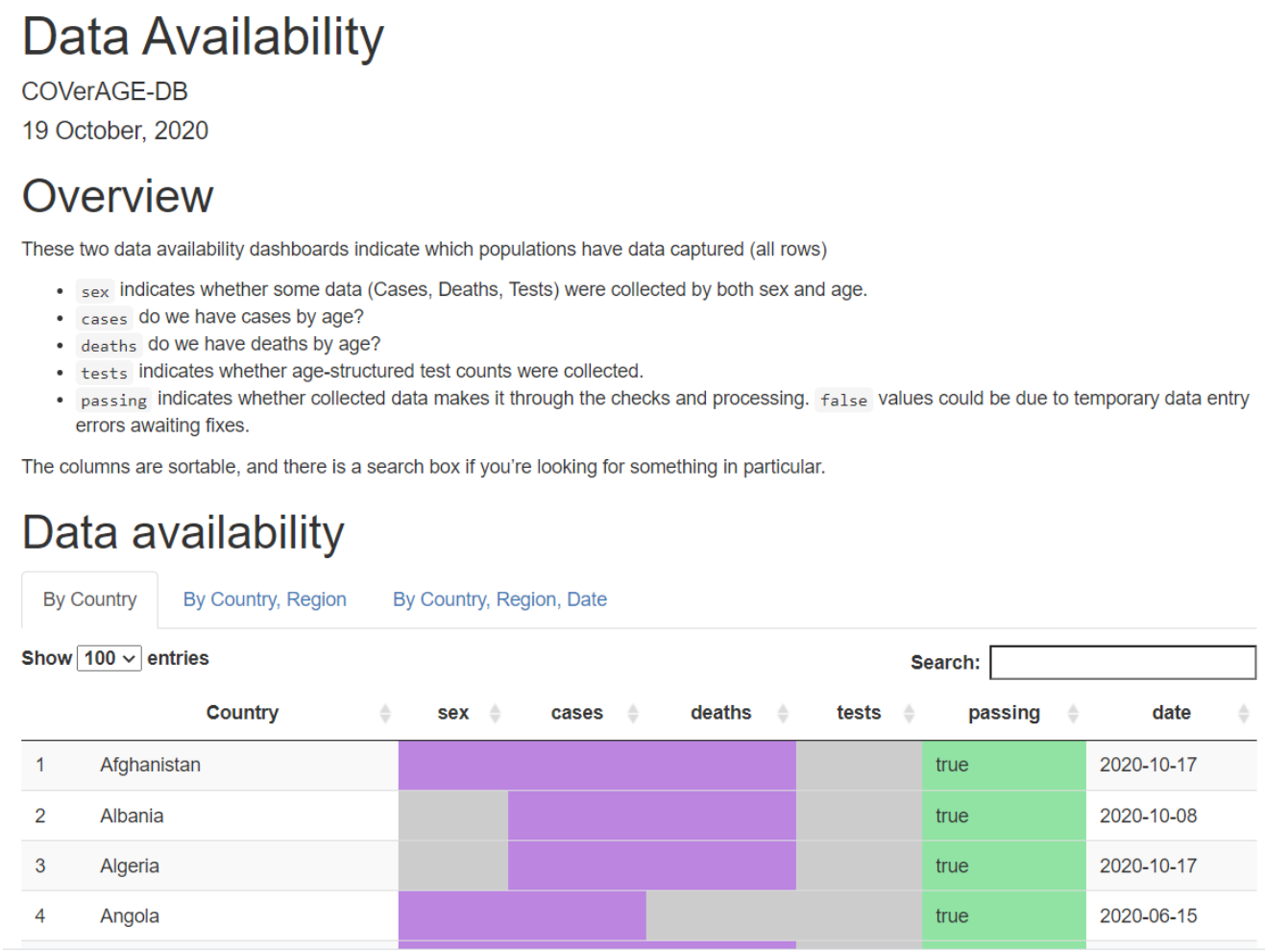
You can get the most up-to-date data at the `OSF` site that we mirror to: https://osf.io/mpwjq/.

Here's an overview of global coverage as of now. A country marked as *forthcoming* means we've identified a source, but that collection is pending for one reason or another. Are you from one of the countries not yet in the collection and want to pitch in? Please reach out, if so by emailing us as `coverage-db < at > demogr.mpg.de`.



Source: `https://github.com/timriffe/covid_age`; accessed on Oct 19, 2020

D_x from the COVerAGE-DB



Data Availability

COVerAGE-DB

19 October, 2020

Overview

These two data availability dashboards indicate which populations have data captured (all rows)

- `sex` indicates whether some data (Cases, Deaths, Tests) were collected by both sex and age.
- `cases` do we have cases by age?
- `deaths` do we have deaths by age?
- `tests` indicates whether age-structured test counts were collected.
- `passing` indicates whether collected data makes it through the checks and processing. `false` values could be due to temporary data entry errors awaiting fixes.

The columns are sortable, and there is a search box if you're looking for something in particular.

Data availability

By Country | By Country, Region | By Country, Region, Date

Show 100 entries Search:

	Country	sex	cases	deaths	tests	passing	date
1	Afghanistan					true	2020-10-17
2	Albania					true	2020-10-08
3	Algeria					true	2020-10-17
4	Angola					true	2020-06-15

Source: `https://timriffe.github.io/covid_age/DataAvail.html`; accessed on Oct 19, 2020; 88 countries covered in COVerAGE-DB

D_x from two sources: pros and cons

Total deaths of JHU CSSE & global age pattern:

- Benefit: COVID-19-related total death counts are available for many countries since Jan 22, 2020
- Drawback: the country-specific age patterns of COVID-19-related death counts are missing and are likely to impact the COVID-19 infection estimates of the demographic scaling model

COVerAGE-DB database:

- Benefit: age-specific death counts caused by COVID-19 are available for many countries over time
- Drawback: data are available for fewer countries and shorter time series (with missing data for some dates), which is likely to reduce the applicability of the demographic scaling model

D and D_x from two sources

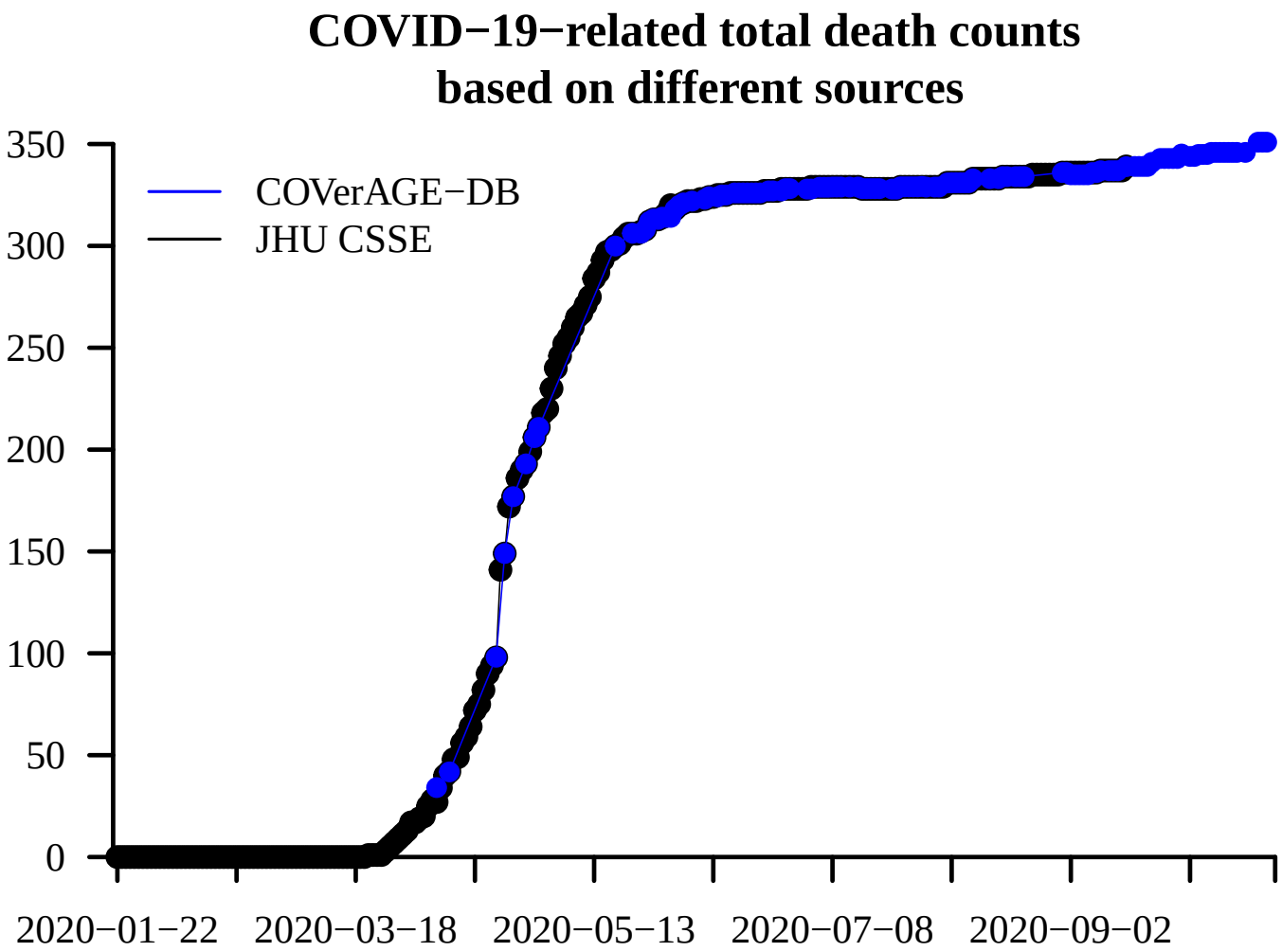
How well do the Finnish data

on COVID-19-related death counts

from the COVerAGE-DB

compare with the death counts from the JHU CSSE?

D estimates for Finland



D estimates for Finland

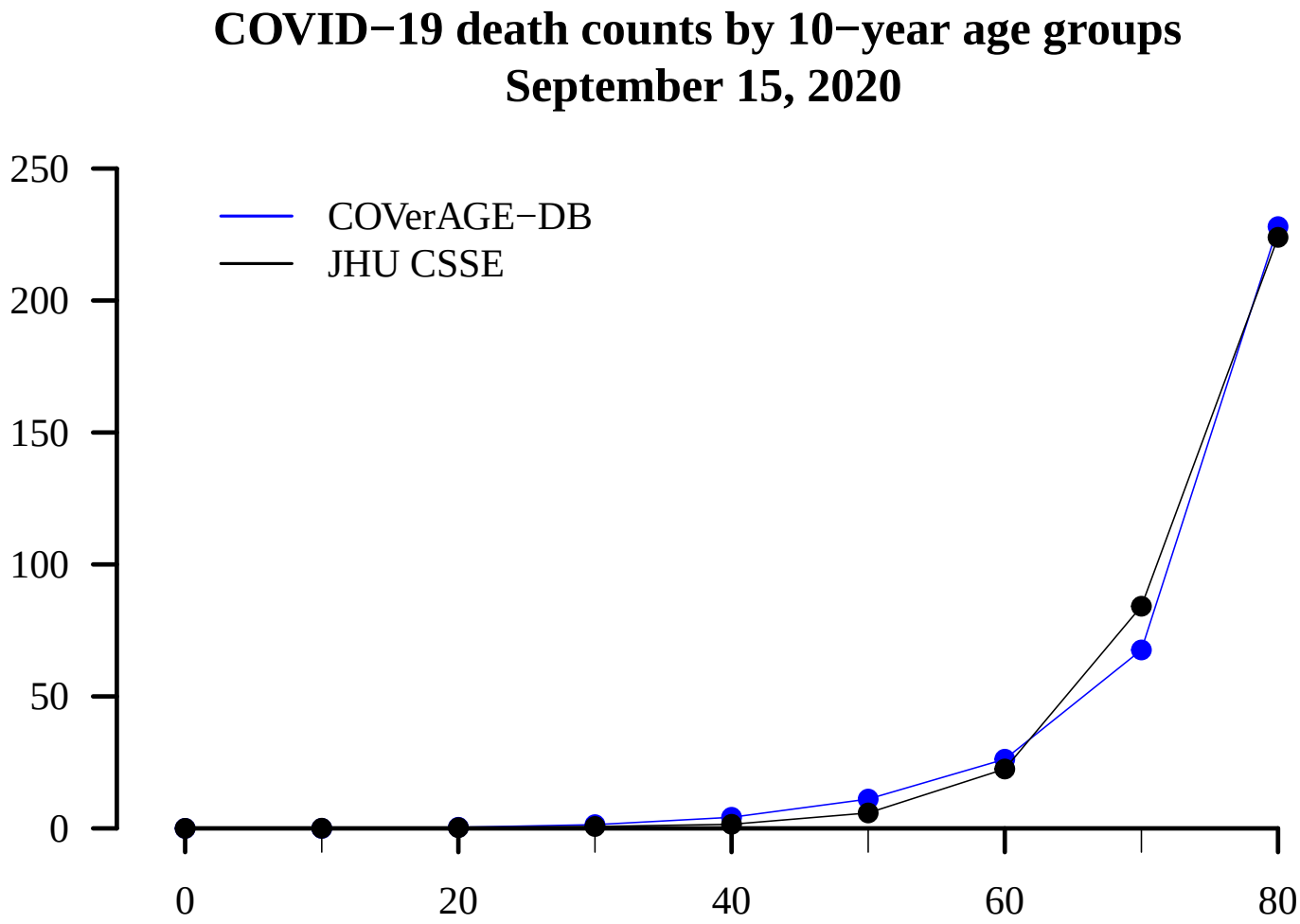
What do you think:

How much do you expect

the death counts to impact the COVID-19 infection estimates

of the demographic scaling model?

D_x estimates for Finland



D_x estimates for Finland

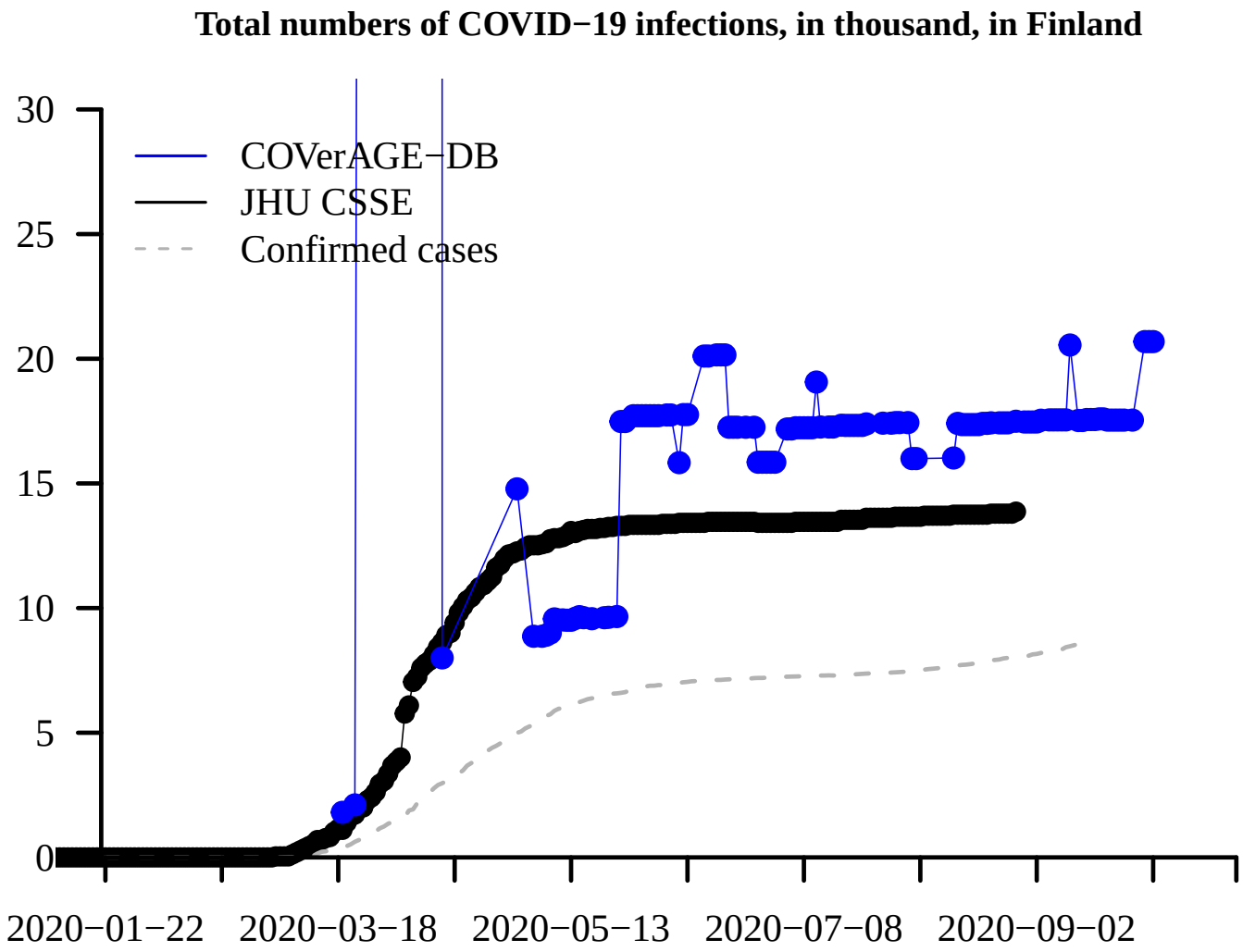
What do you think:

How much do you expect

the death counts to impact the COVID-19 infection estimates

of the demographic scaling model?

Infection estimates for Finland



Infection estimates for Finland

What do you think:

How much do the death counts impact

the COVID-19 infection estimates

of the demographic scaling model?

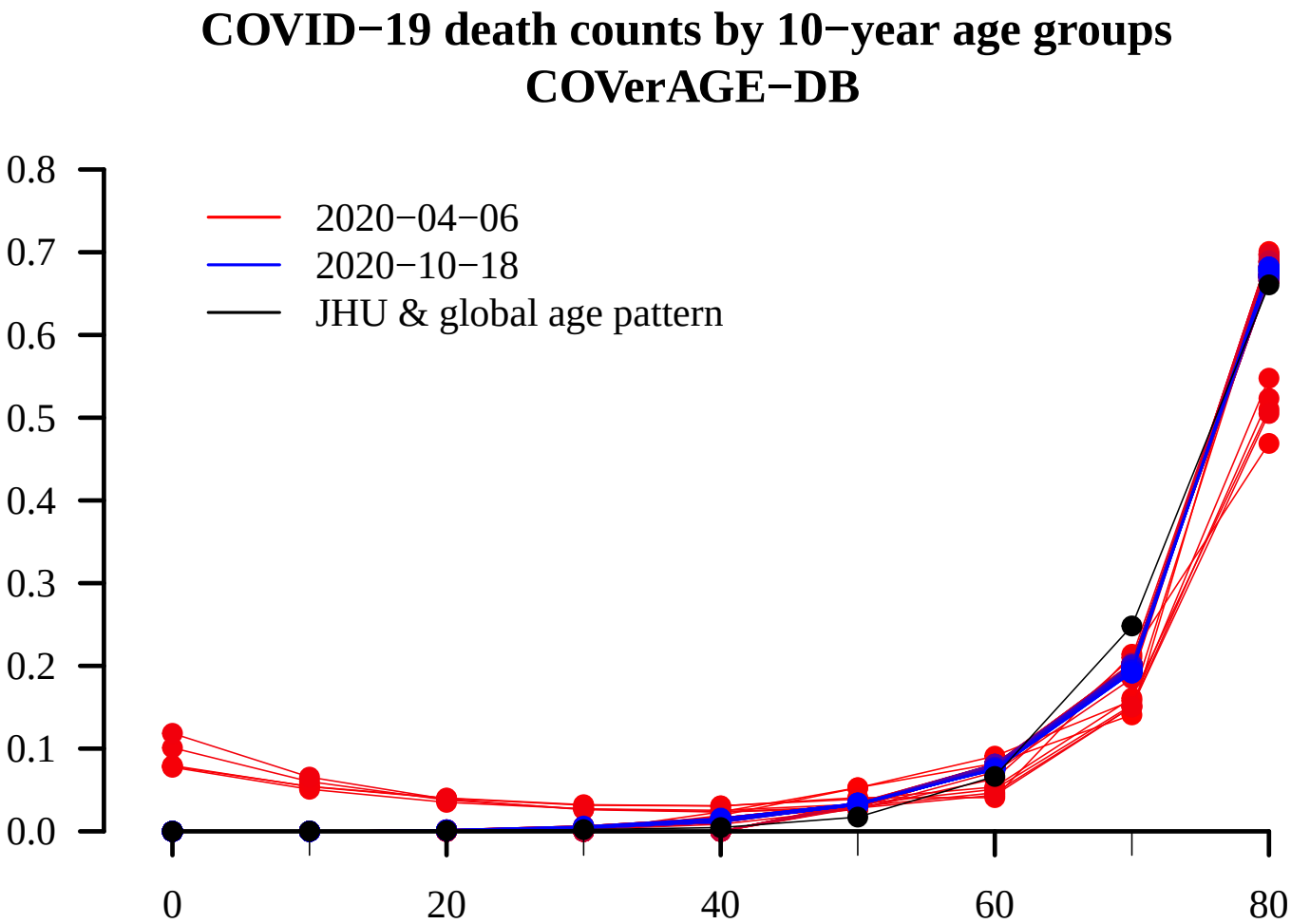
$\Rightarrow$ Gather as many information as possible

Sensitivity of infection estimates wrt D_x for Finland

- COVID-19 infection estimates for Finland appear to be sensitive to D_x.
- COVID-19 infection estimates for Finland appear to be larger when they are based on D_x of COVerAGE-DB than on D_x of JHU CSSE & the global age pattern.
- COVID-19 infection estimates for Finland appear to be erratic and noisy when they are based on D_x of COVerAGE-DB. Why?
  - $\rightarrow$ Small numbers in Finland?
  - $\rightarrow$ Age profile of D_x in COVerAGE-DB unsmooth over time?
  - ...
  - HOW TO TEST FOR THIS?

$\Rightarrow$ Are these results generalizable? Do they vary across countries?

COVerAGE-DB D_x estimates for Finland



$\Rightarrow$ Global age pattern appears to bery similar to recent age pattern of Finnish deaths related to COVID-19

Sensitivity of infection estimates wrt D_x for Finland

The example for Finland also shows:

- D_x estimate appears to have big impact on estimate of total numbers of COVID-19 infections of the demographic scaling model.
- The global age pattern appears to bery similar to the recent age pattern of Finnish deaths related to COVID-19 from the COVer-AGE-DB.
- All COVID-19 infection estimates are consistently larger than the numbers of confirmed cases.

Sensitivity wrt D_x

Think creatively and critically ...and extend this analysis:

- How much do these D_x estimates differ in the total number of COVID-19-related deaths?
- How much do these D_x estimates differ in their age profile?
- In what situations does the global age pattern of COVID-19-related deaths works well and when does it not work well?
- How sensitive are the COVID-19 infection estimates with respect to these D_x estimates in general and how much does this sensitivity vary across countries and over time?
- ...

Infection estimates for Finland

What do you think:

What impact is bigger: the one of IFR_x or D_x estimates?

Or the combined effect of IFR_x and D_x estimates?

$\Rightarrow$ How to test for this?

What you have learned today about assessing the demographic scaling model

- Describe the impact of IFR_x estimates on the COVID-19 infection estimates for Finland.
- Describe different sources for COVID-19-related death counts.
- Explain how to analyze the sensitivity of COVID-19 infection estimates of the demographic scaling model with respect to D_x estimates.
- Describe tentative results regarding the impact of D_x estimates on the COVID-19 infection estimates for Finland.

Course learning materials

Course learning materials on GitHub:

`https://github.com/christina-bohk-ewald/2020-COS-D407-scientific-modeling-and-model-validation`

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`
- **Tim Riffe and Enrique Acosta (2020)**
  COVeAGE-DB: A database of age-structured COVID-19 cases and deaths. medRxiv 2020; published online September 23. `https://doi.org/10.1101/2020.09.18.20197228`

Thank you for your time and attention!

christina.bohk-ewald@helsinki.fi

Fifth week's class in the lab: Sensitivity of demographic scaling model's results.

- Test the sensitivity of the demographic scaling model's results for Finland with respect to death counts from two different sources:
  - Total death counts caused by COVID-19 from the Johns Hopkins University CSSE & global age pattern.
  - Age-specific death counts caused by COVID-19 from the COVerAGE-DB database.
  - Critically think about the plausibility of these D_x estimates and their implications for the results of the demographic scaling model.
- Systematically analyze the combined impact of various IFR_x and D_x estimates on the results of the demographic scaling model.

$\longrightarrow$ Present and discuss your findings in class at the beginning of the next session on Monday.