

Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework

Brieuc Lehmann

@brieuclehmann

Department of Statistical Science

University College London

Turing-RSS Health Data Lab

@turingrss_hdlab

Turing-Roche Knowledge Share

30th May 2022



COVID-19 testing data is biased

“If we did half the testing, we would have half the cases.”

COVID-19 testing data is biased

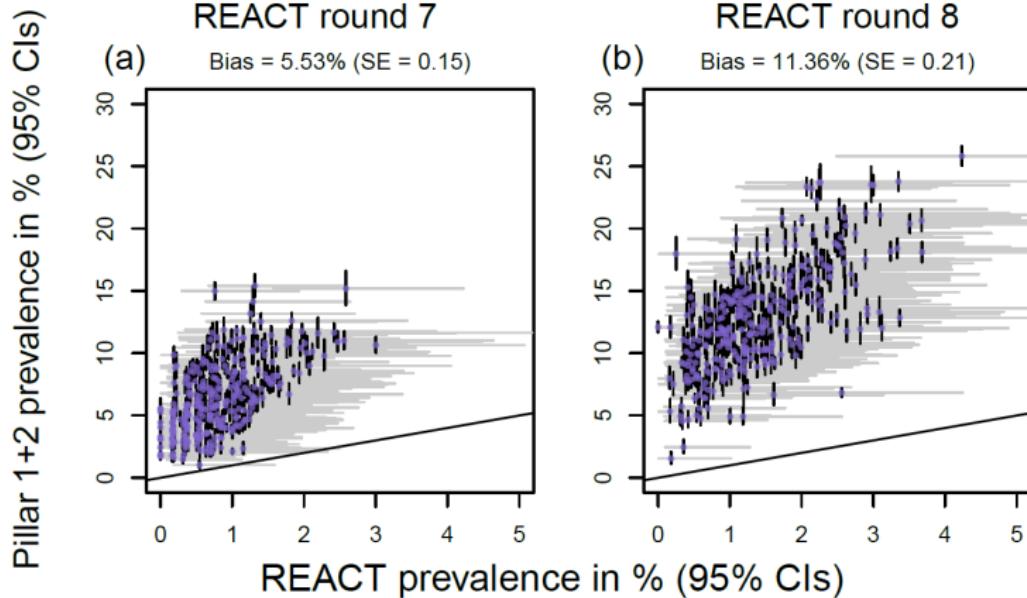
“If we did half the testing, we would have half the cases.”

The 45th President of the United States of America

Donald J. Trump (July, 2020)

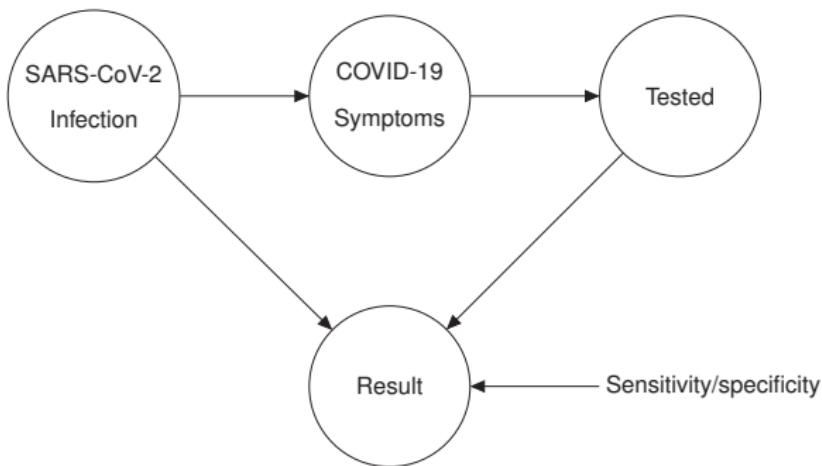


Pillar 1+2 positivity rate vs REACT prevalence



Can we debias fine-scale Pillar1+2 data?

Pillar 1+2 testing is subject to ascertainment bias, for example:



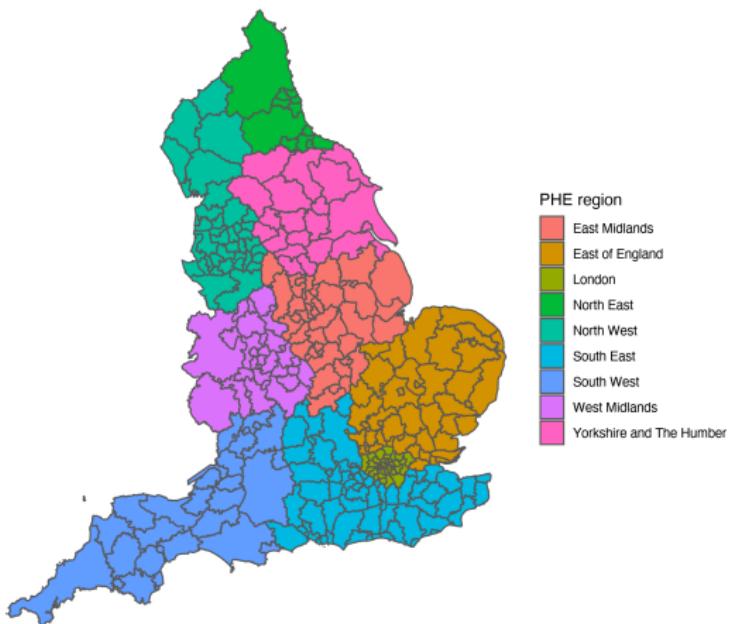
Can we model this bias to estimate fine-scale prevalence accurately?

SARS-CoV-2 testing data – REACT and Pillar 1+2

Test counts (PCR only, positive and total counts):

- REACT, u of U , is randomised surveillance
 - Stratified random sample from England's NHS register
 - Repeated sampling **rounds** of this randomly selected cohort
- Pillar1+2, n of N , is targeted surveillance; roughly speaking:
 - Pillar 1 targets health and care workers
 - Pillar 2 targets the **symptomatic** wider population

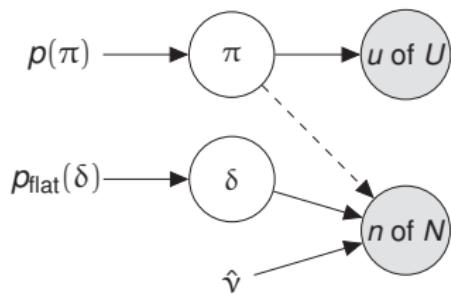
Geographic partitions of England (coarse v fine)



Overview of model

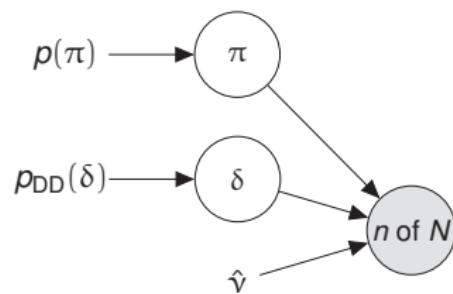
(a) Coarse scale (e.g. PHE region)

Learn δ via “cut” model



(b) Fine scale (e.g. LTLA)

Learn π from Pillar1+2 data



Key

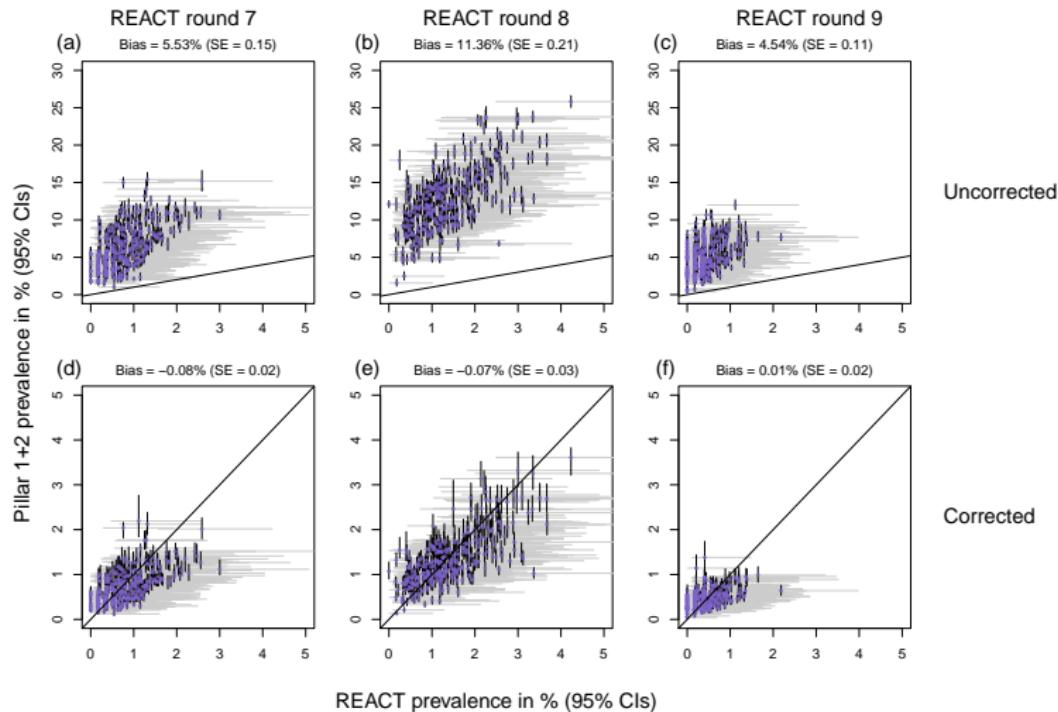
π : prevalence

δ : bias

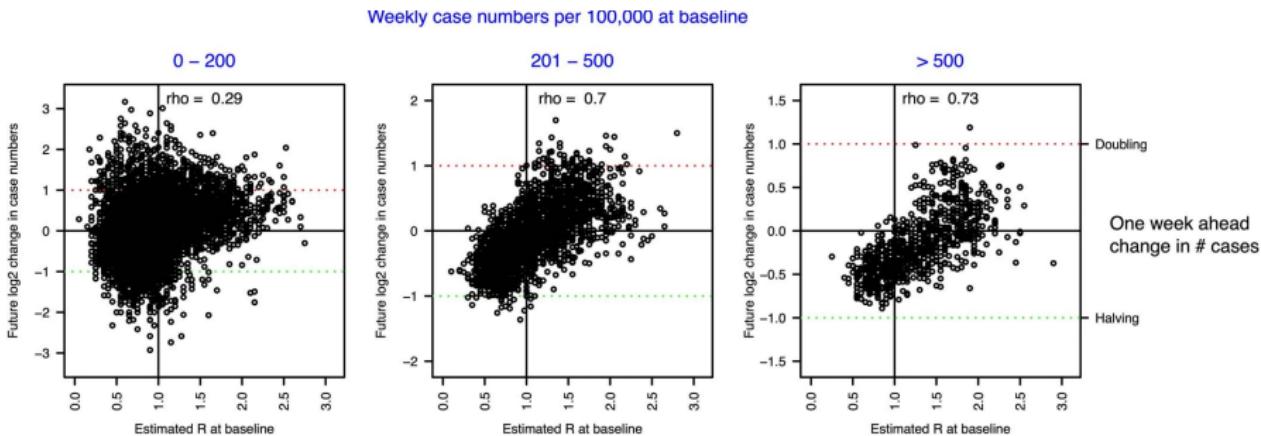
u of U : REACT data

n of N : Pillar 1+2 data

Uncorrected prevalence v. debiased prevalence



Predicting changes in case numbers from estimated \mathcal{R}_t



Take-home messages

Accounting for **ascertainment/sampling bias** in datasets is crucial.

- Inference & prediction can suffer when generalizing results

Take-home messages

Accounting for **ascertainment/sampling bias** in datasets is crucial.

- Inference & prediction can suffer when generalizing results

Randomized studies are extremely valuable!

- REACT and ONS CIS are excellent UK-based studies

Take-home messages

Accounting for **ascertainment/sampling bias** in datasets is crucial.

- Inference & prediction can suffer when generalizing results

Randomized studies are extremely valuable!

- REACT and ONS CIS are excellent UK-based studies

Combining with **observational studies** can bring the best of both worlds.

References ('Lab' papers)

- **Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework**

(Nature Microbiology, 2021)

doi.org/10.1038/s41564-021-01029-0

- **Interoperability of statistical models in pandemic preparedness: principles and reality**

(Statistical Science, 2022)

doi.org/10.1214/22-STS854