Bewertung der Qualität und Glaubwürdigkeit von Beobachtungsstudien

Sabine Hoffmann
StaBLab, LMU München

24.06.2025

1. Challenges in the analysis of retrospective studies
   - Eligibility
   - Data quality
   - Time point alignment
   - Interventions and tests are not random
   - Multiplicity of possible analysis strategies
2. How do flexible AI/ML algorithms change these challenges?
3. Criteria to judge the quality and credibility of observational studies
4. Conclusion

Challenges in the analysis of retrospective studies

Retrospective studies vs. prospective studies

- In prospective studies, the research question informs the study design. In randomised studies, it even shapes the data generating mechanism

Retrospective studies vs. prospective studies

- In prospective studies, the research question informs the study design. In randomised studies, it even shapes the data generating mechanism
- In retrospective studies, there is in general less knowledge of how the data were generated and less control over measurement procedures

Retrospective studies vs. prospective studies

- In prospective studies, the research question informs the study design. In randomised studies, it even shapes the data generating mechanism
- In retrospective studies, there is in general less knowledge of how the data were generated and less control over measurement procedures
- Data from retrospective studies are often analysed using the same methods as prospective studies, but it is important to realise that the analysis of these data sets and the interpretation of results is much more challenging

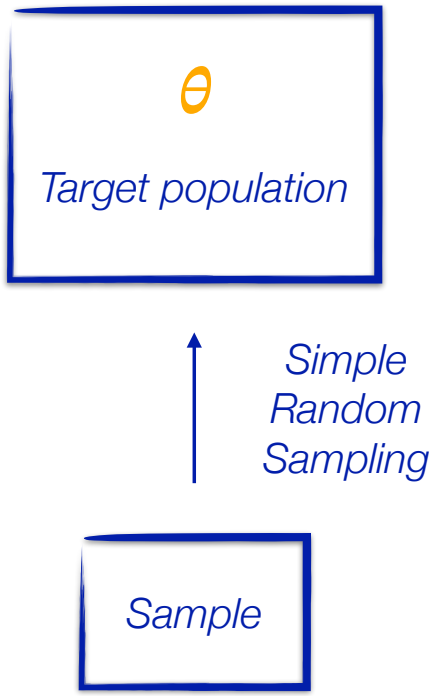
Eligibility

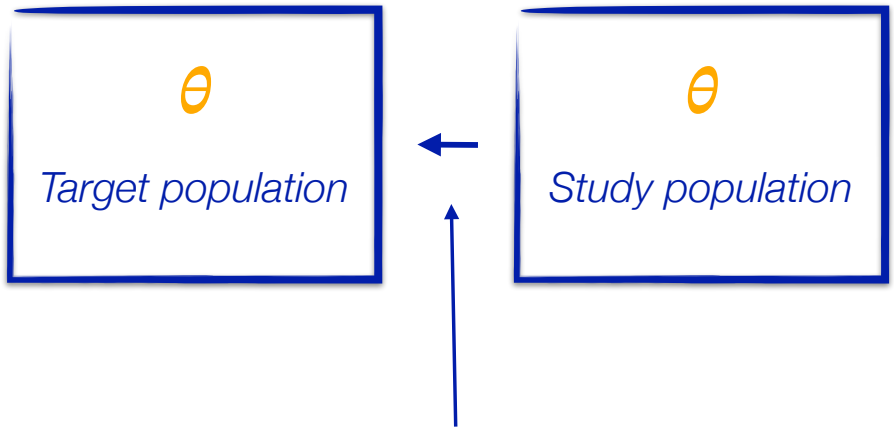
- In contrast to prospective studies, the inclusion and exclusion of patients in retrospective studies is usually dictated, or at least limited, by data availability

- In contrast to prospective studies, the inclusion and exclusion of patients in retrospective studies is usually dictated, or at least limited, by data availability
- Due to disparities in the access and use of health-care services, routinely collected data may under- or mis-represent certain subgroups[1,2,3,4,5,6,7,8], including
  - ethnic minorities
  - patients without medical coverage
  - low-income and rural populations

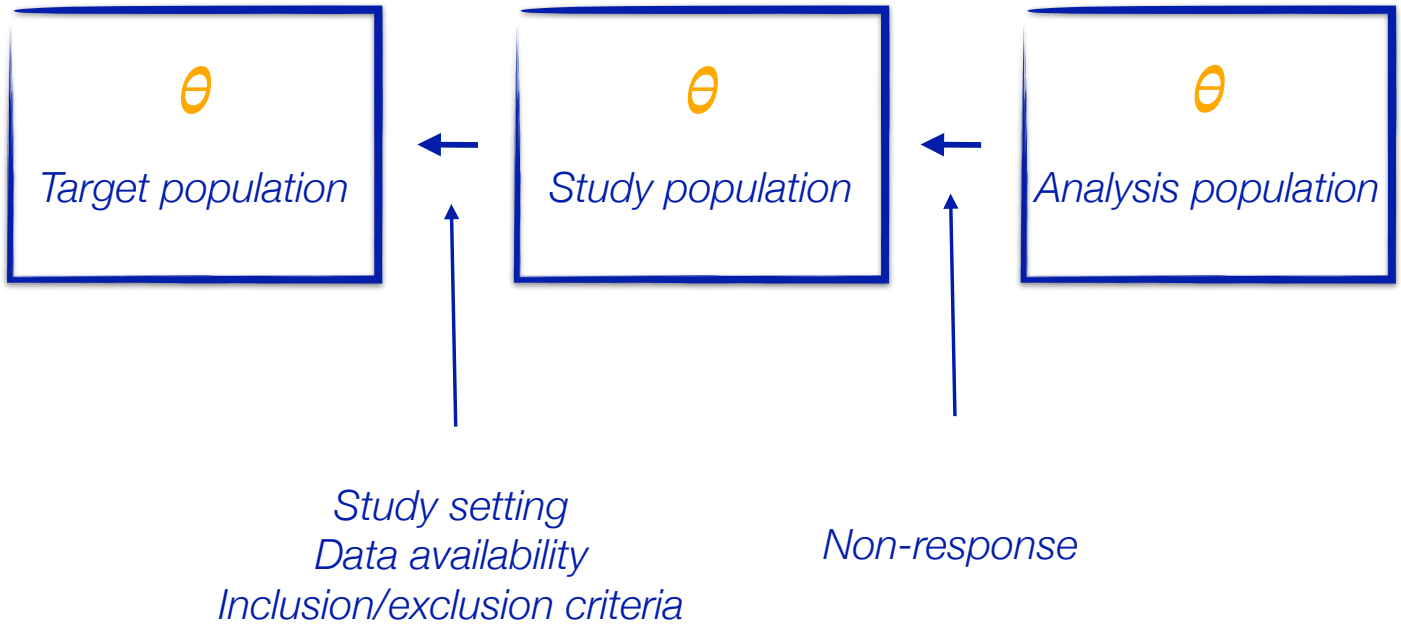
- In contrast to prospective studies, the inclusion and exclusion of patients in retrospective studies is usually dictated, or at least limited, by data availability
- Due to disparities in the access and use of health-care services, routinely collected data may under- or mis-represent certain subgroups[1,2,3,4,5,6,7,8], including
  - ethnic minorities
  - patients without medical coverage
  - low-income and rural populations
- At the same time, there is evidence that women, elderly, more educated patients and patients with a greater burden of disease are overrepresented

- In contrast to prospective studies, the inclusion and exclusion of patients in retrospective studies is usually dictated, or at least limited, by data availability
- Due to disparities in the access and use of health-care services, routinely collected data may under- or mis-represent certain subgroups[1,2,3,4,5,6,7,8], including
  - ethnic minorities
  - patients without medical coverage
  - low-income and rural populations
- At the same time, there is evidence that women, elderly, more educated patients and patients with a greater burden of disease are overrepresented
- If patients from under-represented groups are present in the data, there is a risk that they may be mis-represented, because they are more likely to visit multiple institutions[9,10,11,12] and they receive fewer diagnostics tests and interventions[3]

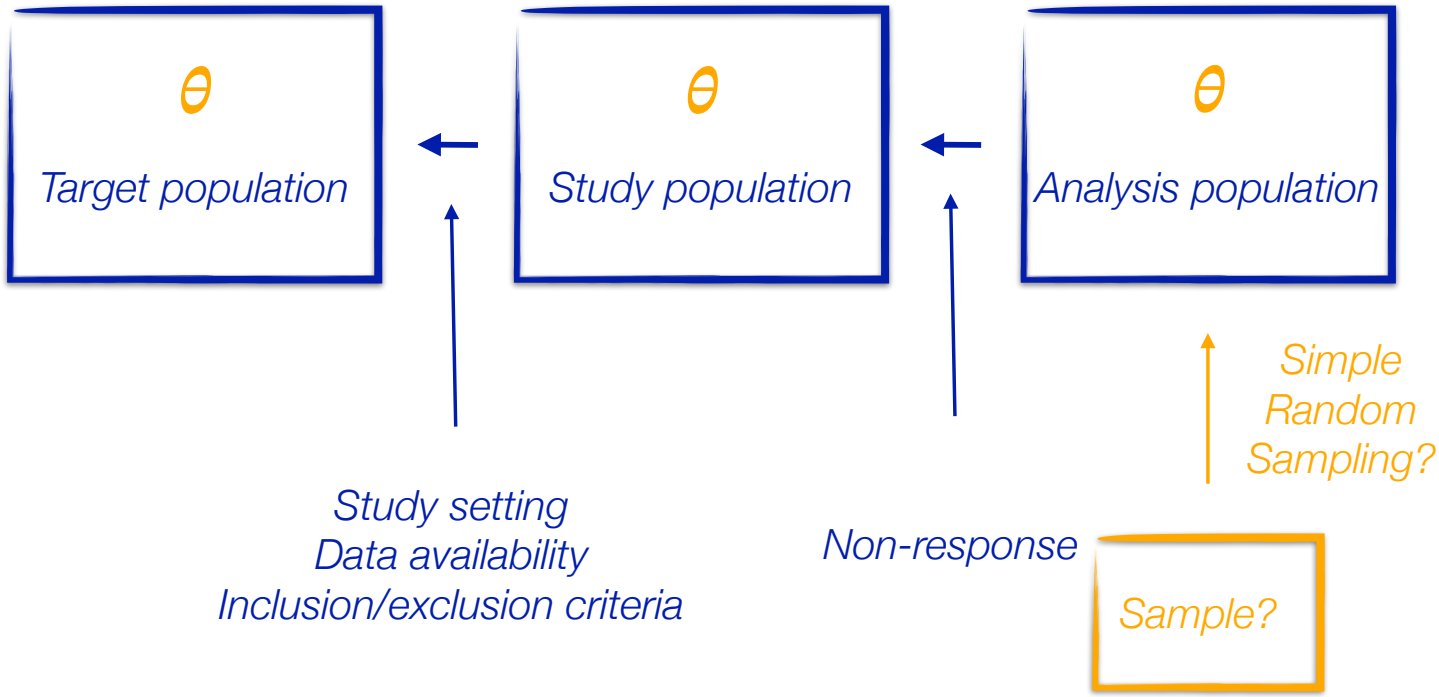

θ
Target population
Simple
Random
Sampling
Sample



Adapted from: Degtiar, I., Rose, S. (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, *10*(1), 501-524.

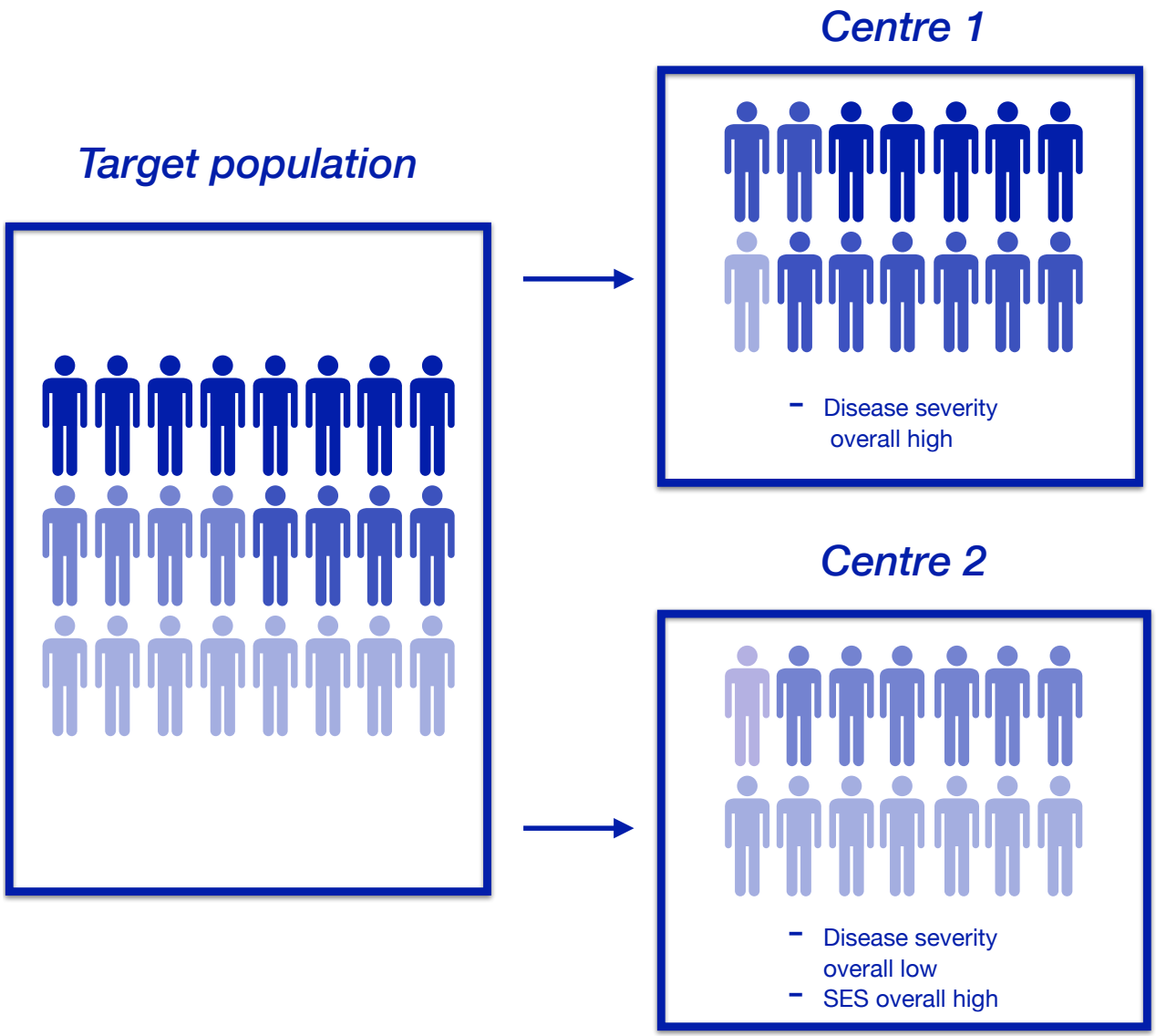


Adapted from: Degtiar, I., Rose, S. (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, *10*(1), 501-524.



Adapted from: Degtiar, I., Rose, S. (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, *10*(1), 501-524.

Differences in eligibility across centres



Addressing issues arising from eligibility

Results will be biased if the data differs from the target population with respect to treatment effect modifiers[13,14]. Selection on a consequence of the exposure and the outcome will induce collider bias[15]

Addressing issues arising from eligibility

Results will be biased if the data differs from the target population with respect to treatment effect modifiers[13,14]. Selection on a consequence of the exposure and the outcome will induce collider bias[15]

Solutions:

- It is important to report patient characteristics overall in the sample, for different centres and to compare these characteristics to the target population

Addressing issues arising from eligibility

Results will be biased if the data differs from the target population with respect to treatment effect modifiers[13,14]. Selection on a consequence of the exposure and the outcome will induce collider bias[15]

Solutions:

- It is important to report patient characteristics overall in the sample, for different centres and to compare these characteristics to the target population
- Approaches like inverse probability weighting and multilevel regression modelling with post-stratification can improve representativeness

Data quality

Data quality

- Retrospective data are not recorded for a specific research purpose at hand

Data quality

- Retrospective data are not recorded for a specific research purpose at hand

$\Rightarrow$ Data entries may be incomplete, inaccurate, inconsistently collected and systematically biased

Data quality

- Retrospective data are not recorded for a specific research purpose at hand

$\Rightarrow$ Data entries may be incomplete, inaccurate, inconsistently collected and systematically biased

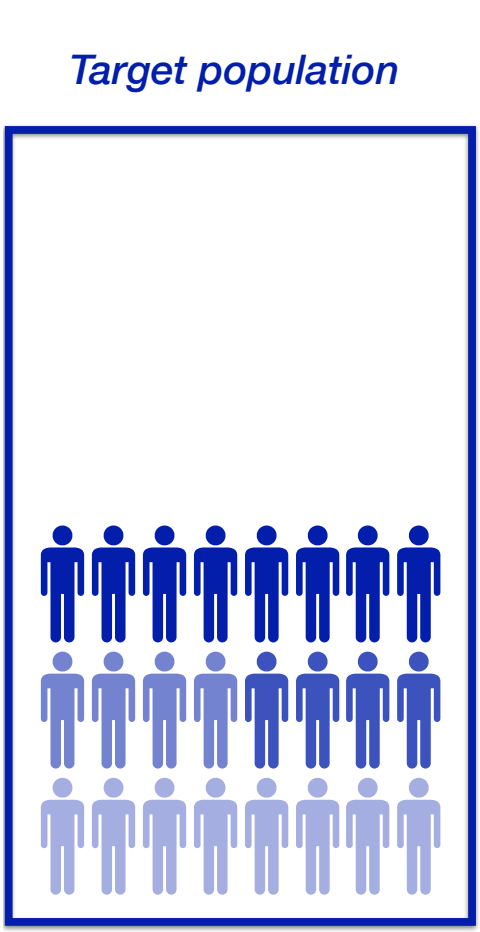
- Examples:
  - On an emergency call, vital parameters may not be measured if they are irrelevant to the clinical question at hand
  - Prescription orders may not be filled or consumed by the patient
  - Temporal changes in the recording of data may produce systematic differences over time

Differences in data quality across centres

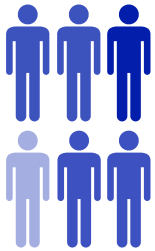
- Documentation practices may vary between different clinical settings, as a function of incentives and of the overall workload of the personnel collecting the data [16,17]

Differences in data quality across centres

- Documentation practices may vary between different clinical settings, as a function of incentives and of the overall workload of the personnel collecting the data [16,17]
- Combined with differences in eligibility, variations in imagining techniques, sensitivity of test kits and coding accuracy can create spurious associations



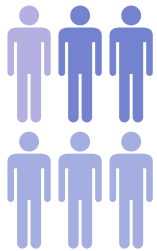
Hospital 1



- Disease severity overall high

- Treatment A, B and C available
- Quality of care overall high
- Blood values measured at hospital admission
- Central laboratory measurements for all patients
- Additional point-of-care blood gas analysis only for patients with high disease severity
- Advanced imagining techniques with high sensitivity

Hospital 2



- Disease severity overall low
- SES overall high

- Treatment A and B are available
- Quality of care overall low
- Blood values measured 8-15 hours after hospital admission
- Reporting of blood values, irrespective of provenance from point-of-care blood gas analysis or central laboratory
- Basic imaging techniques with low sensitivity

ID	Hospital	Treatment	BGA	CL	Diagnosis
1	1	A			1
2	1	B			0
3	1	B			1
4	1	A			1
5	1	B			1
6	1	A			0
7	1	B			1
8	1	C			1
9	1	C			0
10	1	A			0
11	1	C			0
12	2	A			0
13	2	A			0
14	2	B			0
15	2	A			0
16	2				1
17	2	B			1
18	2	A			1
19	2	B			1
20	2				1
21	2	B			1
22	2	B			1
23	2				1

Addressing data quality

Informative missing data patterns and measurement error in treatment and confounders may lead to over- and underestimation of treatment effects

Addressing data quality

Informative missing data patterns and measurement error in treatment and confounders may lead to over- and underestimation of treatment effects

Solutions:

- It is important to perform extensive data quality checks, including plausibility checks and to evaluate multivariate outliers

Addressing data quality

Informative missing data patterns and measurement error in treatment and confounders may lead to over- and underestimation of treatment effects

Solutions:

- It is important to perform extensive data quality checks, including plausibility checks and to evaluate multivariate outliers
- Data audits and validation data:
  - Measurement error and (informative) missing data patterns can be characterised and quantified in validation studies in which prospective measurements of high quality are collected at the same time as the routine data collection

Addressing data quality

Informative missing data patterns and measurement error in treatment and confounders may lead to over- and underestimation of treatment effects

Solutions:

- It is important to perform extensive data quality checks, including plausibility checks and to evaluate multivariate outliers
- Data audits and validation data:
  - Measurement error and (informative) missing data patterns can be characterised and quantified in validation studies in which prospective measurements of high quality are collected at the same time as the routine data collection
- It is possible to account for informative missing data patterns and complex structures of measurement error in a Bayesian hierarchical approach

Time point alignment

Aspects related to timing in retrospective studies

- In retrospective studies, the first entry for a patient typically has no clear clinical meaning and it is often unknown which diagnoses or treatments the patient received before this time point

Aspects related to timing in retrospective studies

- In retrospective studies, the first entry for a patient typically has no clear clinical meaning and it is often unknown which diagnoses or treatments the patient received before this time point
- Data collection is influenced by daily clinical practice $\Rightarrow$ The timing of interventions and measurements may be misreported or missing

Aspects related to timing in retrospective studies

- In retrospective studies, the first entry for a patient typically has no clear clinical meaning and it is often unknown which diagnoses or treatments the patient received before this time point
- Data collection is influenced by daily clinical practice $\Rightarrow$ The timing of interventions and measurements may be misreported or missing
- Example: Nurses may only find the time to record clinical events or changes in medication at the end of their shift

Time point alignment

- In a representative sample of reports of comparative non-randomised studies that assessed the effectiveness and/or safety of drug treatments,[18] found that in 72% of studies eligibility, treatment assignment, and start of follow-up were not aligned.

Time point alignment

- In a representative sample of reports of comparative non-randomised studies that assessed the effectiveness and/or safety of drug treatments,[18] found that in 72% of studies eligibility, treatment assignment, and start of follow-up were not aligned.
- In a data audit on the quality of observational study data in an international HIV research network, treatment regimens and associated dates and the timings of laboratory measurements were especially prone to error with error rates of up to 56% and 42%[19]

Addressing timing in retrospective studies

Problems in time point alignment may result in immortal time bias [20,21] and bias arising from the depletion of susceptibles [22,23]

Addressing timing in retrospective studies

Problems in time point alignment may result in immortal time bias [20,21] and bias arising from the depletion of susceptibles [22,23]

Solutions:

- For every variable that is included in the analysis, it is important to report the time at which it was measured

Addressing timing in retrospective studies

Problems in time point alignment may result in immortal time bias [20,21] and bias arising from the depletion of susceptibles [22,23]

Solutions:

- For every variable that is included in the analysis, it is important to report the time at which it was measured
- Emulated target trial explicitly addresses the timing in retrospective studies by aligning the time at which eligibility criteria are met, treatment assignment and start of follow-up, but it is not a panacea

Aspects related to treatment assignment and inclusion

Intensive Care Med
https://doi.org/10.1007/s00134-025-07805-4

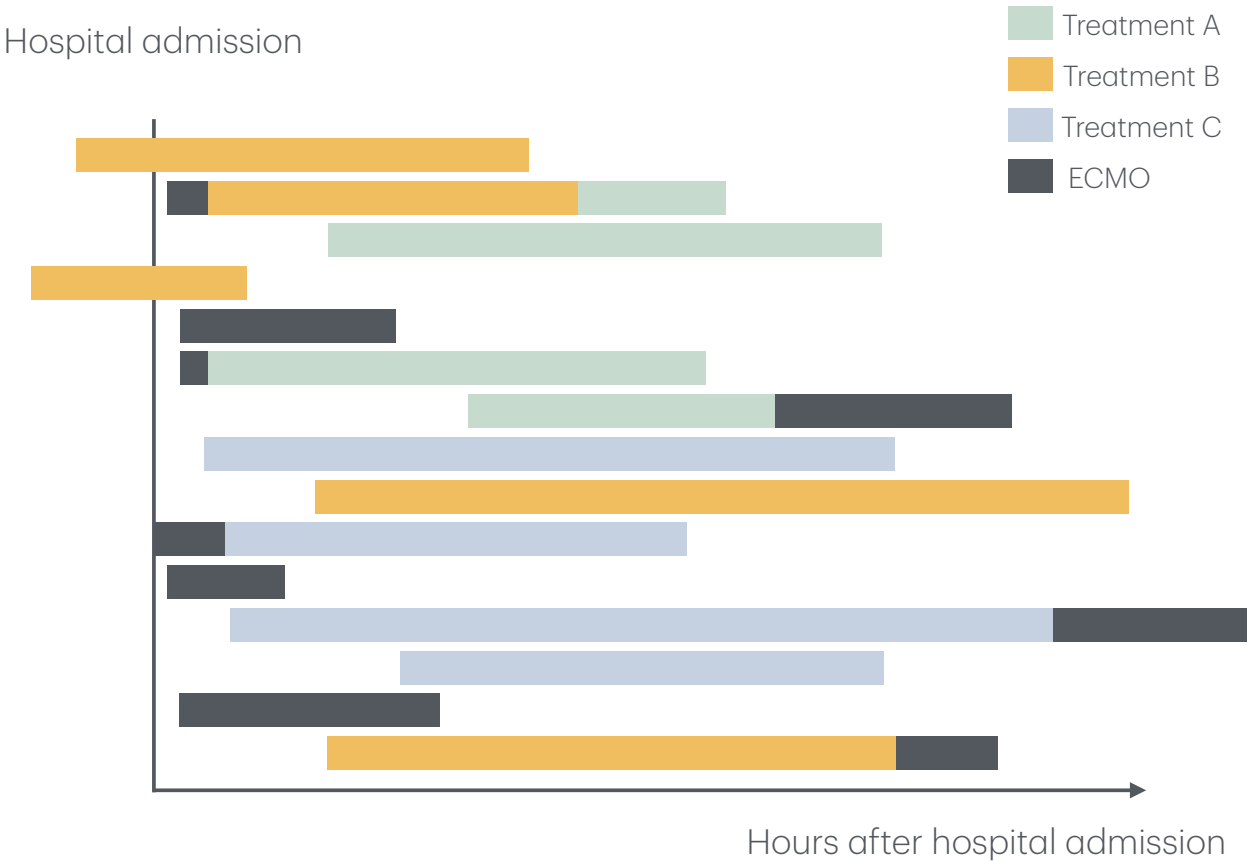
ORIGINAL

Management of high-risk acute pulmonary embolism: an emulated target trial analysis

Andrea Stadlbauer[1], Tom Verbelen[2], Leonhard Binzenhöfer[3], Tomaz Goslar[4], Alexander Supady[5], Peter M. Spieth[6], Marko Noc[4], Andreas Verstraete[2], Sabine Hoffmann[7], Michael Schomaker[41], Julia Höpler[7], Marie Kraft[7], Esther Tautz[5], Daniel Hoyer[8], Jörn Tongers[8], Franz Haertel[9], Aschraf El-Essawi[10], Mostafa Salem[11], Rafael Henrique Rangel[11], Carsten Hullermann[12], Marvin Kriz[13], Benedikt Schrage[13], Jorge Moisés[14], Manel Sabate[14], Federico Pappalardo[15], Lisa Crusius[16], Norman Mangner[16], Christoph Adler[17], Tobias Tichelbäcker[17], Carsten Skurk[18], Christian Jung[19], Sebastian Kufner[20], Tobias Graf[21], Clemens Scherer[3], Laura Villegas Sierra[3], Hannah Billig[22], Nicolas Majunke[23], Walter S. Speidl[24], Robert Zilberszac[24], Luis Chiscano-Camón[25], Aitor Uribarri[26], Jordi Riera[25], Roberto Roncon-Albuquerque Jr[27], Elizabete Terauda[28], Andrejs Erglis[28], Guido Tavazzi[29], Uwe Zeymer[30], Maike Knorr[31], Juliane Kilo[32], Sven Möbius-Winkler[9], Robert H. G. Schwinger[33], Derk Frank[11], Oliver Borst[34], Helene Häberle[35], Frederic De Roeck[36], Christiaan Vrints[36], Christof Schmid[1], Georg Nickenig[22], Christian Hagl[37], Steffen Massberg[3], Andreas Schäfer[38], Dirk Westermann[39], Sebastian Zimmer[22], Alain Combes[40], Daniele Camboni[1*], Holger Thiele[23] and Enzo Lüsebrink[22*] for the High-risk P. E. Investigator Group

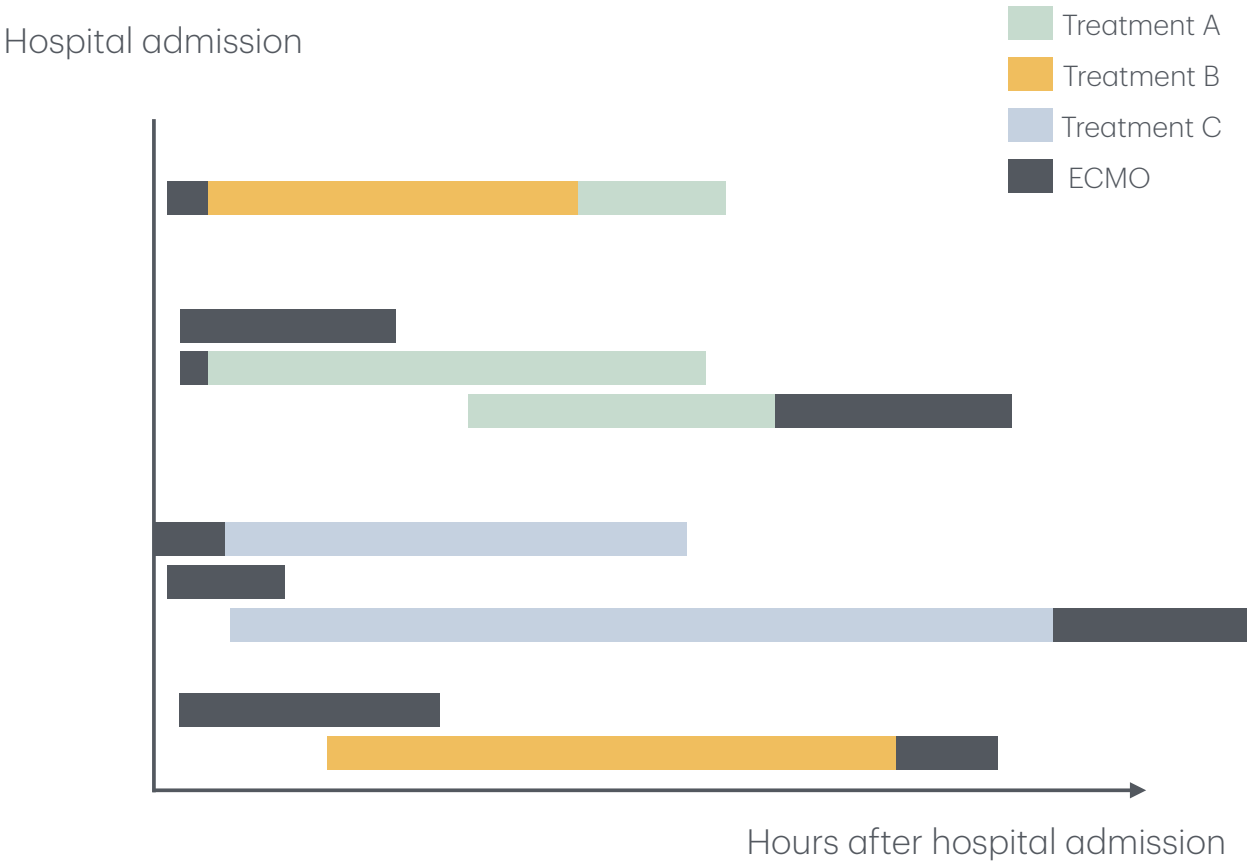
ID	A	B	C	ECMO	Mortality
1	0	1	0	0	1
2	1	1	0	1	0
3	1	0	0	0	1
4	0	1	0	0	1
5	0	0	0	1	1
6	1	0	0	1	0
7	1	0	0	1	1
8	0	0	1	0	1
9	0	1	0	0	0
10	0	0	1	1	0
11	0	0	0	1	1
12	0	0	1	1	1
13	0	0	1	0	0
14	0	0	0	1	1
15	0	1	0	1	1

Aspects related to treatment assignment and inclusion



ID	A	B	C	ECMO	Mortality
1	0	1	0	0	1
2	1	1	0	1	0
3	1	0	0	0	1
4	0	1	0	0	1
5	0	0	0	1	1
6	1	0	0	1	0
7	1	0	0	1	1
8	0	0	1	0	1
9	0	1	0	0	0
10	0	0	1	1	0
11	0	0	0	1	1
12	0	0	1	1	1
13	0	0	1	0	0
14	0	0	0	1	1
15	0	1	0	1	1

Aspects related to treatment assignment and inclusion



ID	A	B	C	ECMO	Mortality
1	0	1	0	0	1
2	1	1	0	1	0
3	1	0	0	0	1
4	0	1	0	0	1
5	0	0	0	1	1
6	1	0	0	1	0
7	1	0	0	1	1
8	0	0	1	0	1
9	0	1	0	0	0
10	0	0	1	1	0
11	0	0	0	1	1
12	0	0	1	1	1
13	0	0	1	0	0
14	0	0	0	1	1
15	0	1	0	1	1

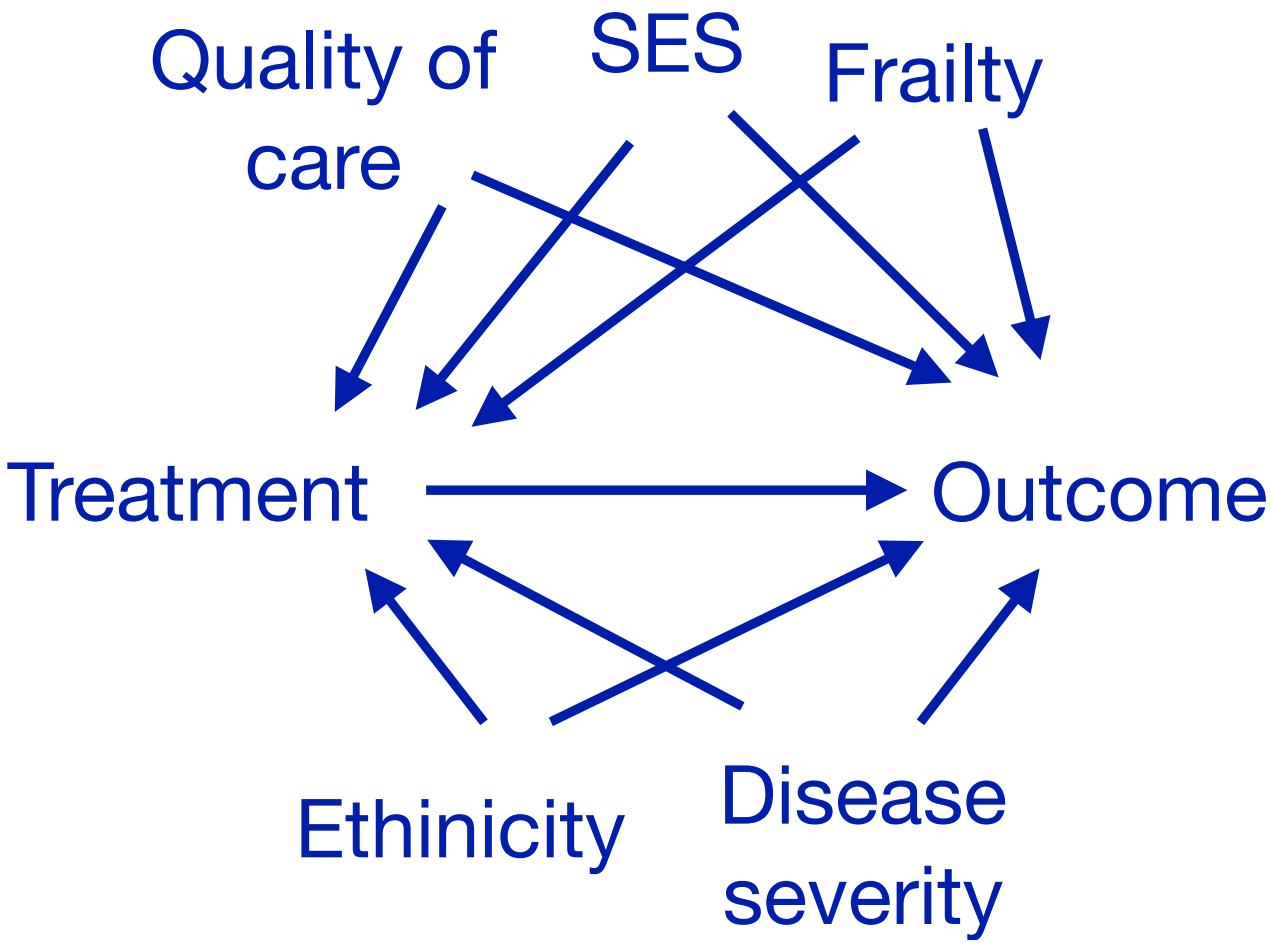
Interventions and tests are not random

Interventions and tests are not random

We often lack important information about which variables guide treatment selection

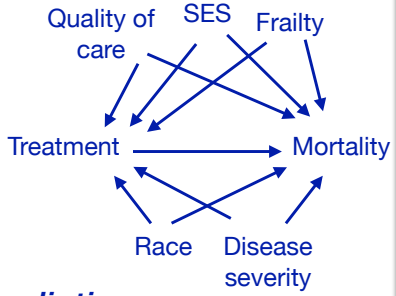
Interventions and tests are not random

We often lack important information about which variables guide treatment selection

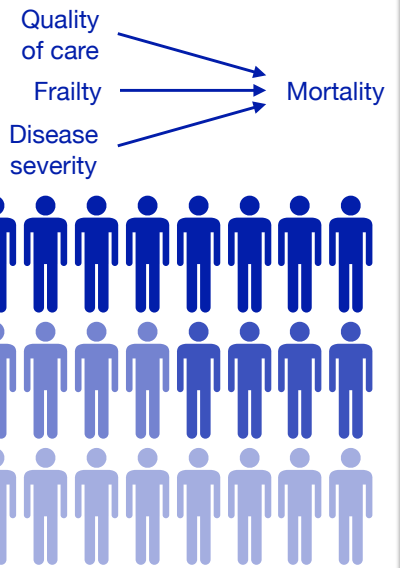


Target population

Explanation:



Prediction:



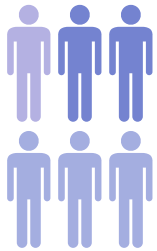
Hospital 1

- Disease severity overall high

- Treatment A, B and C available
- Quality of care overall high
- Immediate treatment initiation

- Frailty, SES and race unmeasured
- Blood values measured at hospital admission
- Central laboratory measurements for all patients
- Additional point-of-care blood gas analysis only for patients with high disease severity
- Advanced imagining techniques

Hospital 2



- Disease severity overall low
- SES overall high

- Treatment A and B are available
- Quality of care overall low
- Delayed treatment initiation

- Frailty and SES unmeasured
- Blood values measured 8-15 hours after hospital admission
- Reporting of blood values, without knowledge of provenance from point-of-care blood gas analysis or central laboratory
- Basic imaging techniques

ID	Hospital	Treatment	BGA	CL	Race	Mortality
1	1	A				1
2	1	B				0
3	1	B				1
4	1	A				1
5	1	B				1
6	1	A				0
7	1	B				1
8	1	C				1
9	1	C				0
10	1	A				0
11	1	C				0
12	2	A			White	0
13	2	A			White	0
14	2	B			White	0
15	2	A			Black	0
16	2				White	1
17	2	B			Black	1
18	2	A			White	1
19	2	B			White	1
20	2				Black	1
21	2	B			White	1
22	2	B			White	1
23	2				Black	1

Addressing confounding by indication

- It is important to identify measured, unmeasured and potentially mismeasured confounders in a directed acyclic graph (DAG)

Addressing confounding by indication

- It is important to identify measured, unmeasured and potentially mismeasured confounders in a directed acyclic graph (DAG)
- Ideally, these DAGs should be informed by interviews with clinicians or by cases with detailed information on all potential confounders to understand the rationale for treatment decisions

Addressing confounding by indication

- It is important to identify measured, unmeasured and potentially mismeasured confounders in a directed acyclic graph (DAG)
- Ideally, these DAGs should be informed by interviews with clinicians or by cases with detailed information on all potential confounders to understand the rationale for treatment decisions
- Statistical techniques to address confounding have often been found to be insufficient to eliminate this bias[24,25,26,27,28].

Addressing confounding by indication

- It is important to identify measured, unmeasured and potentially mismeasured confounders in a directed acyclic graph (DAG)
- Ideally, these DAGs should be informed by interviews with clinicians or by cases with detailed information on all potential confounders to understand the rationale for treatment decisions
- Statistical techniques to address confounding have often been found to be insufficient to eliminate this bias[24,25,26,27,28].
- Bosco et al. (2010), for instance, found that no adjustment method resolved confounding by indication when comparing breast cancer recurrence among women treated with adjuvant chemotherapy with women who did not receive adjuvant chemotherapy[24]

Addressing confounding by indication

- It is important to identify measured, unmeasured and potentially mismeasured confounders in a directed acyclic graph (DAG)
- Ideally, these DAGs should be informed by interviews with clinicians or by cases with detailed information on all potential confounders to understand the rationale for treatment decisions
- Statistical techniques to address confounding have often been found to be insufficient to eliminate this bias[24,25,26,27,28].
- Bosco et al. (2010), for instance, found that no adjustment method resolved confounding by indication when comparing breast cancer recurrence among women treated with adjuvant chemotherapy with women who did not receive adjuvant chemotherapy[24]
- Quantitative bias analysis, instrumental variables approaches and falsification endpoints consisting of negative controls can be used to address unmeasured and mismeasured confounders

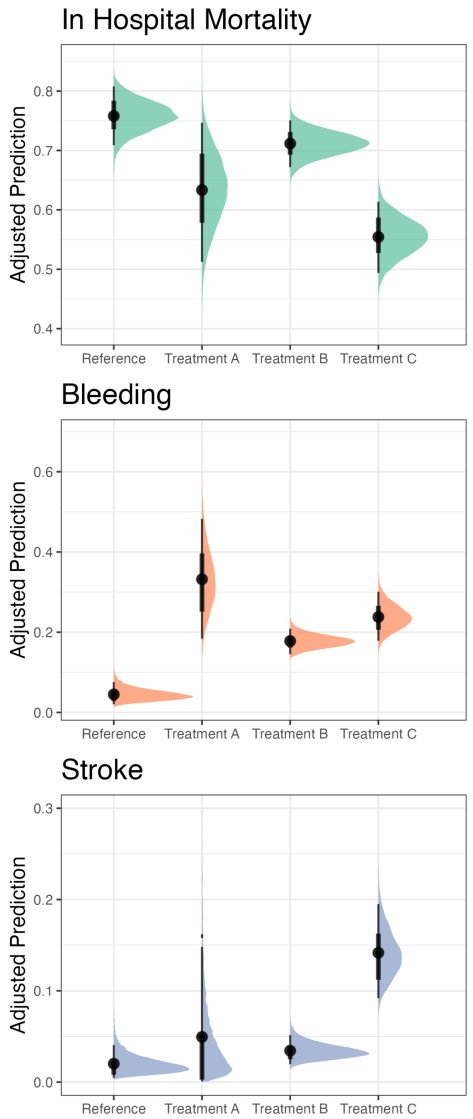
Confounding influences primary and secondary outcomes

Results of Logistic Regression Model

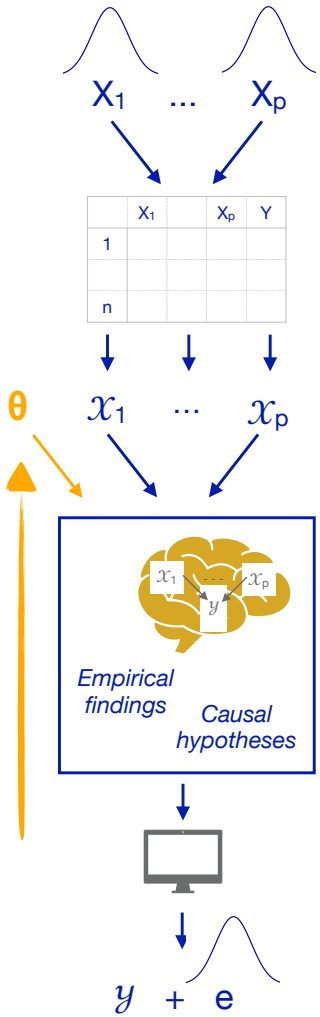
	Reference	Treatment A	Treatment B	Treatment C
OR [95% CI]	1	0.49 [0.24, 0.99]	0.75 [0.5, 1.11]	0.32 [0.21, 0.5]

Raw Prevalences

	Reference	Treatment A	Treatment B	Treatment C
In Hospital Mortality	74.58%	36.17%	77.64%	46.76%
Bleeding	3.81%	21.28%	20.76%	18.52%
Stroke	1.69%	2.13%	3.59%	12.04%



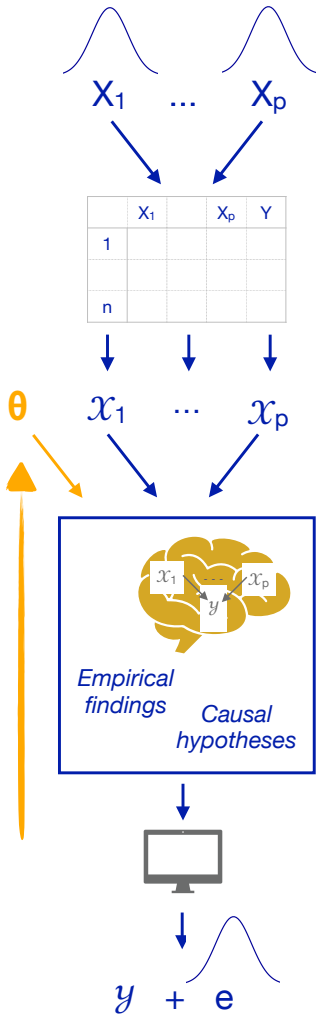
Multiplicity of possible analysis strategies



Measurement uncertainty

- *Missing data*
- *Differential and non-differential measurement error*
- *Measurement heterogeneity across centres*

Hoffmann, S., F. Schönbrodt, R. Elsas, R. Wilson, U. Strasser, Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science 8 201925



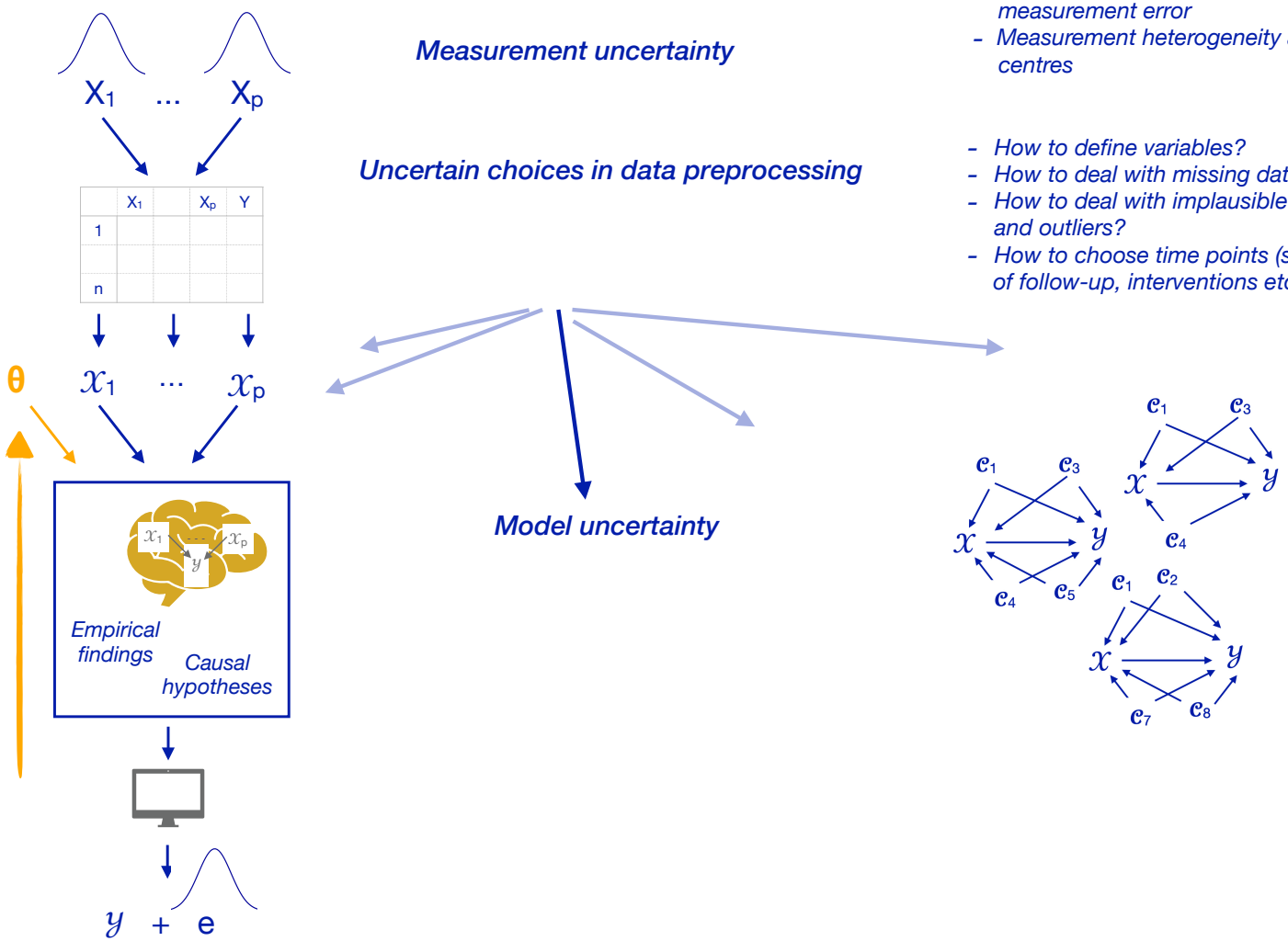
Measurement uncertainty

- *Missing data*
- *Differential and non-differential measurement error*
- *Measurement heterogeneity across centres*

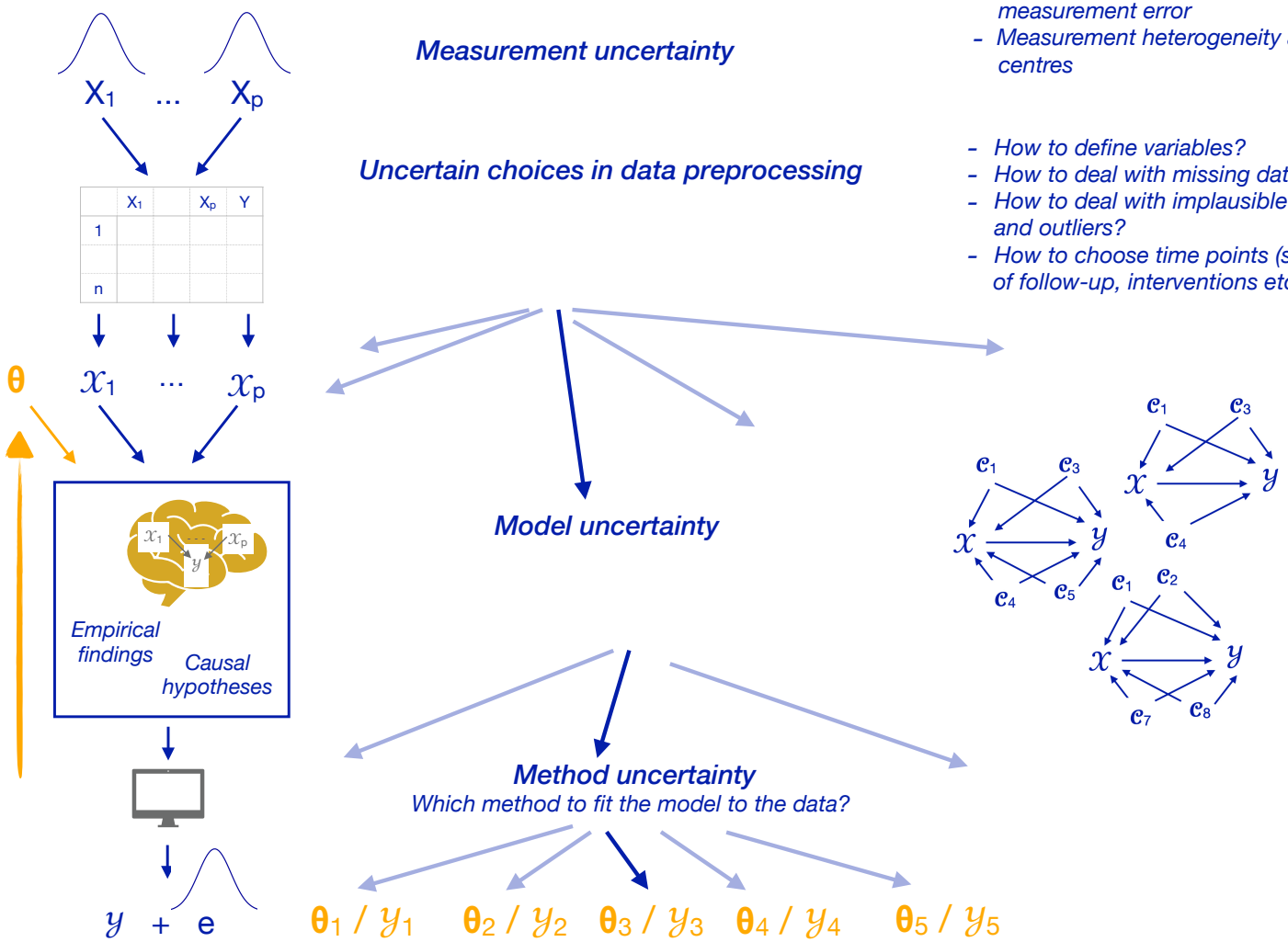
Uncertain choices in data preprocessing

- *How to define variables?*
- *How to deal with missing data?*
- *How to deal with implausible values and outliers?*
- *How to choose time points (start of follow-up, interventions etc.)*

Hoffmann, S., F. Schönbrodt, R. Elsas, R. Wilson, U. Strasser, Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science 8 201925

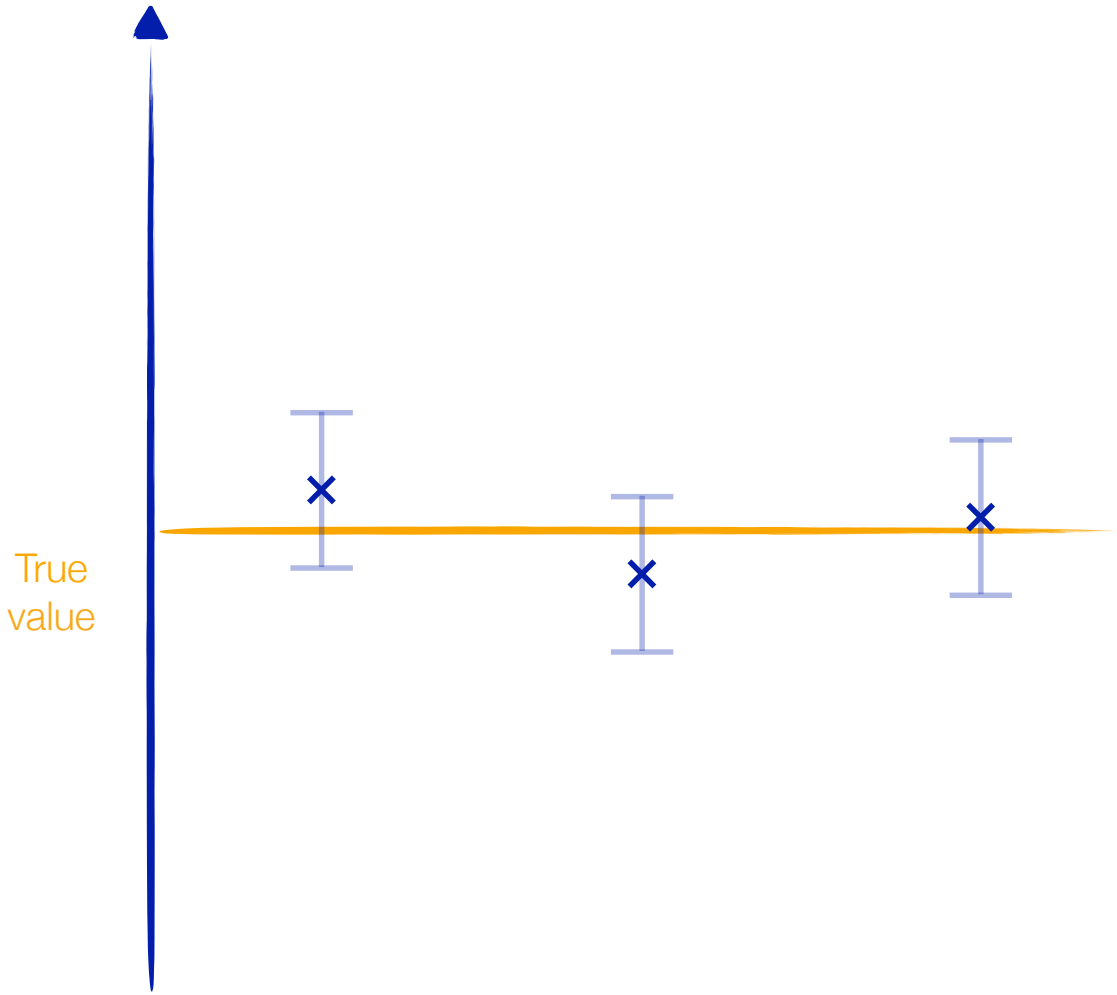


Hoffmann, S., F. Schönbrodt, R. Elsas, R. Wilson, U. Strasser, Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science 8 201925



Hoffmann, S., F. Schönbrodt, R. Elsas, R. Wilson, U. Strasser, Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science 8 201925

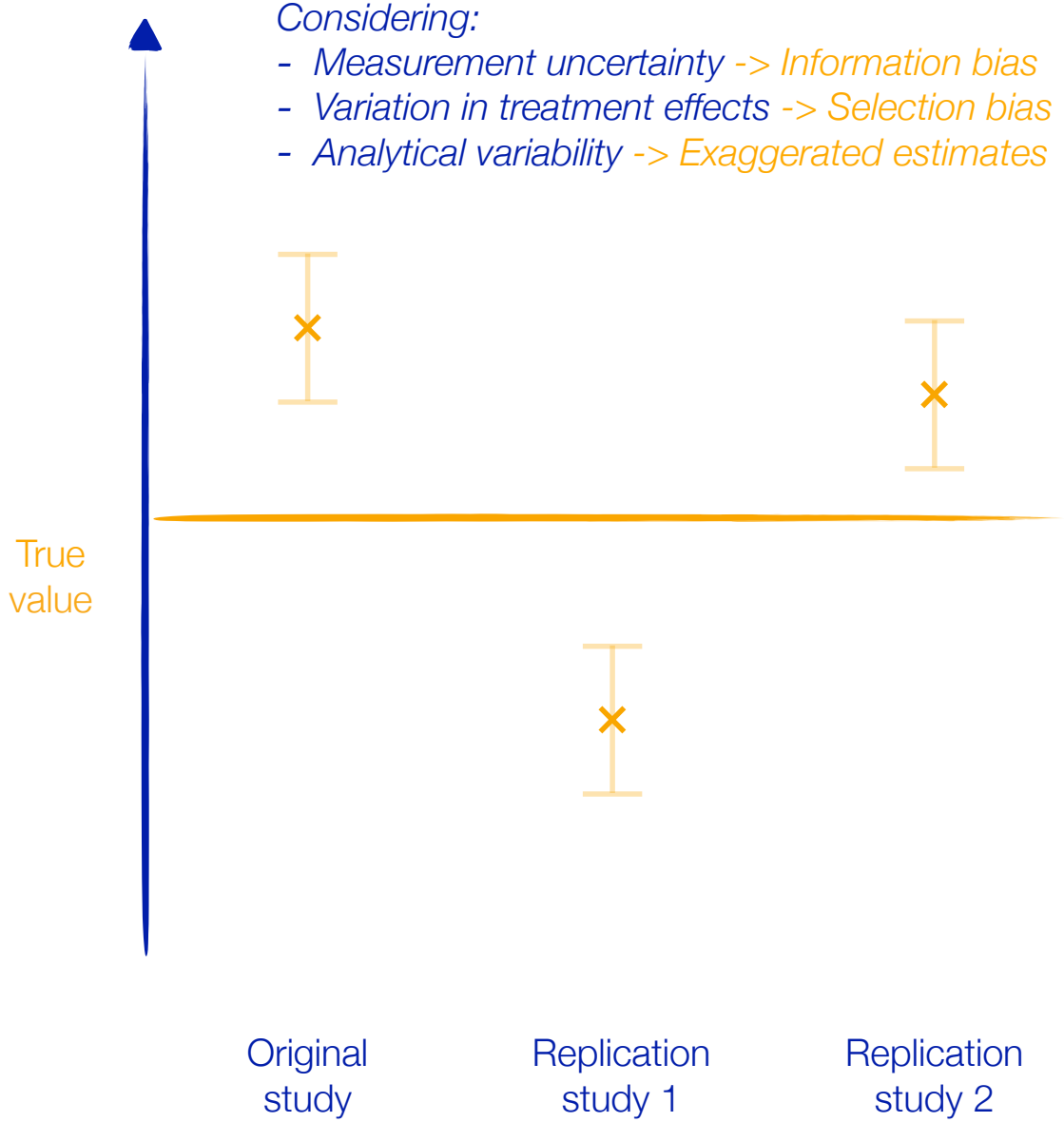
If we ignore uncertainty, it leads to bias



If we ignore uncertainty, it leads to bias

If we ignore uncertainty, it leads to bias

If we ignore uncertainty, it leads to bias



Publication bias and questionable research practices

Received: 16 November 2020 | Revised: 7 March 2021 | Accepted: 9 March 2021

DOI: 10.1111/stan.12241

ORIGINAL ARTICLE

WILEY

The significance filter, the winner's curse and the need to shrink

Erik W. van Zwet[1] | **Eric A. Cator**[2]

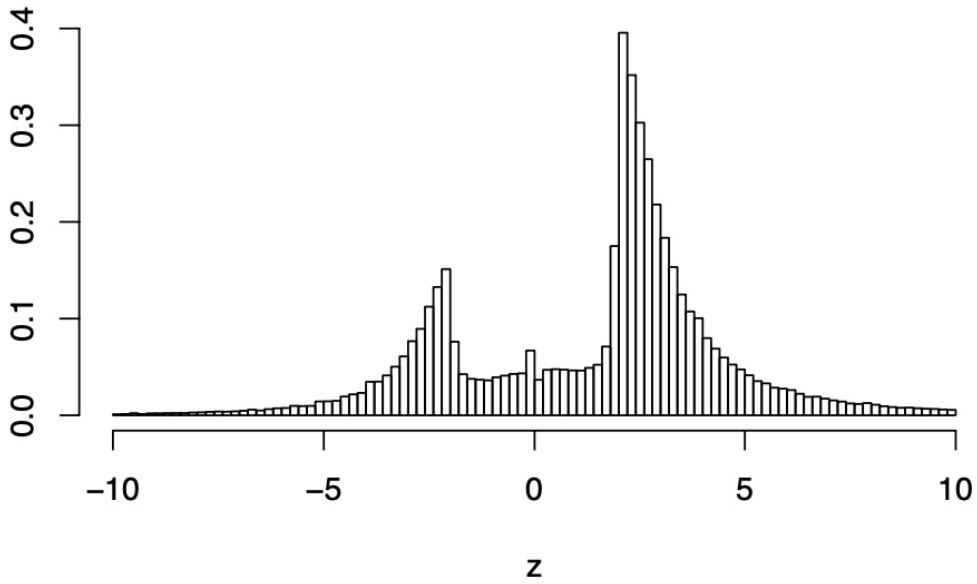


Figure 1: The distribution of more than one million z-values from Medline (1976–2019).

Questionable research practices as a continuum

Conflicting evidence

Researchers may address the same research question on the same data set, but use different inclusion and exclusion criteria, outcome measures, sample sizes, covariates, and operationalization of covariates

Conflicting evidence

Researchers may address the same research question on the same data set, but use different inclusion and exclusion criteria, outcome measures, sample sizes, covariates, and operationalization of covariates [29,30,31]

Surgery 165 (2019) 953–957

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Colon/Rectum

Does retrieval bag use during laparoscopic appendectomy reduce postoperative infection?

Adam C. Fields, MD[a,*], Pamela Lu, MD[a,b], Deanna L. Palenzuela, BS[a], Ronald Bleday, MD[a], Joel E. Goldberg, MD, MPH[a], Jennifer Irani, MD[a], Jennifer S. Davids, MD[c], Nelya Melnitchouk, MD, MSc[a,b,*]

[a] Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[b] Center for Surgery and Public Health, Department of Surgery, Brigha
[c] Department of Surgery, University of Massachusetts Memorial Medic

Surgery 165 (2019) 1199–1202

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Presented at the Academic Surgical Congress 2019

Utilization of a specimen retrieval bag during laparoscopic appendectomy for both uncomplicated and complicated appendicitis is not associated with a decrease in postoperative surgical site infection rates

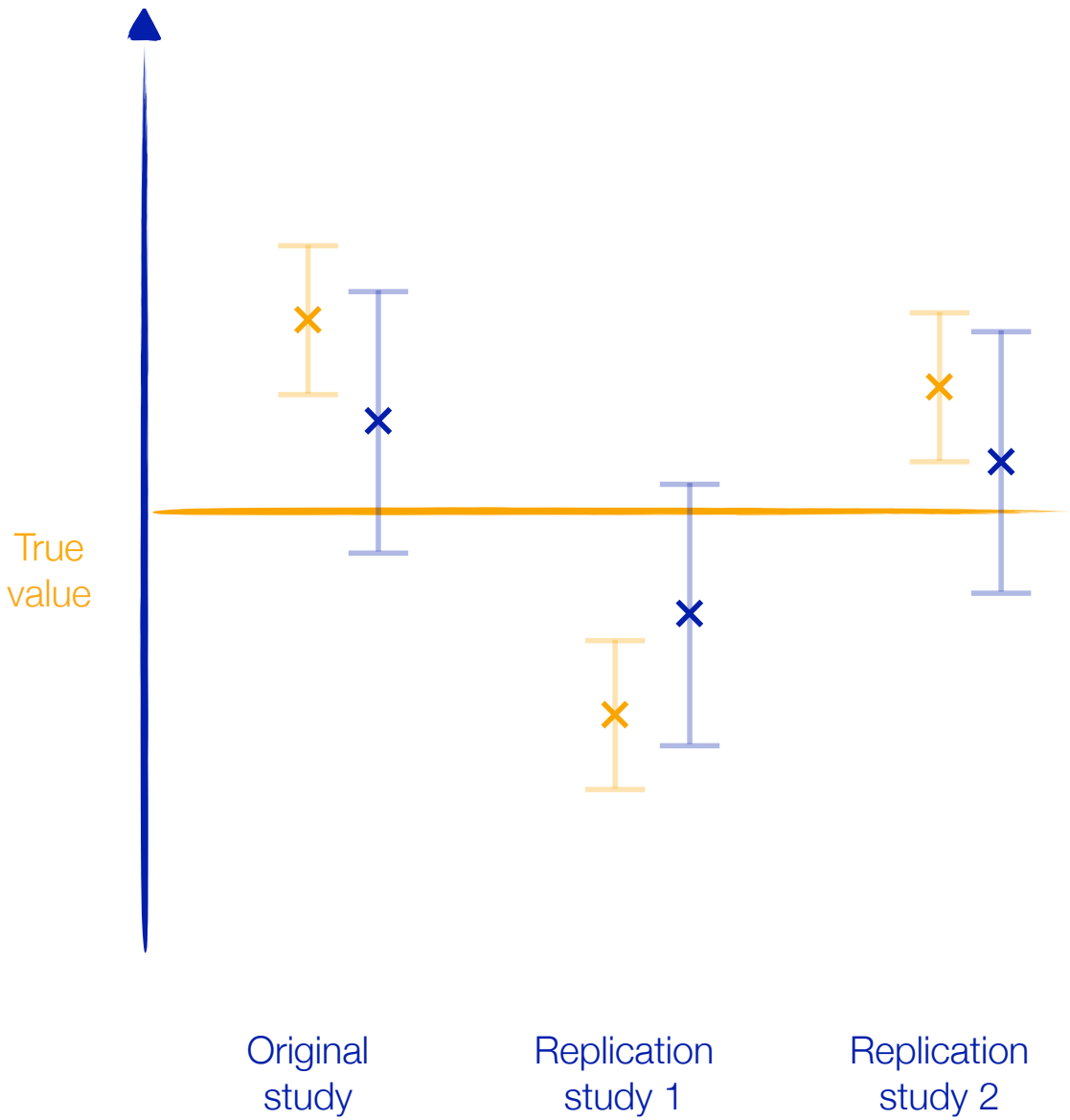
Scott A. Turner, MD*, Hee Soo Jung, MD, FACS, John E. Scarborough, MD, FACS

adison, WI

Table. Comparison of 2 Studies on the Association of a Specimen Retrieval Bag With Surgical Site Infection Rates in Laparoscopic Appendectomy

	Source	
Criteria	Fields et al,[3] 2019	Turner et al,[4] 2019
Inclusion criteria	*CPT* code 44970, not missing information on intra-abdominal abscess	Laparoscopic appendectomy, pathology with acute appendicitis, no additional major procedure
Analytic sample reported, No.	11 475	10 357
Primary outcome	Postoperative intra-abdominal abscess	Any SSI (superficial, deep, organ space)
Primary predictor	Use of retrieval bag	Use of retrieval bag
Covariates included, with operationalization	Age (continuous)	Age (dichotomized at 65 y)
	Sex (dichotomized)	Sex (dichotomized)
	BMI (continuous)	Obesity (categorical: not obese, class I/II/III obesity, missing)
	Race (categorized as: White, Black, Asian, other)	Not included
	Diabetes (dichotomized)	Diabetes (dichotomized)
	Hypertension (dichotomized)	Not included
	COPD (dichotomized)	Not included
	Smoker (dichotomized)	Not included
	Functional status (dichotomized)	Not included
	Steroid use (dichotomized)	Steroid use (dichotomized)
	Weight loss (dichotomized)	Not included
	Preoperative sepsis (dichotomized)	Unclear if included
	Wound class 3/4 (dichotomized)	Not included
	Complicated appendicitis (dichotomized)	2 Indicator variables: presence of abscess and presence of perforation
	ASA class 3/4 (dichotomized)	Not included
	Operative time (continuous)	Operative time dichotomized at 75th percentile
	White blood cell count (continuous)	Not included
Coefficient on primary predictor	OR (95% CI): 0.6 (0.42-0.95) *P* value: .03	OR (95% CI): 1.15 (0.78-1.69) *P* value: .49

Big data paradoxes[32]



Big data paradoxes[32]

Addressing multiplicity: Pre-registration

ANALYSIS

1 CHU Rennes, Inserm, Institut de Recherche en Santé, Environnement et Travail-UMR_S 1085, University of Rennes, Rennes, France

2 Institut Universitaire de France, Paris, France

3 Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA

4 Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

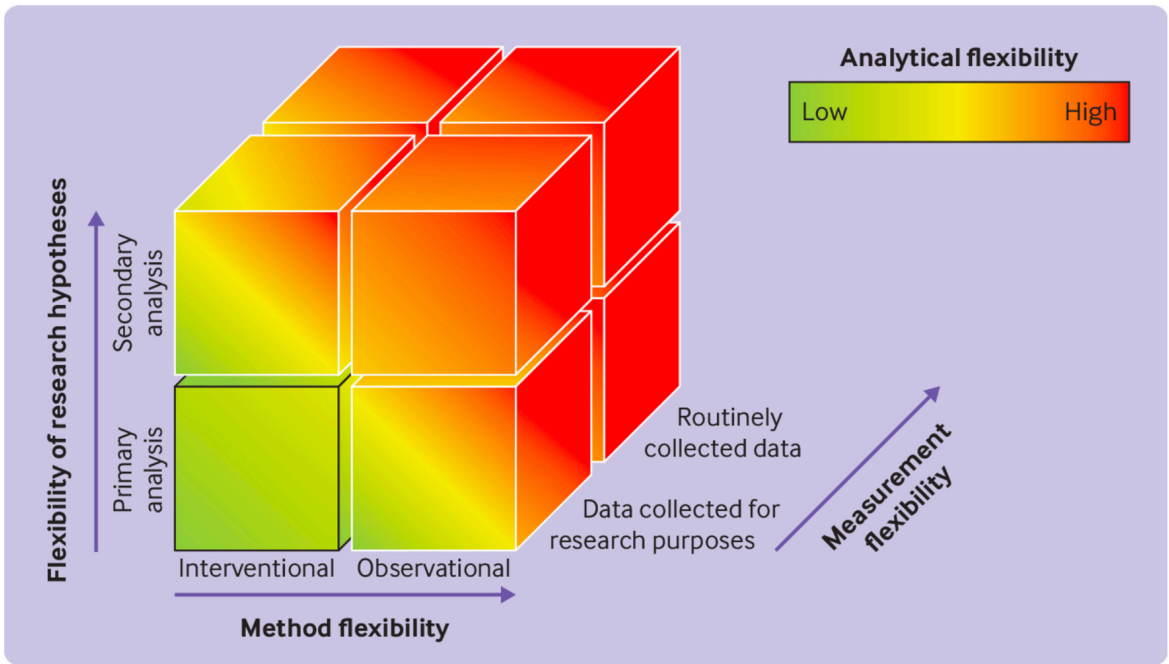
Improving the transparency and reliability of observational studies through registration

Florian Naudet and colleagues argue that routine registration of observational research is needed and suggest how current processes can be adapted to facilitate it

Florian Naudet, [1,2] Chirag J Patel, [3] Nicholas J DeVito, [4] Gérard Le Goff, [5] Ioana A Cristea, [6] Alain Braillon, [7] Sabine Hoffmann [8,9]

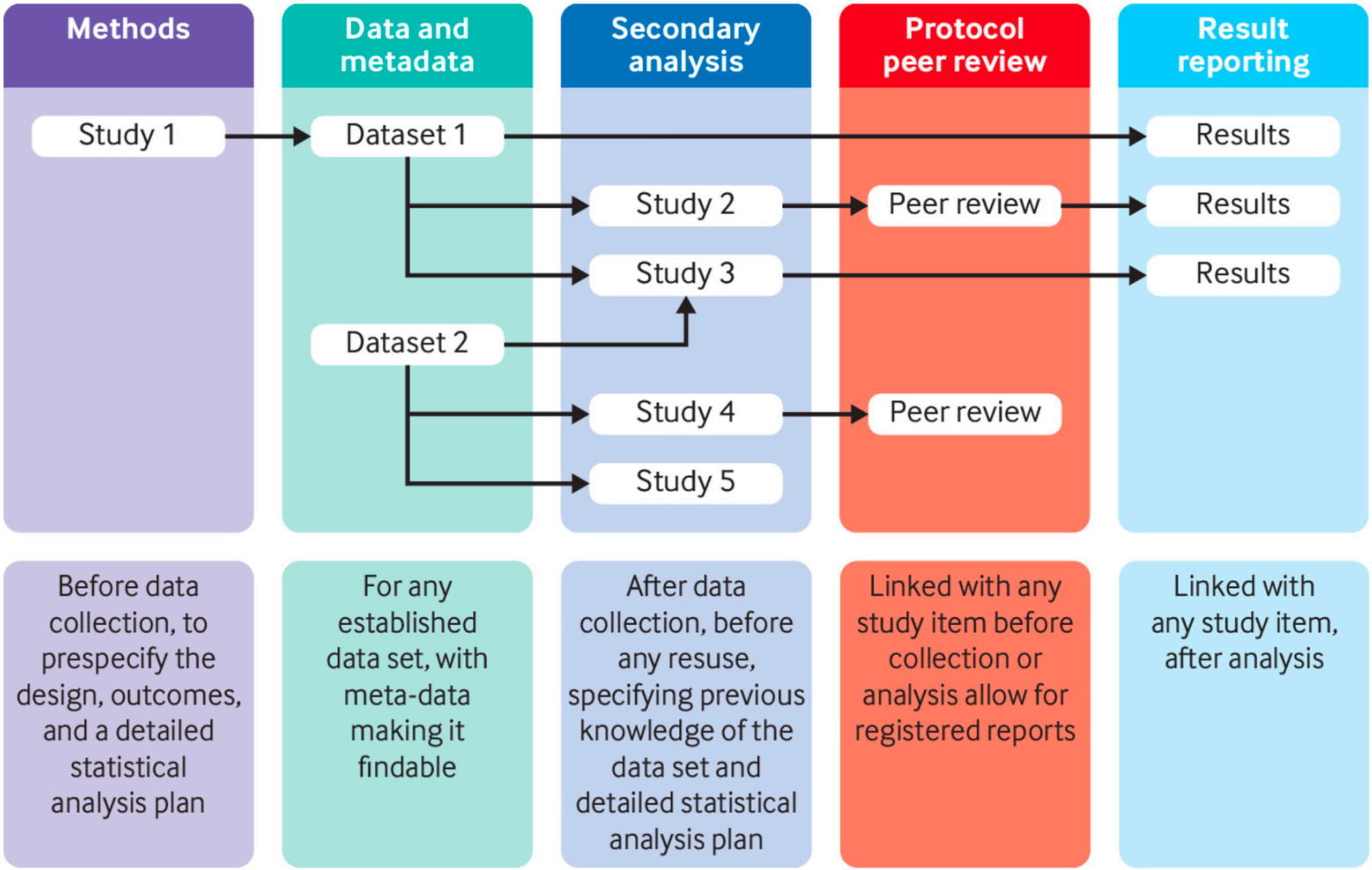
From the use of booster doses against covid-19[1] to approval of novel cancer therapies,[2] health

This already happens for certain studies, such as non-interventional post-authorisation safety studies



BMJ: first published as 10.1136/bmj-2023-076123

Addressing multiplicity: Pre-registration



Addressing multiplicity: Multi-analyst studies

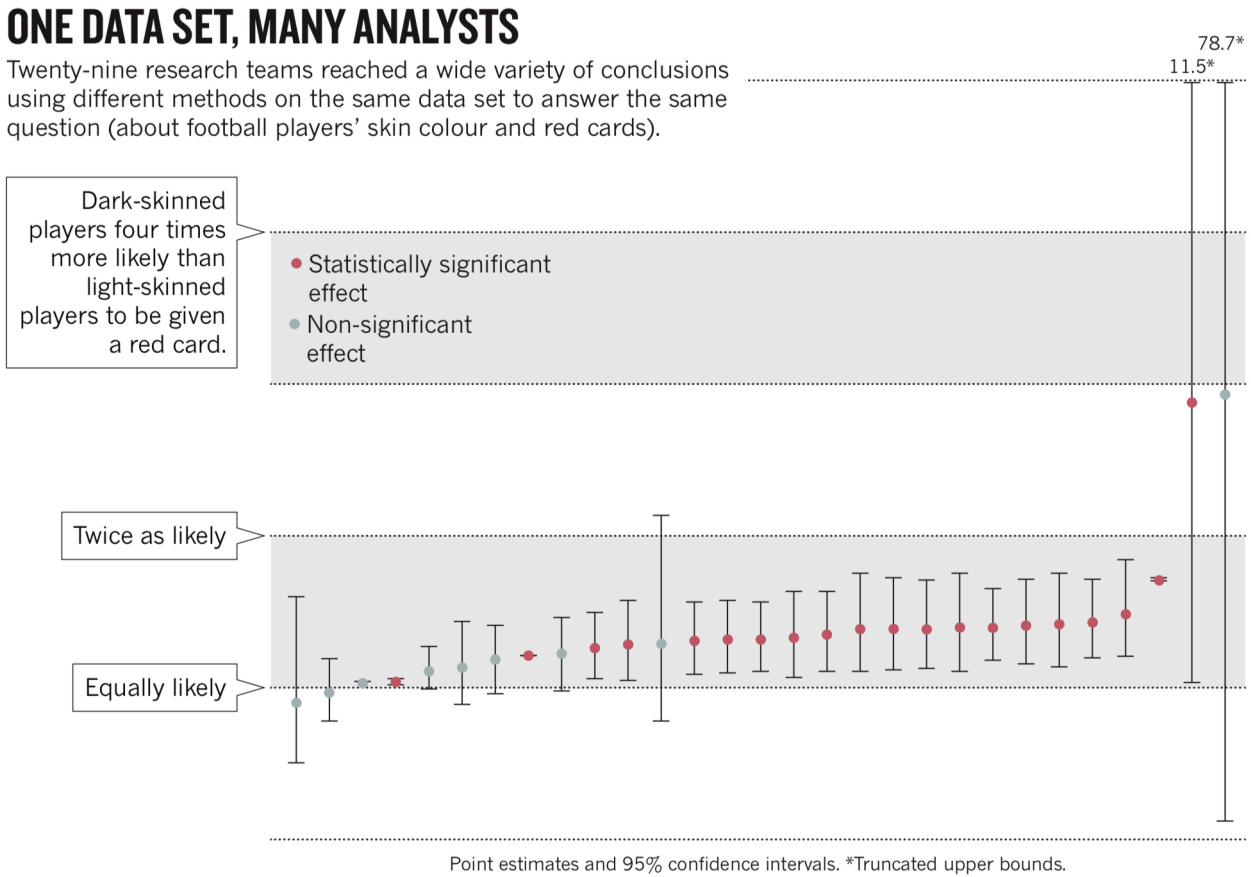
Are football referees more likely to give red cards to players with dark skin than to players with light skin?[33]

MICHAEL REGAN/GETTY

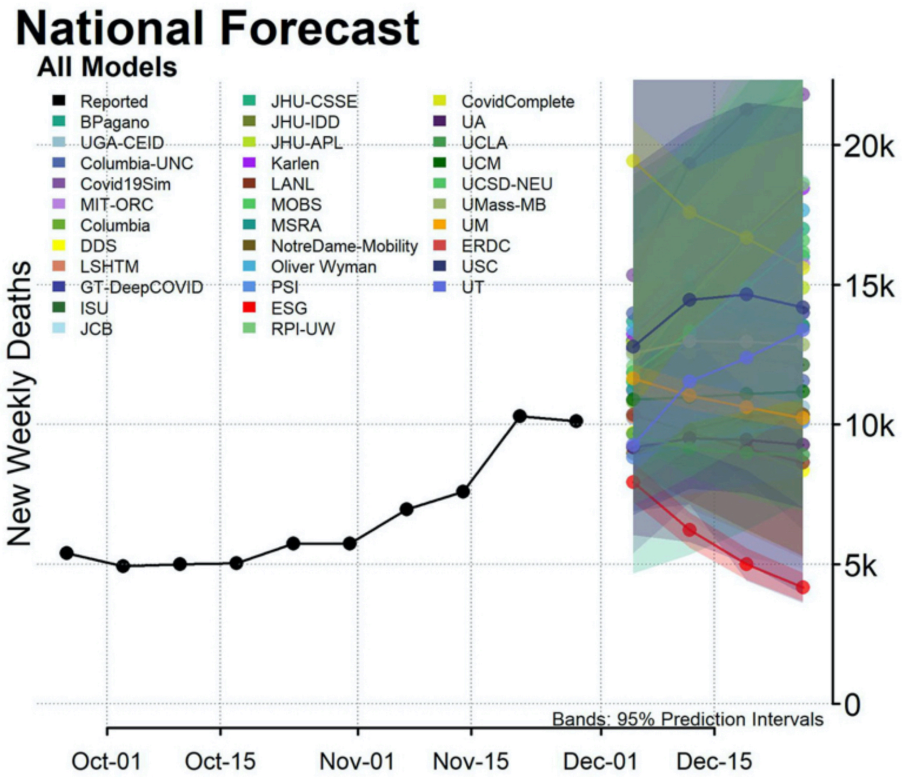
Mario Balotelli, playing for Manchester City, is shown a red card during a match against Arsenal.

Addressing multiplicity: Multi-analyst studies

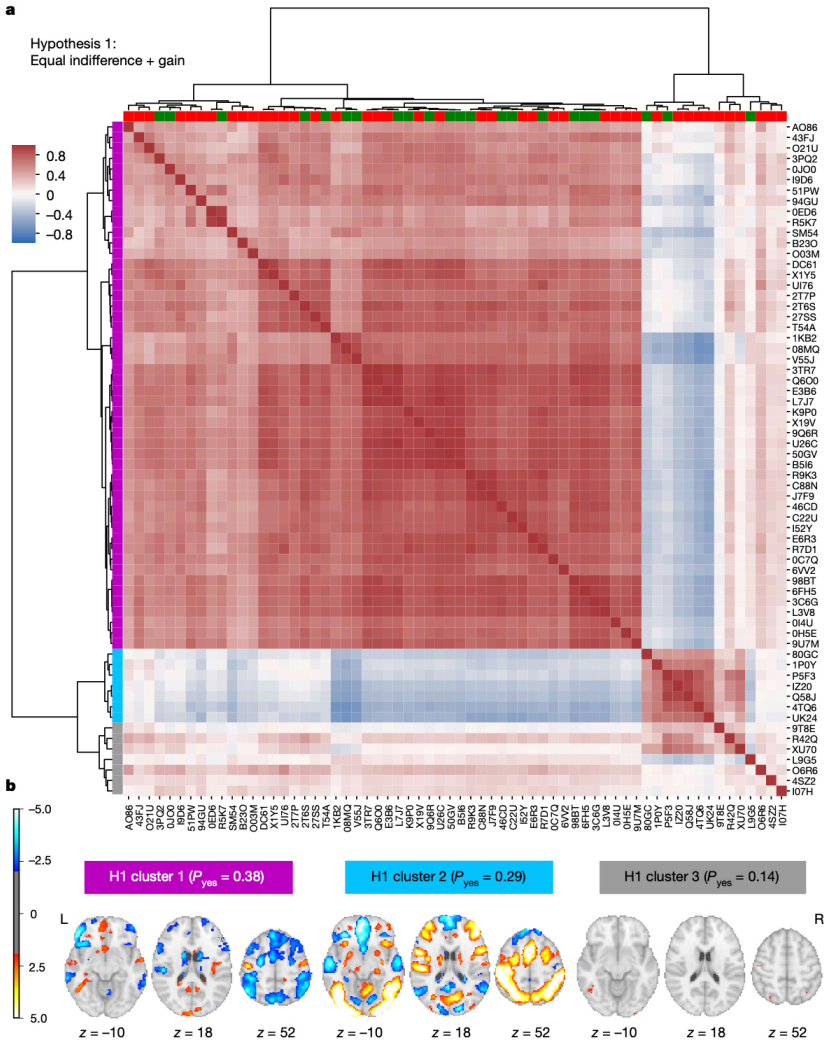
Are football referees more likely to give red cards to players with dark skin than to players with light skin?[33]



Addressing multiplicity: Multi-analyst studies



COVID-19 modelling



[34]

Addressing multiplicity: Vibration of effects

Meta-Psychology, 2023, vol 7, MP.2020.2556
https://doi.org/10.15626/MP.2020.2556
Article type: Original Article
Published under the CC-BY4.0 license

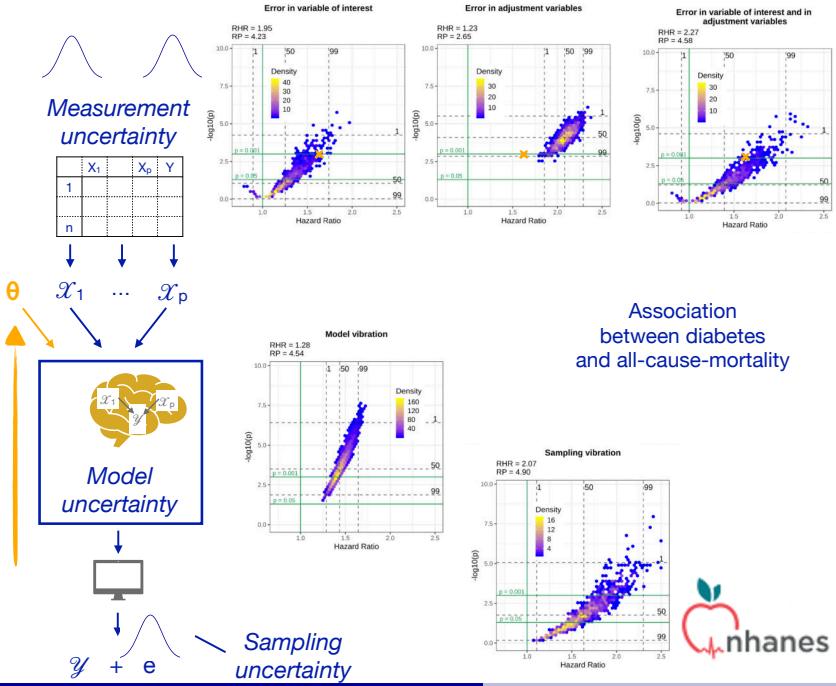
Open data: Not Applicable
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Danielsson, H., Carlsson, R.
Reviewed by: Parsons, S., Young, C., Antonoplis, S.
Analysis reproduced by: Batinović, L., Tobias, C.
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/8MZV3

Comparing the vibration of effects due to model, data pre-processing, and sampling uncertainty on a large data set in personality psychology

Simon Klau[1,2], Felix D. Schönbrodt[3,4], Chirag J. Patel[5], John P.A. Ioannidis[6,7,8,9], Anne-Laure Boulesteix[1,4], and Sabine Hoffmann[1,4]

[1]Institute for Medical Information Processing, Biometry, and Epidemiology, Munich, Germany
[2]Leibniz-Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany
[3]Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany
[4]LMU Open Science Center, Ludwig-Maximilians-Universität München, Munich, Germany
[5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[6]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA
[7]Department of Epidemiology and Population Health, Stanford University, Stanford, CA, USA
[8]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
[9]Department of Statistics, Stanford University, Stanford, CA, USA



IEA International Epidemiological Association

International Journal of Epidemiology, 2021, 266–278
doi: 10.1093/ije/dyaa164
Advance Access Publication Date: 5 November 2020
Original article

Methods

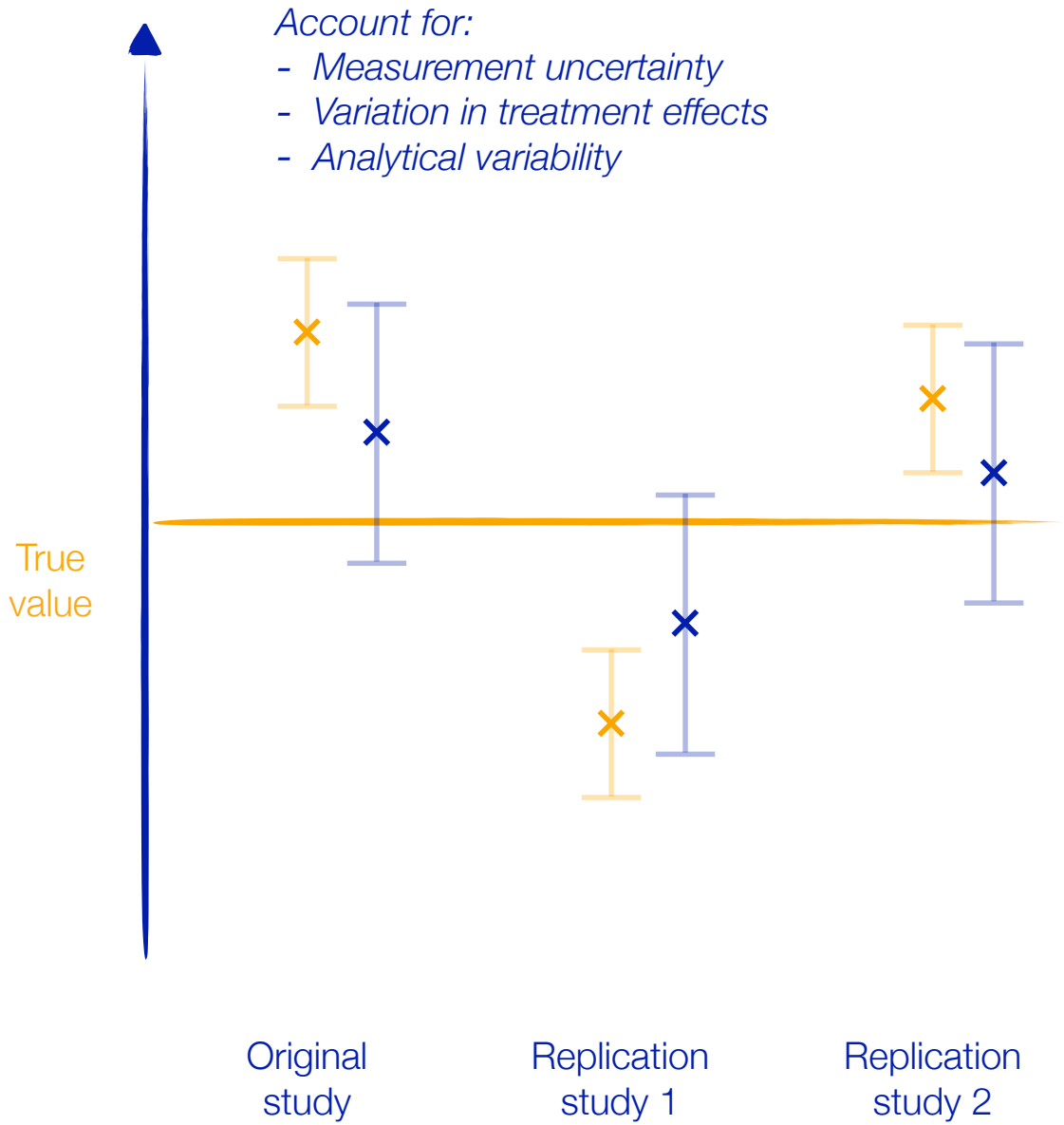
Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework

Simon Klau, [1,2,*,†] Sabine Hoffmann,[1,3†] Chirag J Patel,[4] John PA Ioannidis,[5,6,7,8,9] and Anne-Laure Boulesteix[1,3]

[1]Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany, [2]Leibniz Institute for Prevention Research and Epidemiology–BIPS, Bremen, Germany, [3]LMU Open Science Center, Ludwig-Maximilians-Universität München, Munich, Germany, [4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA, [5]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA, [6]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA, [7]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA, [8]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA and [9]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

[†]These authors contributed equally to this work.
*Corresponding author: Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-Universität München, Marchioninistr. 15, D-81377 Munich, Germany. E-mail: simon.klau@yahoo.de.

Addressing multiplicity: Account for analytical variability



How do flexible AI/ML algorithms change these challenges?

Eligibility

- Digitalization and AI may reduce health disparities by providing broader access to care and reducing costs

Eligibility

- Digitalization and AI may reduce health disparities by providing broader access to care and reducing costs
- The black box nature of algorithms, which may only partly be resolved through explainable and interpretable AI, makes it difficult to detect biases and to evaluate generalisability.

Eligibility

- Digitalization and AI may reduce health disparities by providing broader access to care and reducing costs
- The black box nature of algorithms, which may only partly be resolved through explainable and interpretable AI, makes it difficult to detect biases and to evaluate generalisability.
- Transfer learning and modern foundational models may facilitate the generalization to other settings.

Data quality

- Digital health technology, including sensors, wearable remote monitoring and AI assistance in data collection, extraction, cleaning and validation may improve quality

Data quality

- Digital health technology, including sensors, wearable remote monitoring and AI assistance in data collection, extraction, cleaning and validation may improve quality
- Flexible AI algorithms may be more vulnerable to poor data quality and make so-called shortcuts where they use spurious features for prediction rather than to detect clinically meaningful differences

Data quality

- Digital health technology, including sensors, wearable remote monitoring and AI assistance in data collection, extraction, cleaning and validation may improve quality
- Flexible AI algorithms may be more vulnerable to poor data quality and make so-called shortcuts where they use spurious features for prediction rather than to detect clinically meaningful differences
- Spurious associations may be difficult to detect when the interpretability of results is limited and when models are trained across distributed data sets in federated learning

Data quality

- Digital health technology, including sensors, wearable remote monitoring and AI assistance in data collection, extraction, cleaning and validation may improve quality
- Flexible AI algorithms may be more vulnerable to poor data quality and make so-called shortcuts where they use spurious features for prediction rather than to detect clinically meaningful differences
- Spurious associations may be difficult to detect when the interpretability of results is limited and when models are trained across distributed data sets in federated learning
- Deliberate random noise injection can increase robustness and generalizability

Time point alignment

- When modern algorithms are used to analyse large, heterogenous data sets without filtering based on clinical expertise, it is difficult to prevent problems arising from time point alignment.

Time point alignment

- When modern algorithms are used to analyse large, heterogenous data sets without filtering based on clinical expertise, it is difficult to prevent problems arising from time point alignment.
- Overoptimism due to problems in time point alignment may be difficult to detect when the interpretability of results is limited.

Interventions and tests are not random

- Causal machine learning may offer robustness to model misspecification, requiring less modelling decisions and capturing non-linear associations and interactions

Interventions and tests are not random

- Causal machine learning may offer robustness to model misspecification, requiring less modelling decisions and capturing non-linear associations and interactions
- When flexible algorithms are trained on clinician-initiated data, they may show strong predictive performance that only expresses the actions of clinicians

Interventions and tests are not random

- Causal machine learning may offer robustness to model misspecification, requiring less modelling decisions and capturing non-linear associations and interactions
- When flexible algorithms are trained on clinician-initiated data, they may show strong predictive performance that only expresses the actions of clinicians
- Flexible algorithms may learn and reproduce differential treatments and diagnoses based on socio-demographic factors and thereby replicate and exacerbate existing biases in the data, including underdiagnosis bias for historically under-served populations

Interventions and tests are not random

- Causal machine learning may offer robustness to model misspecification, requiring less modelling decisions and capturing non-linear associations and interactions
- When flexible algorithms are trained on clinician-initiated data, they may show strong predictive performance that only expresses the actions of clinicians
- Flexible algorithms may learn and reproduce differential treatments and diagnoses based on socio-demographic factors and thereby replicate and exacerbate existing biases in the data, including underdiagnosis bias for historically under-served populations
- These problems may be more difficult to detect when the interpretability of results is limited.

Multiplicity of analysis strategies

- AI may exacerbate problems arising from the mulitplicity of possible analysis strategies on retrospective data by making it easier to generate research findings and by making it harder to discover questionable research practices and fraud
- The reproducibility of results may be limited by cost considerations and by a large number of parameters that may be poorly documented

Criteria to judge the quality and credibility of observational studies

Criteria to assure the quality of retrospective studies

	Minimal requirements	**Acceptable**	**Ideal approach**
Eligibility	Detailed comparison of patient characteristics with target population and across centres		
Data quality	Extensive data quality checks and reporting of data pre-processing and missing patterns		
Time point alignment	Report exact timing of all measurements and of treatment trajectories		
Interventions and tests not random	Directed acyclic graph including unmeasured and mismeasured confounders		
Multiplicity of analysis strategies	Pre-registration of statistical analysis plan		

Criteria to assure the quality of retrospective studies

	Minimal requirements	**Acceptable**	**Ideal approach**
Eligibility	Detailed comparison of patient characteristics with target population and across centres	Inverse probability weighting or multilevel regression modelling with post-stratification	
Data quality	Extensive data quality checks and reporting of data pre-processing and missing patterns	Audits, standardisation and training in data collection, validation data to quantify errors	
Time point alignment	Report exact timing of all measurements and of treatment trajectories	Target trial emulation	
Interventions and tests not random	Directed acyclic graph including unmeasured and mismeasured confounders	Quantitative bias analysis and falsification endpoints of negative controls	
Multiplicity of analysis strategies	Pre-registration of statistical analysis plan	Multi-analyst studies or extensive multiverse analysis to report robustness of results	

Criteria to assure the quality of retrospective studies

	Minimal requirements	**Acceptable**	**Ideal approach**
Eligibility	Detailed comparison of patient characteristics with target population and across centres	Inverse probability weighting or multilevel regression modelling with post-stratification	Combination of data with more representative data sources
Data quality	Extensive data quality checks and reporting of data pre-processing and missing patterns	Audits, standardisation and training in data collection, validation data to quantify errors	Account for complex measurement error and informative missing data
Time point alignment	Report exact timing of all measurements and of treatment trajectories	Target trial emulation	Validate accuracy of time stamps, report reasons for treatment switches
Interventions and tests not random	Directed acyclic graph including unmeasured and mismeasured confounders	Quantitative bias analysis and falsification endpoints of negative controls	Combined analysis with RCT, high quality documentation of treatment decisions
Multiplicity of analysis strategies	Pre-registration of statistical analysis plan	Multi-analyst studies or extensive multiverse analysis to report robustness of results	Uncertainty intervals that account for analytical variability

Conclusion

Conclusion

- Retrospective studies must meet high standards to provide meaningful insights and to genuinely provide benefit by advancing patient care

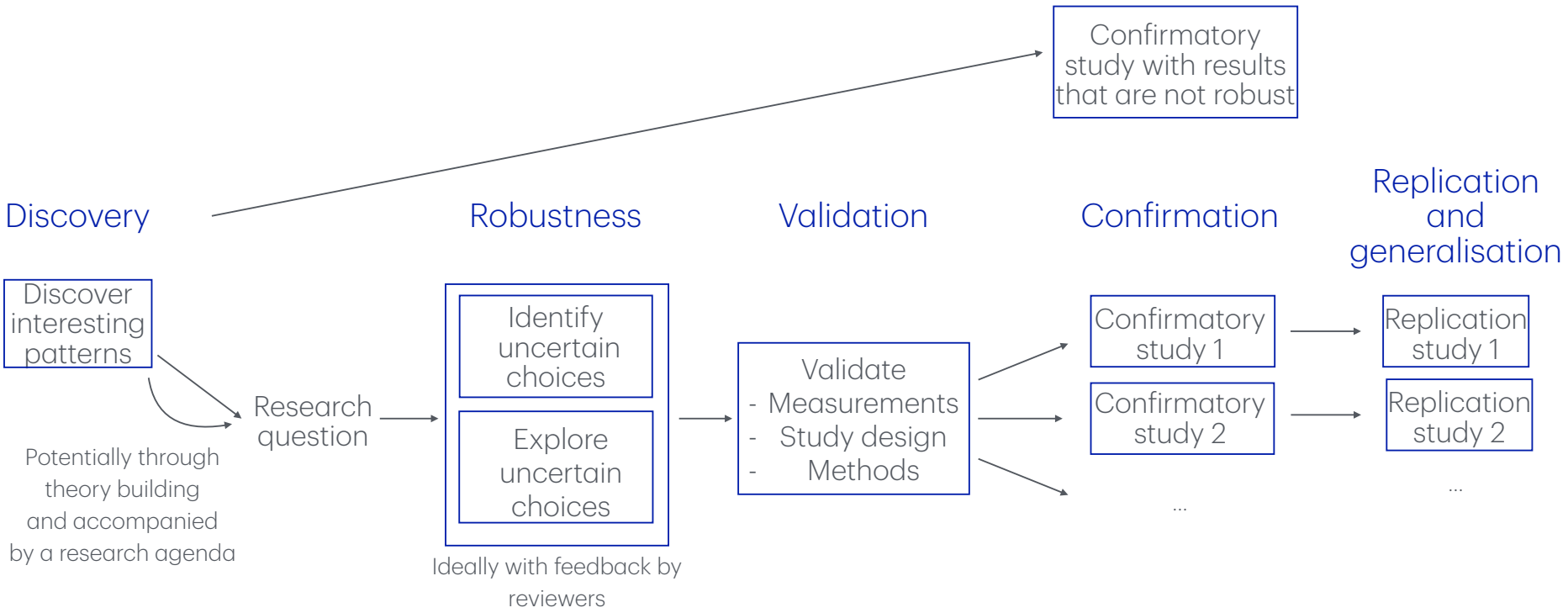
Conclusion

- Retrospective studies must meet high standards to provide meaningful insights and to genuinely provide benefit by advancing patient care
- Using a retrospective design is rarely straightforward and never automatic. Simply "doing the best we can" using retrospective data is often not good enough.

Conclusion

- Retrospective studies must meet high standards to provide meaningful insights and to genuinely provide benefit by advancing patient care
- Using a retrospective design is rarely straightforward and never automatic. Simply "doing the best we can" using retrospective data is often not good enough.
- Credible and reliable evidence from retrospective studies requires good knowledge of how the data are collected, ideally gained through extensive validation studies and interviews with clinicians to understand the rationale for treatment decisions and with the personnel collecting the data

Can retrospective studies be seen as more than hypothesis-generating?



Thank you for your attention!

[1] Benjamin A Goldstein, Nrupen A Bhavsar, Matthew Phelan, and Michael J Pencina. Controlling for informed presence bias due to the number of health encounters in an electronic health record. *American journal of epidemiology*, 184(11):847–855, 2016.

[2] Sarah Miller and Laura R Wherry. Health and access to care during the first 2 years of the aca medicaid expansions. *New England Journal of Medicine*, 376(10):947–956, 2017.

[3] Milena A Gianfrancesco and Neal D Goldstein. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC medical research methodology*, 21(1):234, 2021.

[4] Himabindu Reddy, Shiv Joshi, Abhishek Joshi, and Vasant Wagh. A critical review of global digital divide and the role of technology in healthcare. *Cureus*, 14(9), 2022.

[5] Joanna Harton, Nandita Mitra, and Rebecca A Hubbard. Informative presence bias in analyses of electronic health records-derived data: a cautionary note. *Journal of the American Medical Informatics Association*, 29(7):1191–1199, 2022.

[6] Ariana Mihan and Harriette GC Van Spall. Interventions to enhance digital health equity in cardiovascular care. *Nature Medicine*, 30(3):628–630, 2024.

[7] Ariana Mihan, Ambarish Pandey, and Harriette GC Van Spall. Mitigating the risk of artificial intelligence bias in cardiovascular care. *The Lancet Digital Health*, 6(10):e749–e754, 2024.

[8] Ban Al-Sahab, Alan Leviton, Tobias Loddenkemper, Nigel Paneth, and Bo Zhang. Biases in electronic health records data for generating real-world evidence: An overview. *Journal of Healthcare Informatics Research*, 8(1):121–139, 2024.

[9] Nicholas C Arpey, Anne H Gaglioti, and Marcy E Rosenbaum. How socioeconomic status affects patient perceptions of health care: a qualitative study. *Journal of primary care & community health*, 8(3):169–175, 2017.

[10] Judy H Ng, Faye Ye, Lauren M Ward, Samuel C,ÄúChris,Äù Haffer, and Sarah Hudson Scholle. Data on race, ethnicity, and language largely incomplete for managed care plan members. *Health Affairs*, 36(3):548–552, 2017.

[11] Milena A Gianfrancesco, Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk. Potential biases in machine learning algorithms using electronic health record data. *JAMA internal medicine*, 178(11):1544–1547, 2018.

[12] Marco Ramoni and Paola Sebastiani. Robust learning with missing data. *Machine Learning*, 45:147–170, 2001.

[13] Niels Keiding and Thomas A Louis. Perils and potentials of self-selected entry to epidemiological studies and surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2):319–376, 2016.

[14] Irina Degtiar and Sherri Rose. A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, 10(1):501–524, 2023.

[15] Marcus R Munafò, Kate Tilling, Amy E Taylor, David M Evans, and George Davey Smith. Collider scope: when selection bias can substantially influence observed associations. *International journal of epidemiology*, 47(1):226–235, 2018.

[16] Kaiwen Ni, Hongling Chu, Lin Zeng, Nan Li, and Yiming Zhao. Barriers and facilitators to data quality of electronic health records used for clinical research in china: a qualitative study. *BMJ open*, 9(7):e029314, 2019.

[17] Robert A Verheij, Vasa Curcin, Brendan C Delaney, and Mark M McGilchrist. Possible sources of bias in primary care electronic health record data use and reuse. *Journal of medical Internet research*, 20(5):e185, 2018.

[18] Sally Yaacoub, Raphael Porcher, Anna Pellat, Hillary Bonnet, Viet-Thi Tran, Philippe Ravaud, and Isabelle Boutron. Characteristics of non-randomised studies of drug treatments: cross sectional study. *BMJ medicine*, 3(1):e000932, 2024.

[19] Stephany N Duda, Bryan E Shepherd, Cynthia S Gadd, Daniel R Masys, and Catherine C McGowan. Measuring the quality of observational study data in an international hiv research network. *PLoS one*, 7(4):e33908, 2012.

[20] Samy Suissa. Immortal time bias in pharmacoepidemiology. *American journal of epidemiology*, 167(4):492–499, 2008.

[21] Miguel A Hernán, Jonathan AC Sterne, Julian PT Higgins, Ian Shrier, and Sonia Hernández-Díaz. A structural description of biases that generate immortal time. *Epidemiology*, 36(1):107–114, 2025.

[22] Moride Yola and Abenhaim Lucien. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *Journal of clinical epidemiology*, 47(7):731–737, 1994.

[23] Christel Renoux, Sophie Dell'Aniello, Benjamin Brenner, and Samy Suissa. Bias from depletion of susceptibles: the example of hormone replacement therapy and the risk of venous thromboembolism. *Pharmacoepidemiology and drug safety*, 26(5):554–560, 2017.

[24] Jaclyn LF Bosco, Rebecca A Silliman, Soe Soe Thwin, Ann M Geiger, Diana SM Buist, Marianne N Prout, Marianne Ulcickas Yood, Reina Haque, Feifei Wei, and Timothy L Lash. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *Journal of Clinical Epidemiology*, 63(1):64–74, 2010.

[25] Sharon H Giordano, Yong-Fang Kuo, Zhigang Duan, Gabriel N Hortobagyi, Jean Freeman, and James S Goodwin. Limits of

observational data in determining outcomes from cancer therapy. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 112(11):2456–2466, 2008.

[26] Diederick E Grobbee and Arno W Hoes. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *Bmj*, 315(7116):1151–1154, 1997.

[27] Demetrios N Kyriacou and Roger J Lewis. Confounding by indication in clinical research. *Jama*, 316(17):1818–1819, 2016.

[28] Rory Collins, Louise Bowman, Martin Landray, and Richard Peto. The magic of randomization versus the myth of real-world evidence, 2020.

[29] Adam C Fields, Pamela Lu, Deanna L Palenzuela, Ronald Bleday, Joel E Goldberg, Jennifer Irani, Jennifer S Davids, and Nelya Melnitchouk. Does retrieval bag use during laparoscopic appendectomy reduce postoperative infection? *Surgery*, 165(5):953–957, 2019.

[30] Scott A Turner, Hee Soo Jung, and John E Scarborough. Utilization of a specimen retrieval bag during laparoscopic appendectomy for both uncomplicated and complicated appendicitis is not associated

with a decrease in postoperative surgical site infection rates. *Surgery*, 165(6):1199–1202, 2019.

[31] Christopher P Childers and Melinda Maggard-Gibbons. Same data, opposite results?: A call to improve surgical database research. *JAMA surgery*, 156(3):219–220, 2021.

[32] Xiao-Li Meng et al. Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, 12(2):685–726, 2018.

[33] Raphael Silberzahn and Eric L Uhlmann. Crowdsourced research: Many hands make tight work. *Nature*, 526(7572):189–91, Oct 2015.

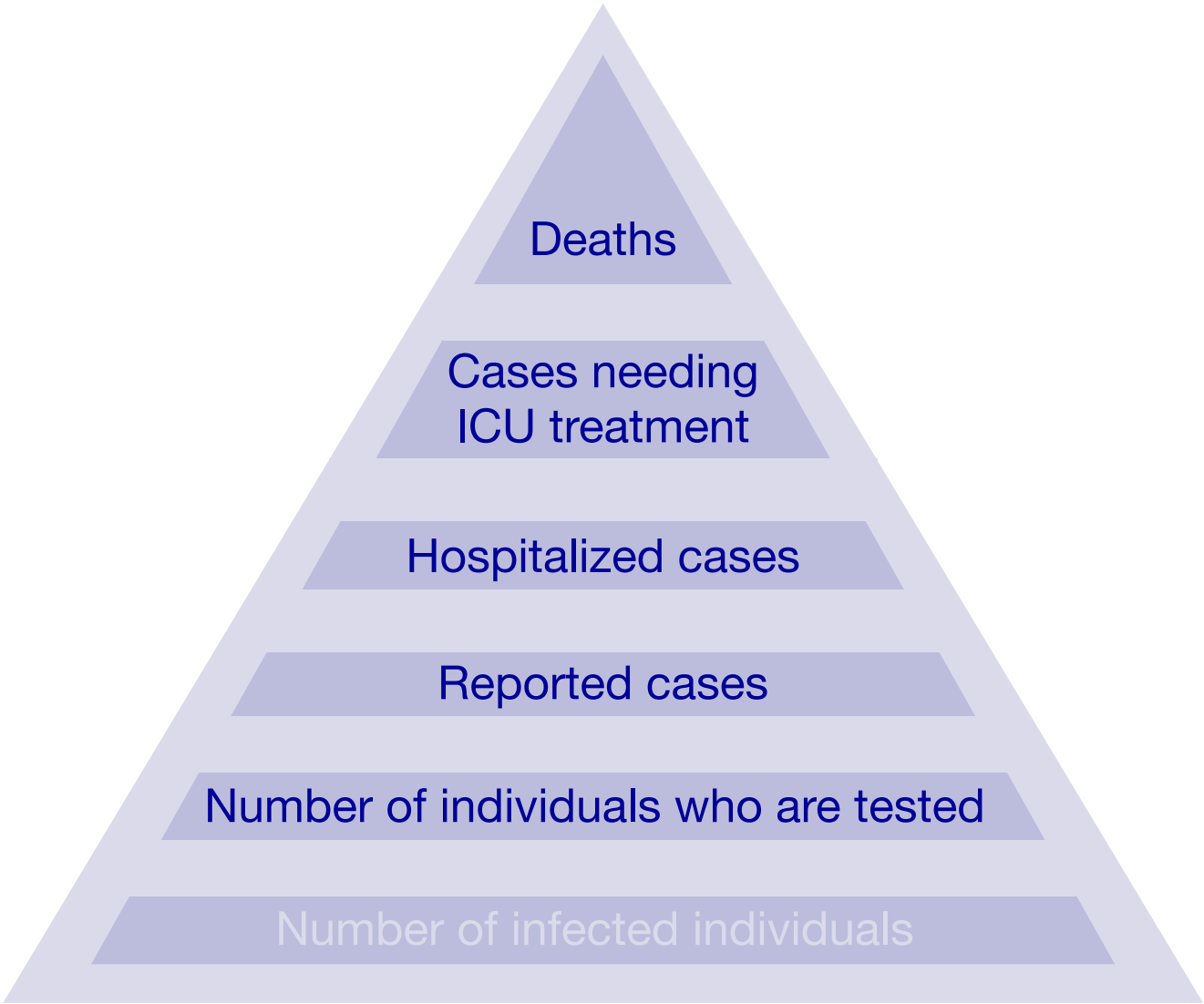
[34] Rotem Botvinik-Nezer, Felix Holzmeister, Colin F Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A Mumford, R Alison Adcock, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, pages 1–7, 2020.

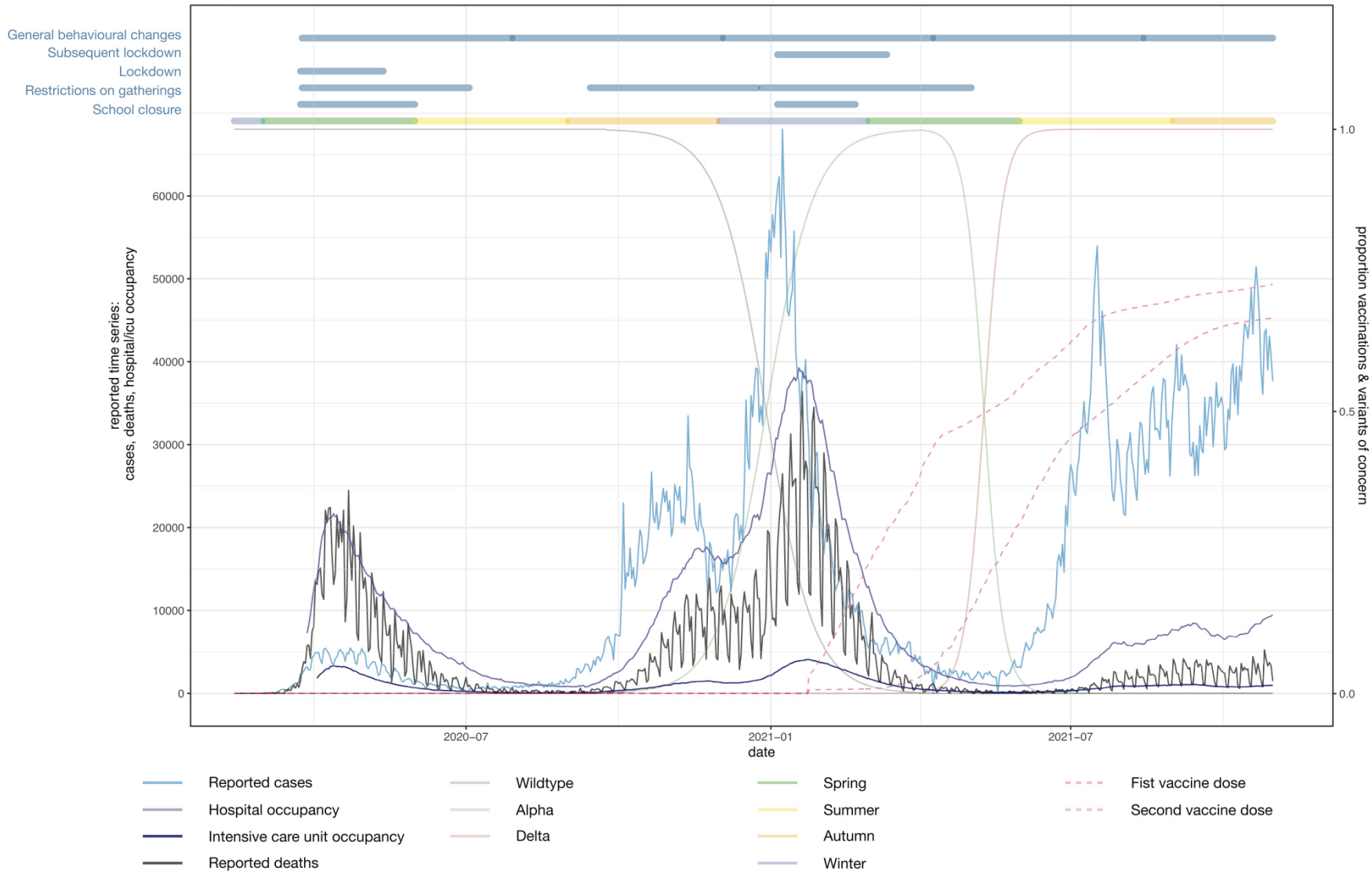
[35] Yeganeh Khazaei, Helmut Küchenhoff, Sabine Hoffmann, Diella Syliqi, and Raphael Rehms. Using a bayesian hierarchical approach to study

the association between non-pharmaceutical interventions and the spread of covid-19 in germany. *Scientific Reports*, 13(1):18900, 2023.

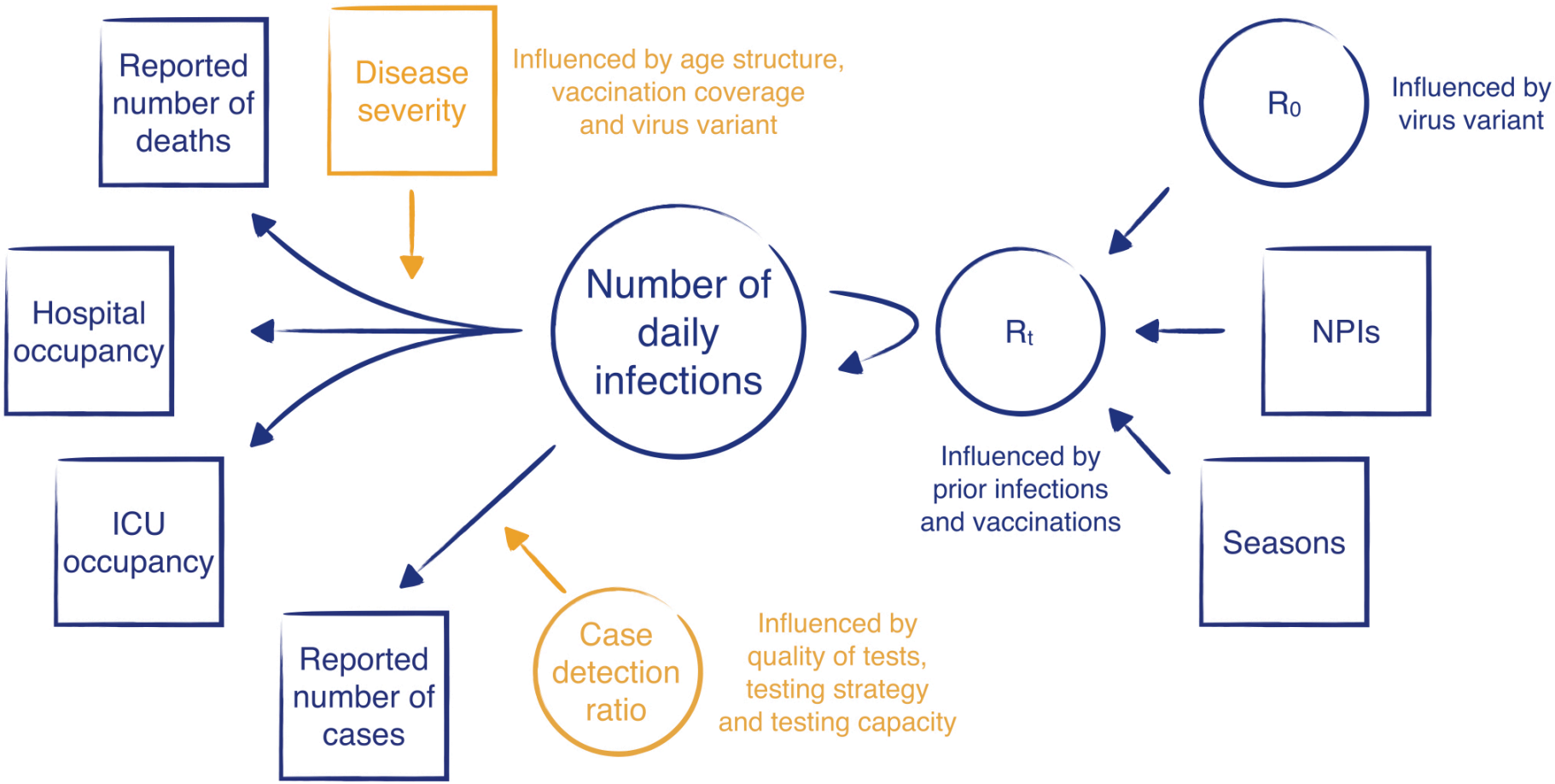
[36] Raphael Rehms, Nicole Ellenbach, Eva Rehfuess, Jacob Burns, Ulrich Mansmann, and **Hoffmann, Sabine**. **A Bayesian hierarchical approach to account for evidence and uncertainty in the modeling of infectious diseases: An application to COVID-19**. *Biometrical Journal*, 66(1):2200341, 2024.

Sources of information in the modelling of COVID-19




General behavioural changes
Subsequent lockdown
Lockdown
Restrictions on gatherings
School closure
reported time series:
cases, deaths, hospital/icu occupancy
proportion vaccinations & variants of concern
60000
50000
40000
30000
20000
10000
0
1.0
0.5
0.0
2020–07
2021–01
2021–07
date
Reported cases
Hospital occupancy
Intensive care unit occupancy
Reported deaths
Wildtype
Alpha
Delta
Spring
Summer
Autumn
Winter
Fist vaccine dose
Second vaccine dose

[35,36]


Reported number of deaths
Disease severity
Influenced by age structure, vaccination coverage and virus variant
R_0
Influenced by virus variant
Hospital occupancy
Number of daily infections
R_t
NPIs
ICU occupancy
Influenced by prior infections and vaccinations
Seasons
Reported number of cases
Case detection ratio
Influenced by quality of tests, testing strategy and testing capacity

[35,36]

A hierarchical model of COVID-19 propagation

- The disease model:

$$C_{t,m} = \sum_{u<t} I_{u,m}(F_{\xi^C}(t-u+1) - F_{\xi^C}(t-u))$$

A hierarchical model of COVID-19 propagation

- The disease model:

$$C_{t,m} = \sum_{u<t} I_{u,m}(F_{\xi^C}(t-u+1) - F_{\xi^C}(t-u))$$

- The death model:
  $D_{t,m} \sim$
  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi^D}(t-u+1) - F_{\xi^D}(t-u)), \phi_d\right)$

A hierarchical model of COVID-19 propagation

- The disease model:

$$C_{t,m} = \sum_{u<t} I_{u,m}(F_{\xi^C}(t-u+1) - F_{\xi^C}(t-u))$$

- The death model:

$D_{t,m} \sim$
Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi^D}(t-u+1) - F_{\xi^D}(t-u)), \phi_d\right)$

- The reporting model:

$C_{t,m}^R \sim$
Negative Binomial $\left(\rho_{t,m} \sum_{u<t} C_{u,m}(F_{\xi^R}(t-u+1) - F_{\xi^R}(t-u)), \phi_c\right)$

A hierarchical model of COVID-19 propagation

- The disease model:

$$C_{t,m} = \sum_{u<t} I_{u,m}(F_{\xi^C}(t-u+1) - F_{\xi^C}(t-u))$$

- The death model:
  $D_{t,m} \sim$
  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$
- The reporting model:
  $C_{t,m}^R \sim$
  Negative Binomial $\left(\rho_{t,m} \sum_{u<t} C_{u,m}(F_{\xi_m^R}(t-u+1) - F_{\xi_m^R}(t-u)), \phi_c\right)$

A hierarchical model of COVID-19 propagation

- The disease model:

$$C_{t,m} = \sum_{u<t} I_{u,m}(F_{\xi^C}(t-u+1) - F_{\xi^C}(t-u))$$

- The death model:

$D_{t,m} \sim$

Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$

- The reporting model:

$C_{t,m}^R \sim$

Negative Binomial $\left(\rho_{t,m} \sum_{u<t} C_{u,m}(F_{\xi_m^R}(t-u+1) - F_{\xi_m^R}(t-u)), \phi_c\right)$

- The hospitalization model:

$H_{t,m} \sim$

Negative Binomial $\left(\pi_m^H \sum_{u<t} C_{u,m}(F_{\xi^H}(t-u+1) - F_{\xi^H}(t-u)), \phi_h\right)$

A hierarchical model of COVID-19 propagation

- The death model:
  $D_{t,m} \sim$
  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$
- The reporting model:
  $C_{t,m}^R \sim$
  Negative Binomial $\left(\rho_{t,m} \sum_{u<t} C_{u,m}(F_{\xi_m^R}(t-u+1) - F_{\xi_m^R}(t-u)), \phi_c\right)$
- The hospitalization model:
  $H_{t,m} \sim$
  Negative Binomial $\left(\pi_m^H \sum_{u<t} C_{u,m}(F_{\xi^H}(t-u+1) - F_{\xi^H}(t-u)), \phi_h\right)$
- The renewal model:
  $I_{t,m} \sim$
  Negative Binomial $\left(R_{t,m} \sum_{u<t} I_{u,m}(F_{\gamma}(t-u+1) - F_{\gamma}(t-u)), \phi_i\right)$

A hierarchical model of COVID-19 propagation

- The death model:

  $D_{t,m} \sim$

  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$

- The renewal model:

  $I_{t,m} \sim$

  Negative Binomial $\left(R_{t,m} \sum_{u<t} I_{u,m}(F_\gamma(t-u+1) - F_\gamma(t-u)), \phi_i\right)$

$$R_{t,m} = R_m^0 \cdot \exp\left(-\sum_{k=1}^{K} \alpha_{k,m} In_{k,t,m}\right)$$

$$R_m^0 \sim \mathcal{N}(R^0, \sigma_R)$$

$$\alpha_{k,m} \sim \mathcal{N}(\alpha_k, \sigma_{\alpha_k})$$

A hierarchical model of COVID-19 propagation

- The death model:
  $D_{t,m} \sim$
  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$

- The renewal model:
  $I_{t,m} \sim$ Negative Binomial (τ_m, ϕ_i) for t = 1
  $I_{t,m} \sim$
  Negative Binomial $\left(R_{t,m} \sum_{u<t} I_{u,m}(F_\gamma(t-u+1) - F_\gamma(t-u)), \phi_i\right)$

$$R_{t,m} = R_m^0 \cdot \exp\left(-\sum_{k=1}^{K} \alpha_{k,m} In_{k,t,m}\right)$$

A hierarchical model of COVID-19 propagation

- The death model:
  $D_{t,m} \sim$
  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$

- The renewal model:
  $I_{t,m} \sim$
  Negative Binomial $\left(R_{t,m} \sum_{u<t} I_{u,m}(F_\gamma(t-u+1) - F_\gamma(t-u)), \phi_i\right)$

$$R_{t,m} = R^0_{t,m} \cdot \exp\left(-\sum_{k=1}^{K} \alpha_{k,m} In_{k,t,m}\right)$$

$$R^0_{t,m} = R^0_m \cdot (1 - p^\alpha_{t,m} - p^\delta_{t,m}) + (1+\beta^\alpha) \cdot R^0_m \cdot p^\alpha_{t,m} + (1+\beta^\delta) \cdot R^0_m \cdot p^\delta_{t,m}$$

A hierarchical model of COVID-19 propagation

- The death model:

  $D_{t,m} \sim$

  Negative Binomial $\left(\pi^{D}_{m,t} \sum_{u<t} C_{u,m}(F_{\xi^{D}_{m}}(t-u+1) - F_{\xi^{D}_{m}}(t-u)), \phi_d\right)$

- The renewal model:

  $I_{t,m} \sim$

  Negative Binomial $\left(R_{t,m} \sum_{u<t} I_{u,m}(F_{\gamma}(t-u+1) - F_{\gamma}(t-u)), \phi_i\right)$

$$R_{t,m} = R^{0}_{t,m} \cdot \exp\left(-\sum_{k=1}^{K} \alpha_{k,m} In_{k,t,m}\right) \cdot (1 - c^{1}_{t,m} - c^{2}_{t,m} \cdot (1 - c^{1}_{t,m}))$$

$$c^{1}_{t,m} = \frac{\sum_{u<t} I_{u,m}}{N_m} \cdot (1 - \beta^{reinf}) \text{ and}$$

$$c^{2}_{t,m} = \frac{\sum_{u<t} Vacc^{1}_{u,m} \cdot \beta^{v1} + Vacc^{2}_{u,m} \cdot \beta^{v2}}{N_m}$$

A hierarchical model of COVID-19 propagation

- The death model:

  $D_{t,m} \sim$

  Negative Binomial $\left(\pi^{D}_{m,t} \sum_{u<t} C_{u,m}(F_{\xi^{D}_{m}}(t-u+1) - F_{\xi^{D}_{m}}(t-u)), \phi_d\right)$

- The renewal model:

  $I_{t,m} \sim$

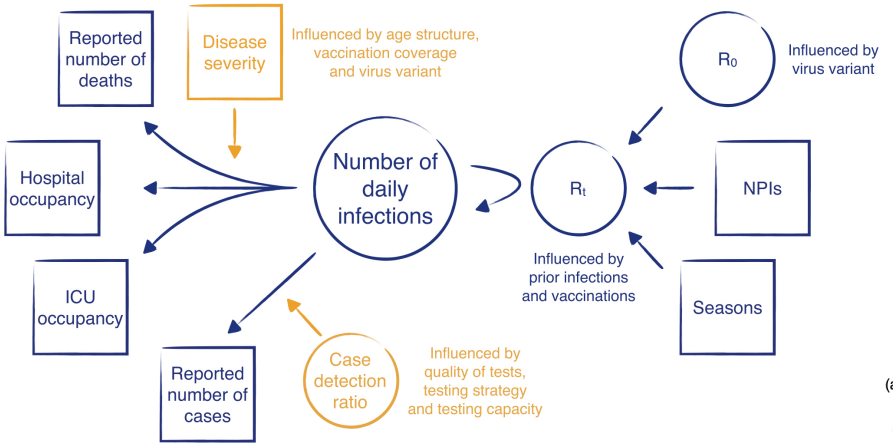
  Negative Binomial $\left(R_{t,m} \sum_{u<t} I_{u,m}(F_{\gamma}(t-u+1) - F_{\gamma}(t-u)), \phi_i\right)$

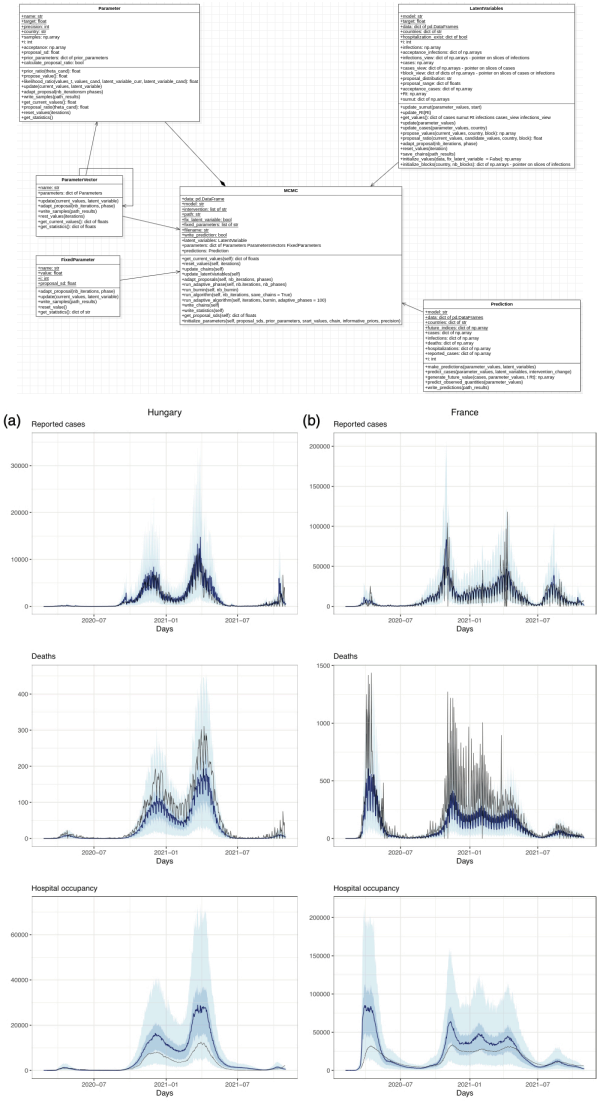
$$R_{t,m} = R^{0}_{t,m} \cdot \exp\left(-\sum_{k=1}^{K} \alpha_{k,m} In_{k,t,m}\right) \cdot (1 - c^{1}_{t,m} - c^{2}_{t,m} \cdot (1 - c^{1}_{t,m}))$$

$$c^{1}_{t,m} = \frac{\sum_{u<t} I_{u,m}}{N_m} \cdot (1 - \beta^{reinf}) \text{ and}$$

$$c^{2}_{t,m} = \frac{\sum_{u<t} Vacc^{1}_{u,m} \cdot \beta^{v1} + Vacc^{2}_{u,m} \cdot \beta^{v2}}{N_m}$$

Integrate uncertainty in infectious disease modelling

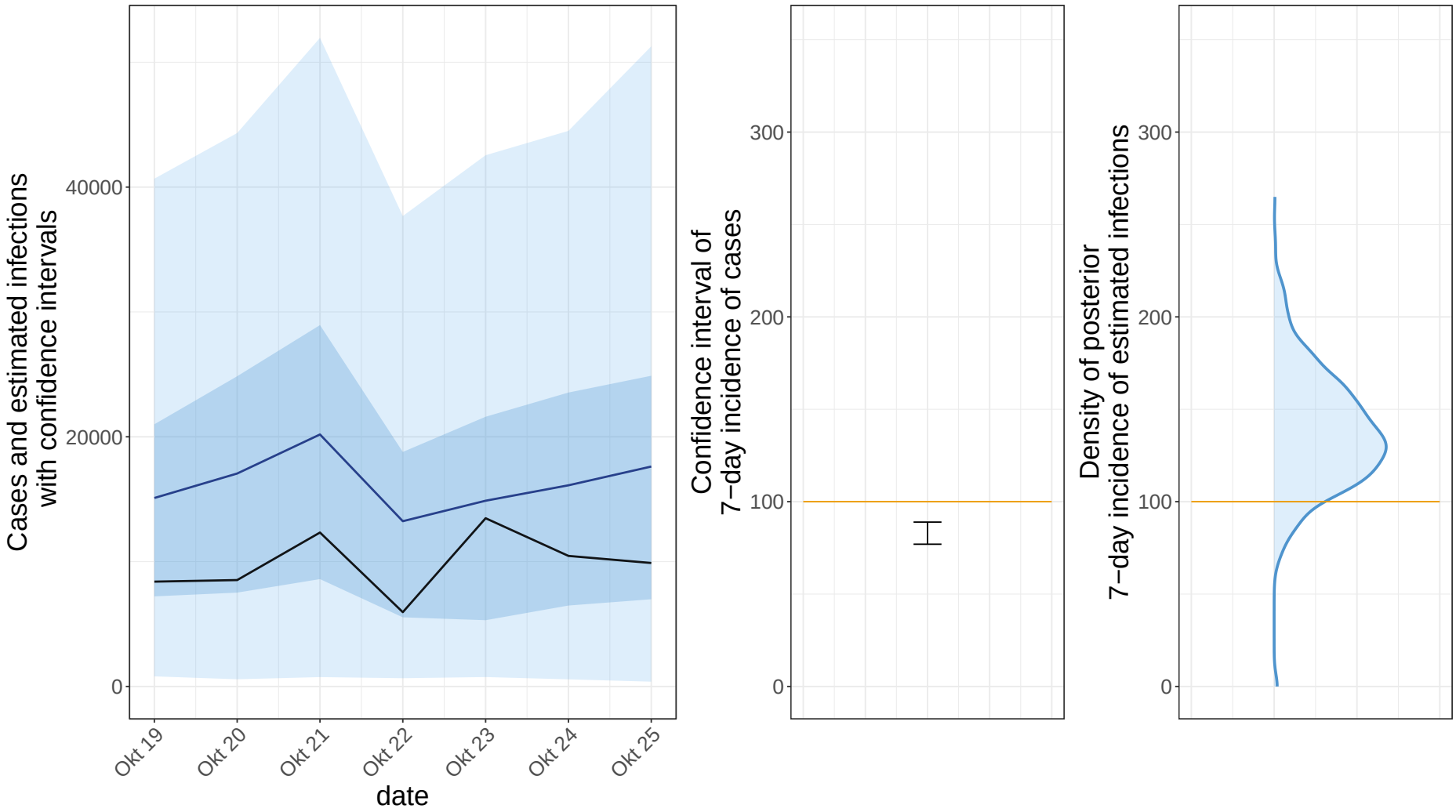




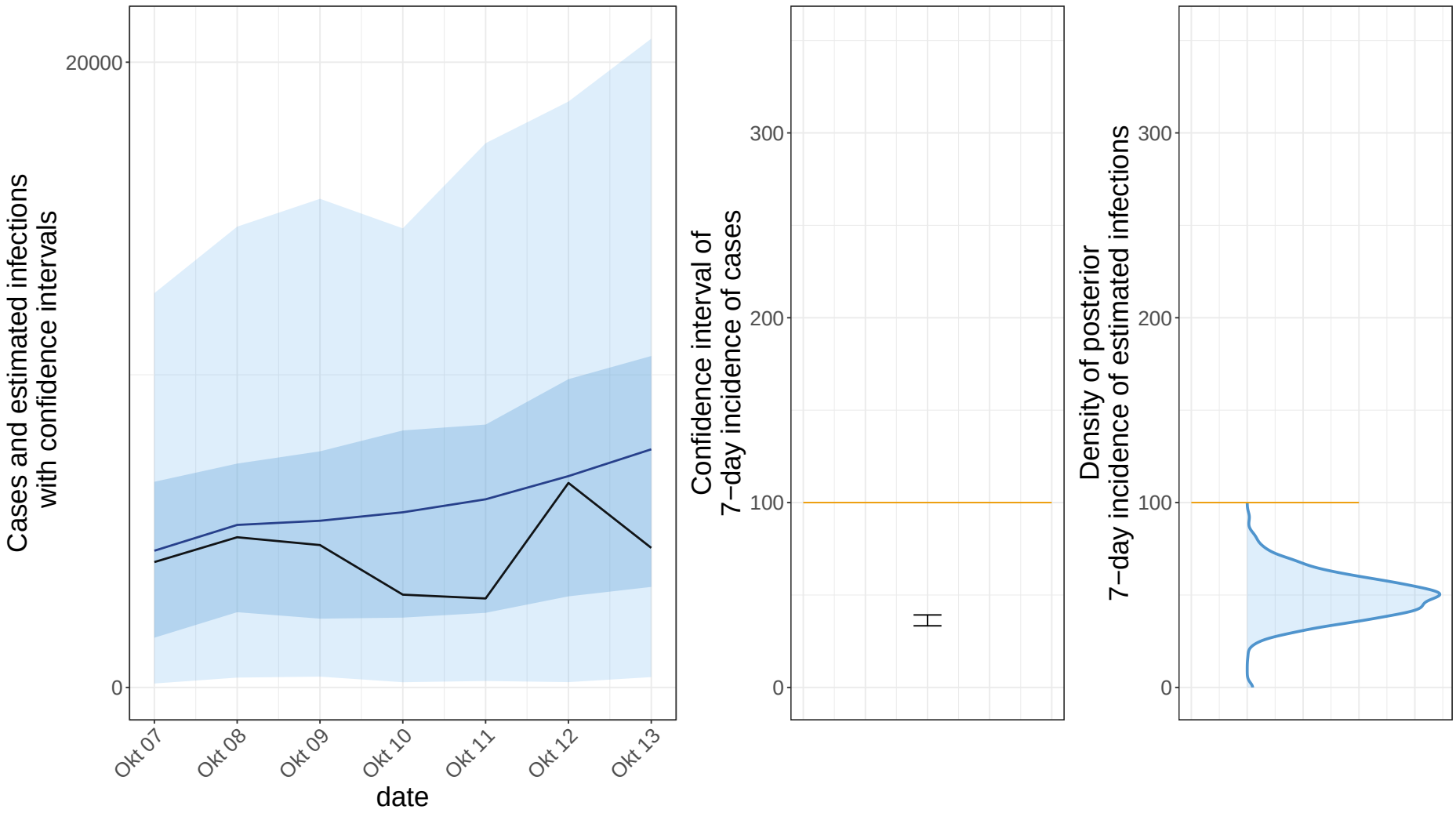
- The disease model:

$$C_{t,m} = \sum_{u<t} I_{u,m}(F_{\xi^C}(t-u+1) - F_{\xi^C}(t-u))$$

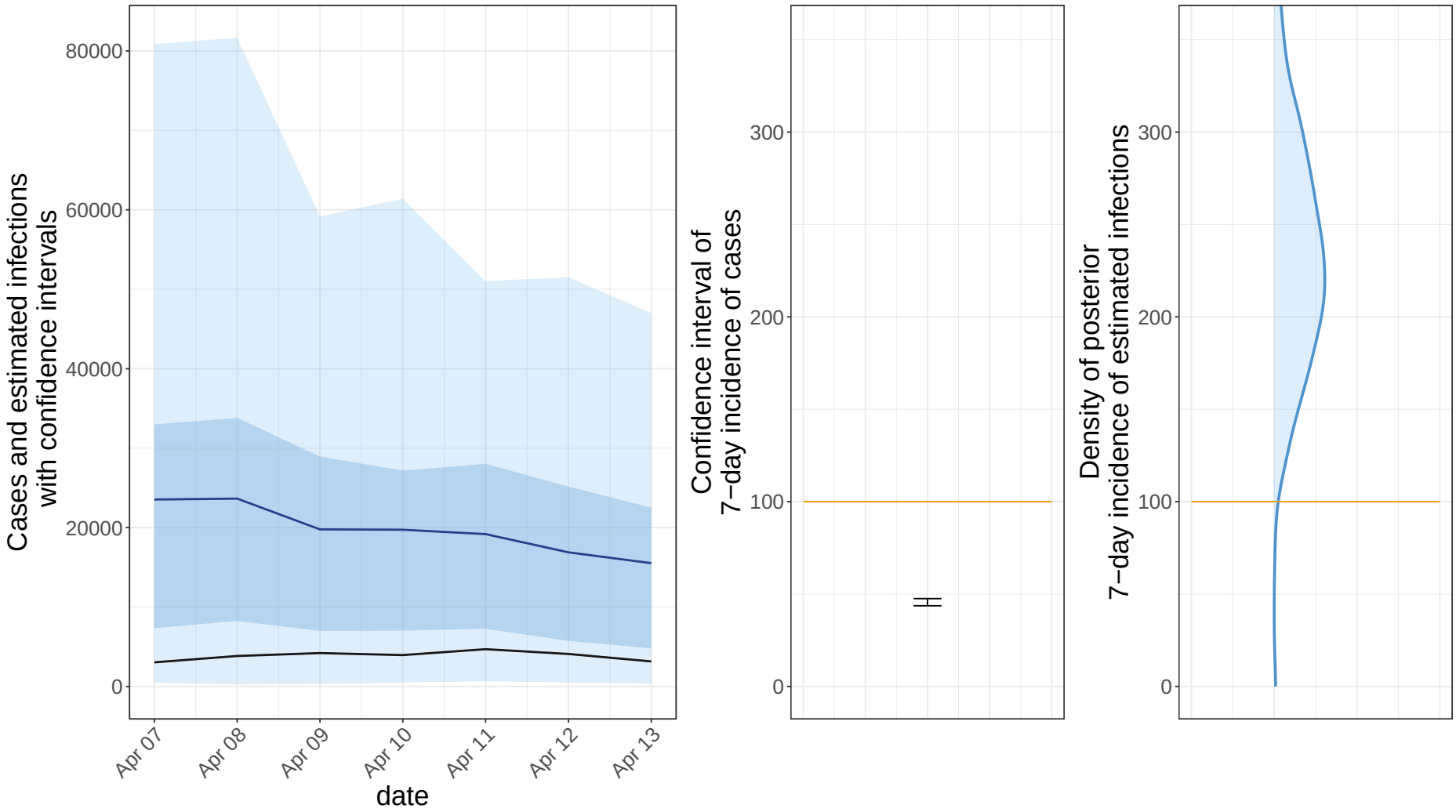
- The death model:
  $D_{t,m} \sim$
  Negative Binomial $\left(\pi_m^D \sum_{u<t} C_{u,m}(F_{\xi_m^D}(t-u+1) - F_{\xi_m^D}(t-u)), \phi_d\right)$
- The reporting model:
  $C_{t,m}^R \sim$
  Negative Binomial $\left(\rho_{t,m} \sum_{u<t} C_{u,m}(F_{\xi_m^R}(t-u+1) - F_{\xi_m^R}(t-u)), \phi_c\right)$
- The hospitalization model:
  $H_{t,m} \sim$
  Negative Binomial $(\pi_m^H \sum_{u<t} C_{u,m}(F_{\xi^H}(t-u+1) - F_{\xi^H}(t-u)), \phi_h)$


Cases and estimated infections with confidence intervals
40000
20000
0
Okt 19
Okt 20
Okt 21
Okt 22
Okt 23
Okt 24
Okt 25
date
Confidence interval of 7−day incidence of cases
300
200
100
0
Density of posterior 7−day incidence of estimated infections
300
200
100
0

[35,36]


Cases and estimated infections with confidence intervals
20000
0
Okt 07
Okt 08
Okt 09
Okt 10
Okt 11
Okt 12
Okt 13
date
Confidence interval of 7−day incidence of cases
300
200
100
0
H
Density of posterior 7−day incidence of estimated infections
300
200
100
0

[35,36]


Cases and estimated infections with confidence intervals
80000
60000
40000
20000
0
Apr 07
Apr 08
Apr 09
Apr 10
Apr 11
Apr 12
Apr 13
date
Confidence interval of 7−day incidence of cases
300
200
100
0
Density of posterior 7−day incidence of estimated infections
300
200
100
0

[35,36]