Bayesian hierarchical modelling to relax auxiliary assumptions

Sabine Hoffmann

24.03.2026

Ludwig-Maximilians-Universität München

1. Auxiliary assumptions in statistical modelling: Making the problem fit the tools

2. Bayesian hierarchical models to relax auxiliary assumptions
   - Modelling infectious diseases
   - Accounting for exposure measurement error in occupational cohorts
   - Estimating the prevalence of drug use
   - Addressing incommensurability in meta-analysis of randomised controlled trials on treatments for depression
   - Big data paradoxes in routinely collected data

3. Summary

Main message

- Stephen Senn's view:

  "A Bayesian is one who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes that he has seen a mule"

Main message

- Stephen Senn's view:

  "A Bayesian is one who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes that he has seen a mule"

- In this talk, I will try to convince you that:

  Frequentists do not need to be worried about stubborn mules.

Main message

- Stephen Senn's view:

  "A Bayesian is one who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes that he has seen a mule"

- In this talk, I will try to convince you that:

  Frequentists do not need to be worried about stubborn mules.

  A frequentist is one who strongly expecting a horse and clearly viewing a donkey, insists that he has indeed seen a horse

Making the problem fit the tools: Auxiliary assumptions in statistical modelling

The double underdetermination problem

Underdetermination of scientific theory by evidence:

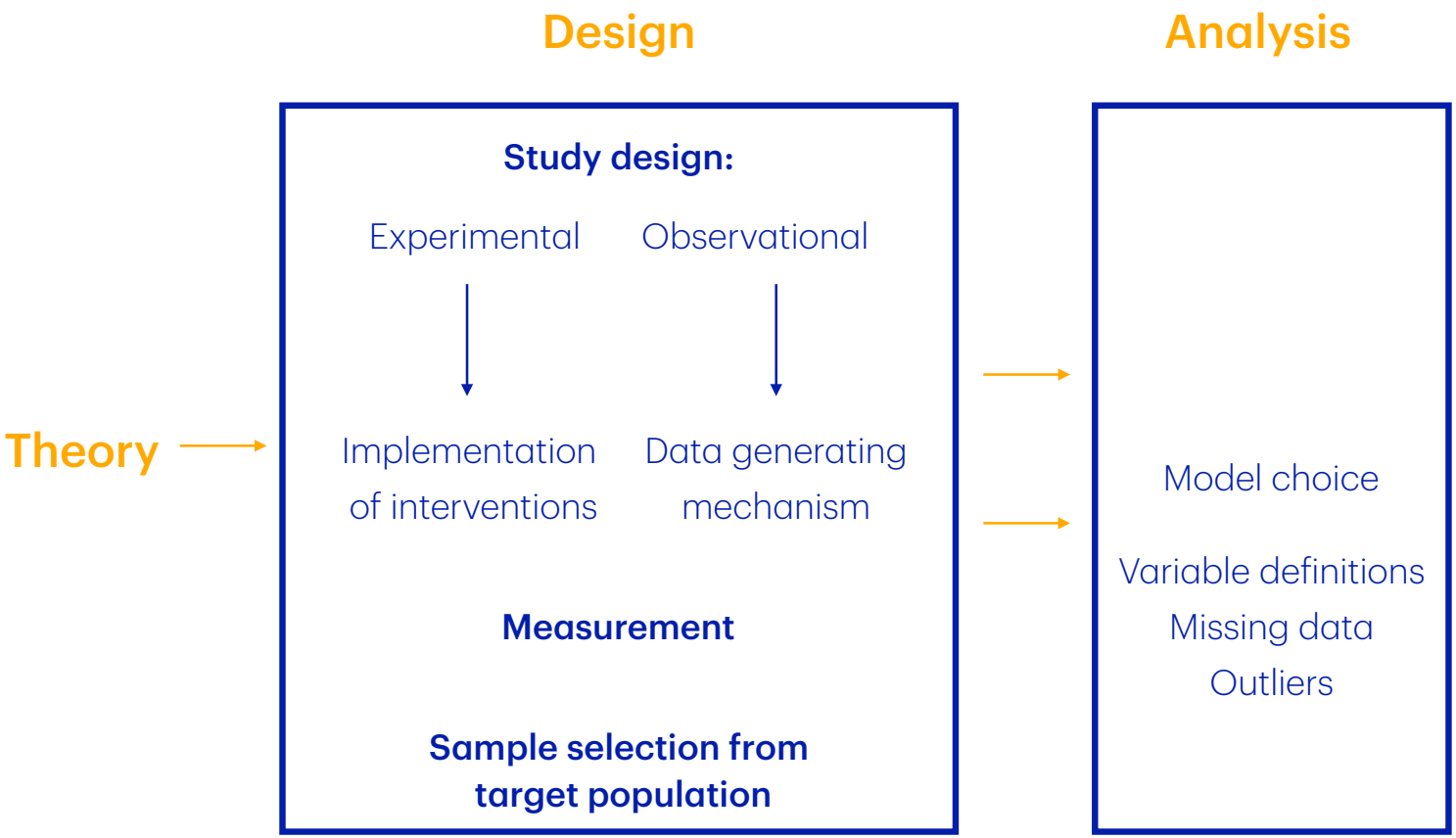
It is not possible to unambiguously falsify a theory, because theories are always tested in a bundle with various auxiliary assumptions.
When a prediction fails, we never know whether we should blame the theory or one of the auxiliary assumptions.

(Duhem-Quine problem)

The double underdetermination problem

Underdetermination of scientific theory by evidence

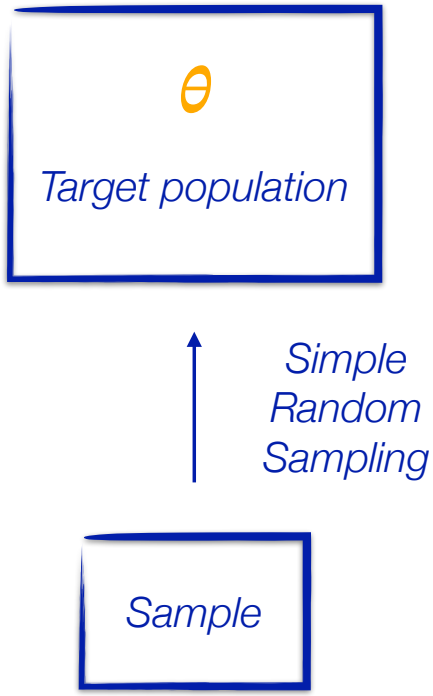
Type 2: Underdetermination of evidence by scientific theory



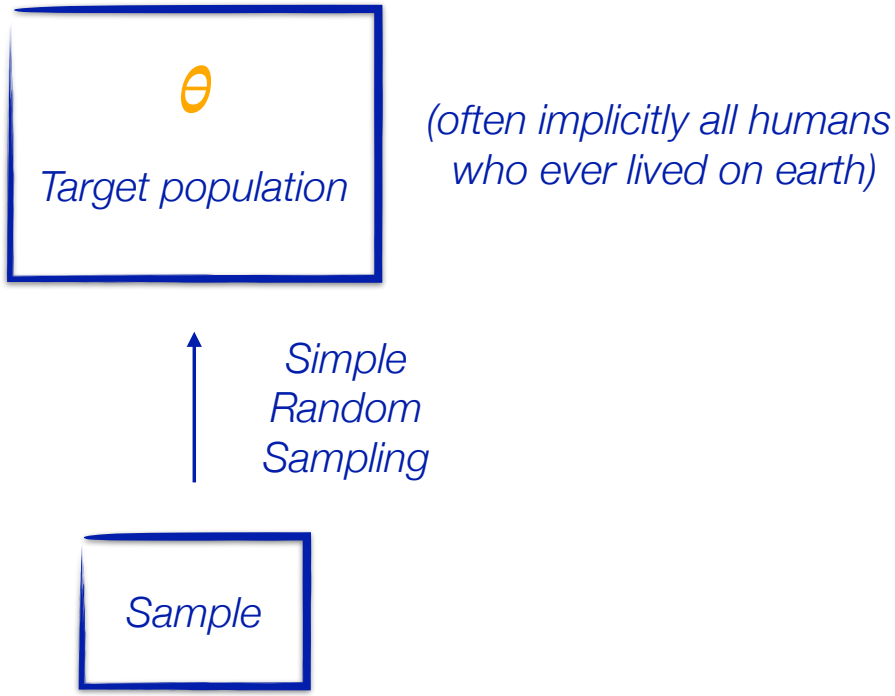
Landy et al. (2020, Psychological Bulletin)
Almaatouq et al. (2022, Behavioural and Brain Sciences)
Huber et al. (2023, PNAS)

Silberzahn et al. (2015, Nature)
Botvinik-Nezer et al. (2020, Nature)
Aczel et al. (2026, in press in Nature)

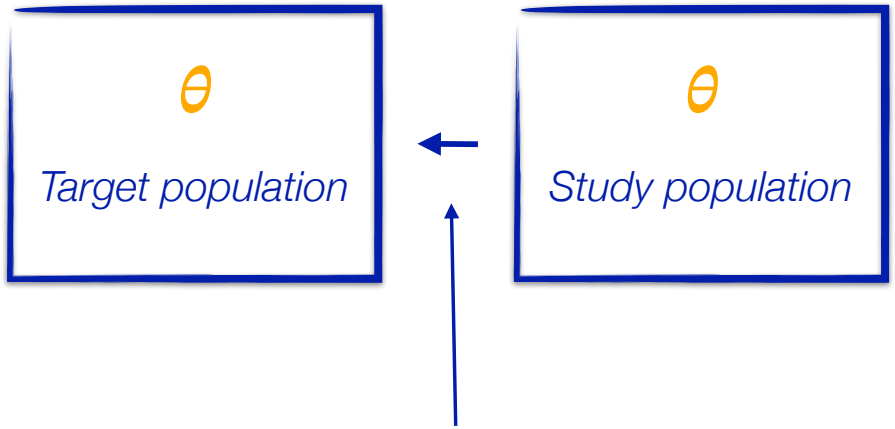
Auxiliary assumptions in statistical modelling



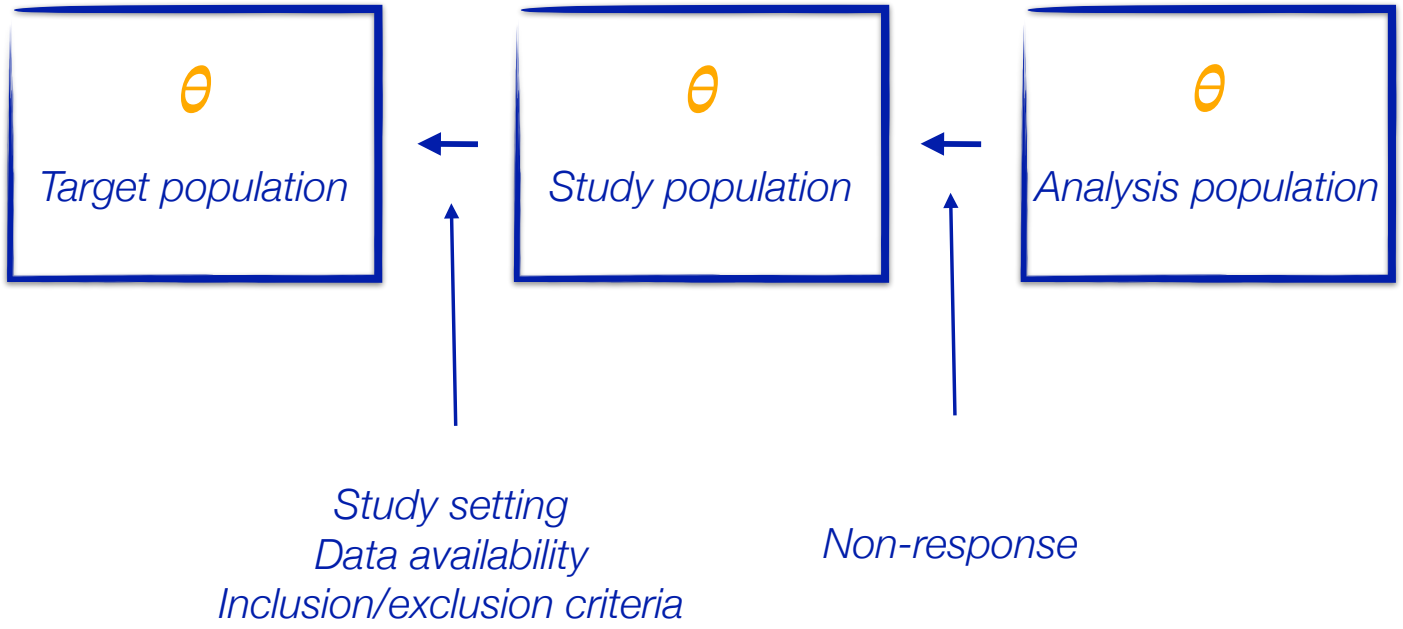
Auxiliary assumptions in statistical modelling



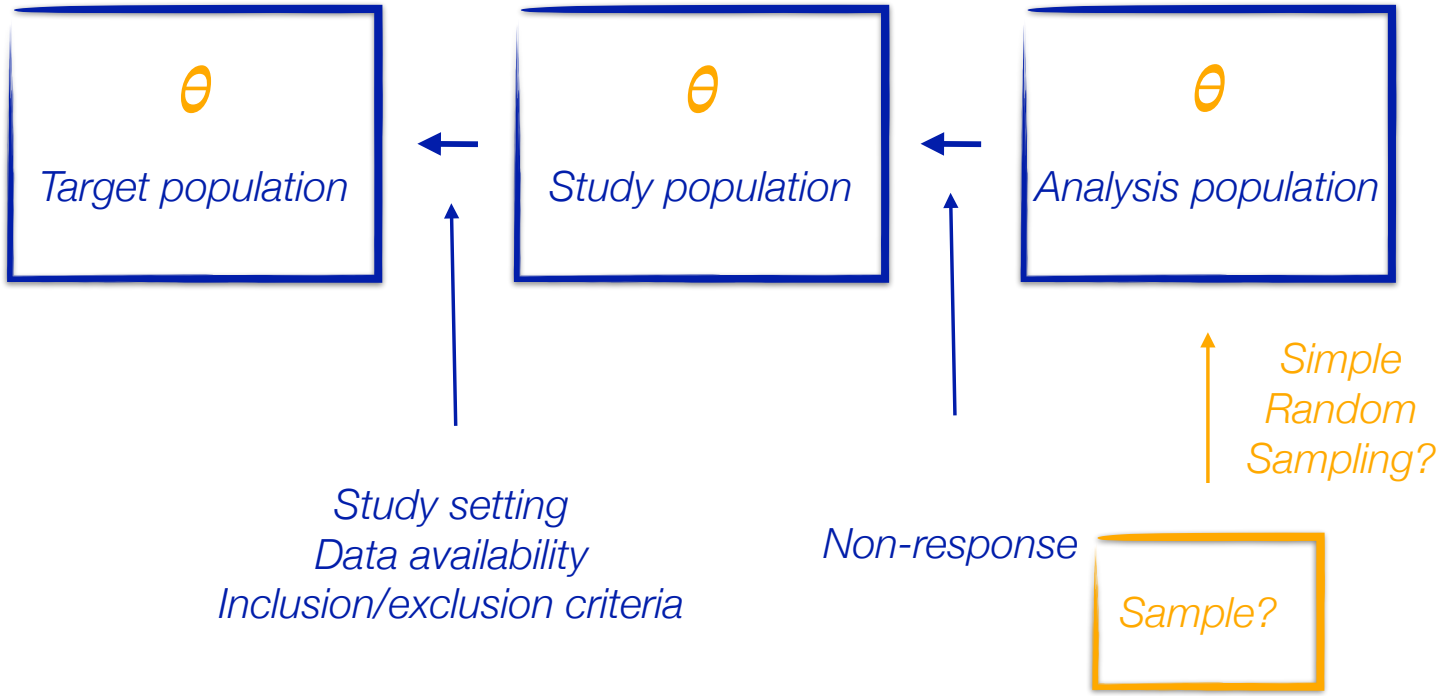
Auxiliary assumptions in statistical modelling



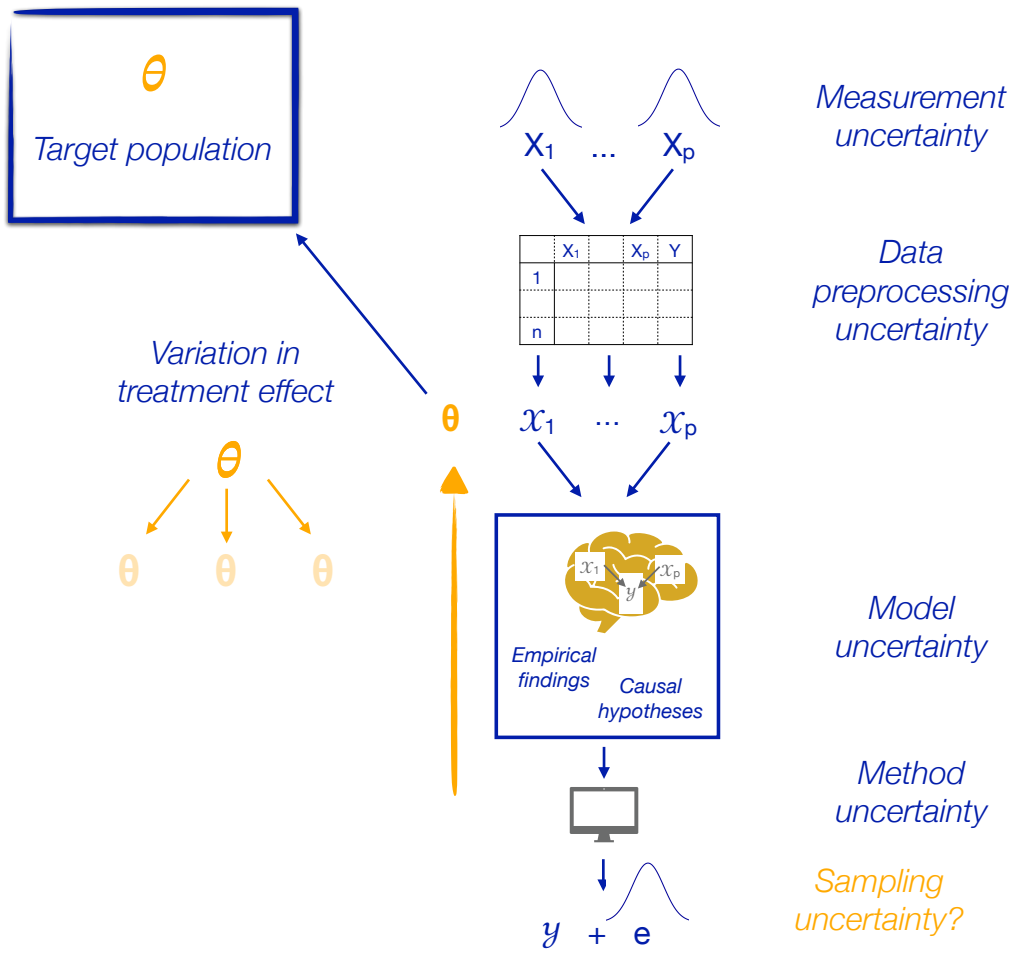
Auxiliary assumptions in statistical modelling



Auxiliary assumptions in statistical modelling

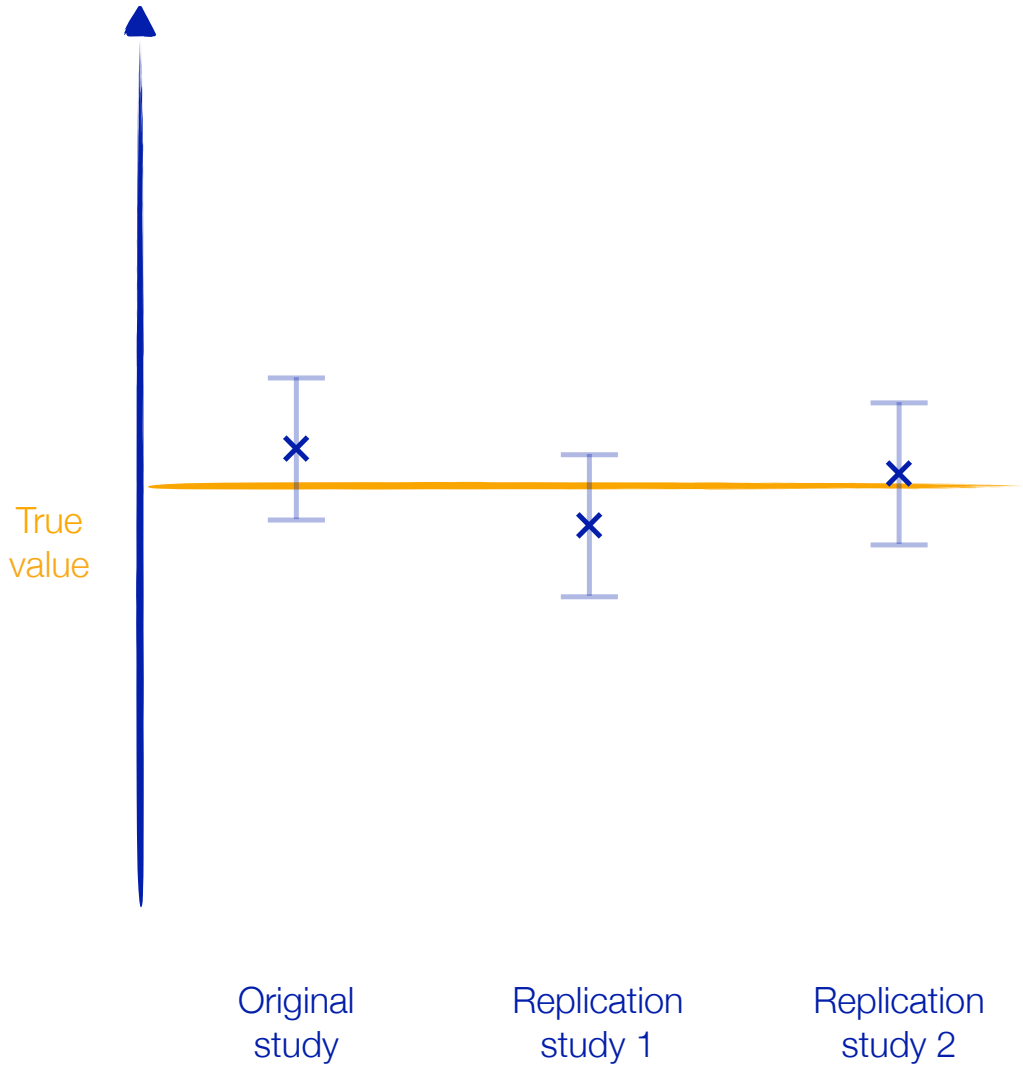


Existing and nonexistant sources of uncertainty

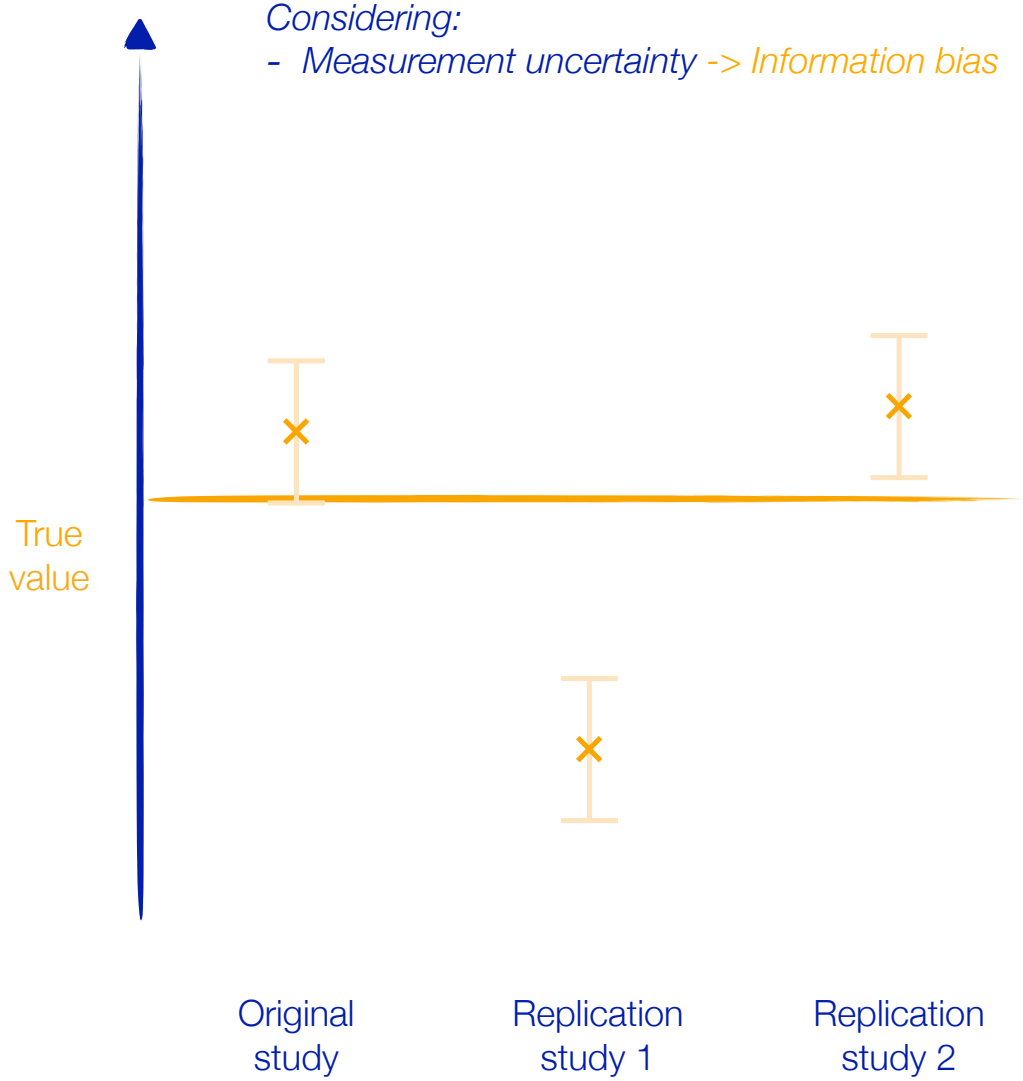


Hoffmann, S., F. Schönbrodt, R. Elsas, R. Wilson, U. Strasser, Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science 8 201925

Impact on statistical inference

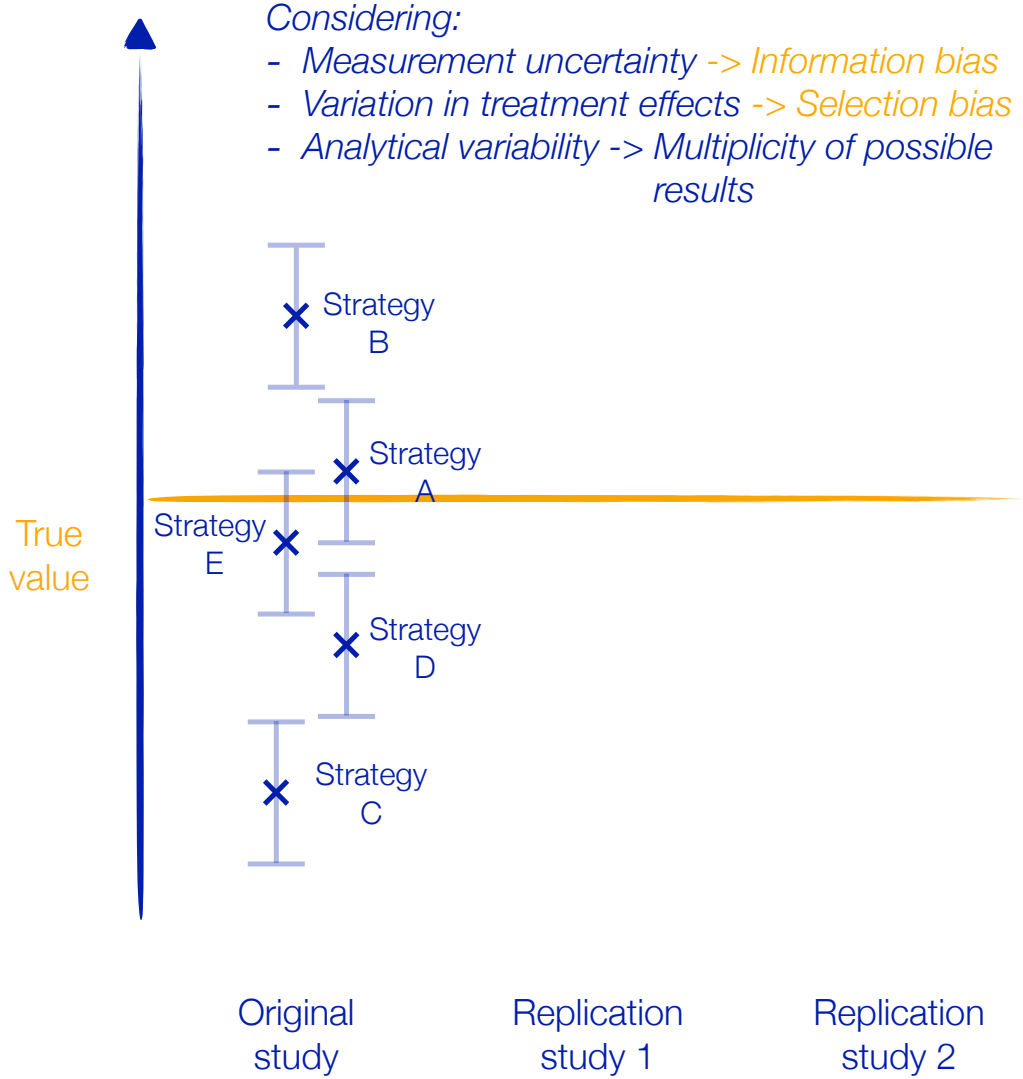


Unjustified auxiliary assumptions lead to bias

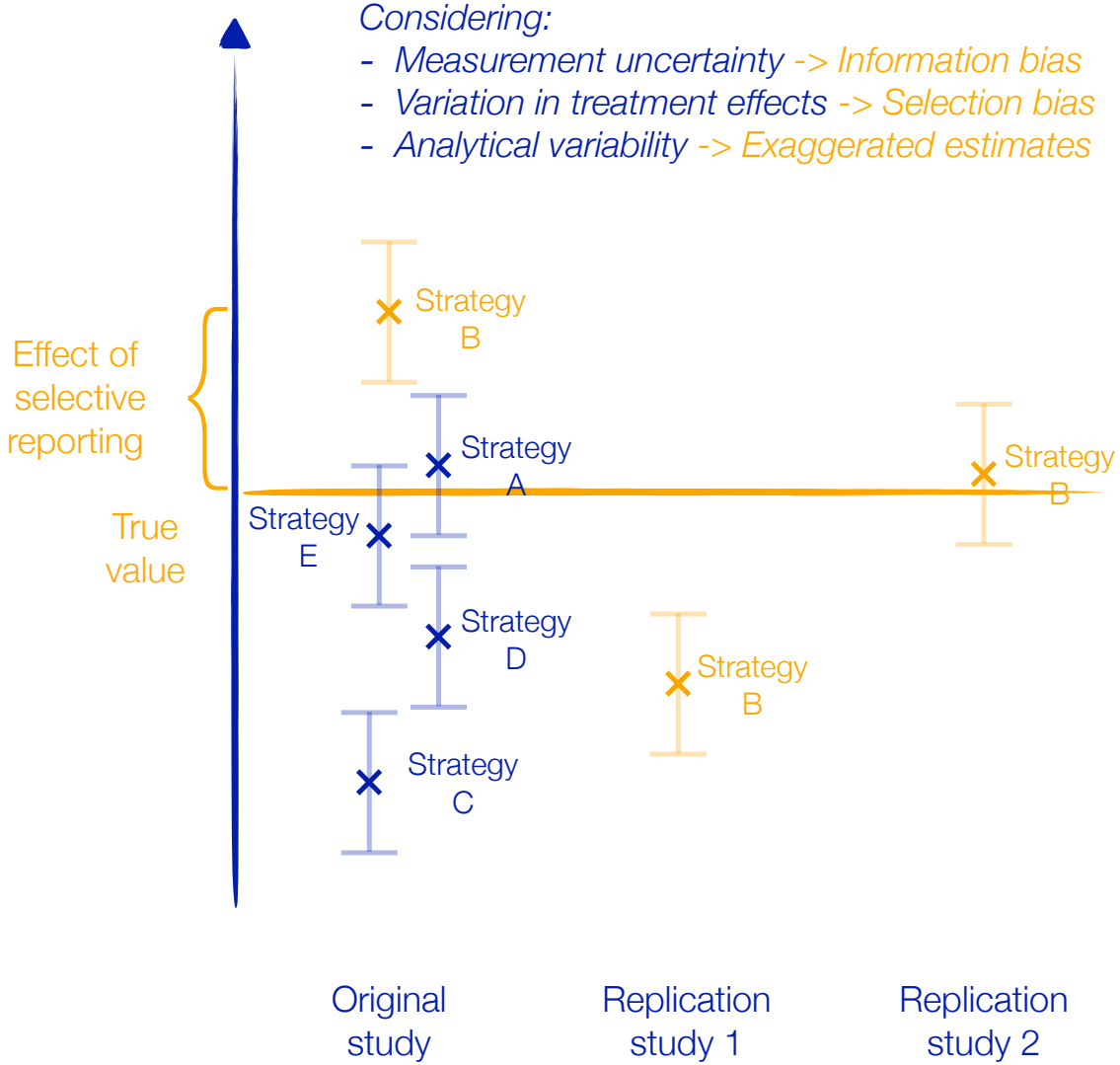


Unjustified auxiliary assumptions lead to bias

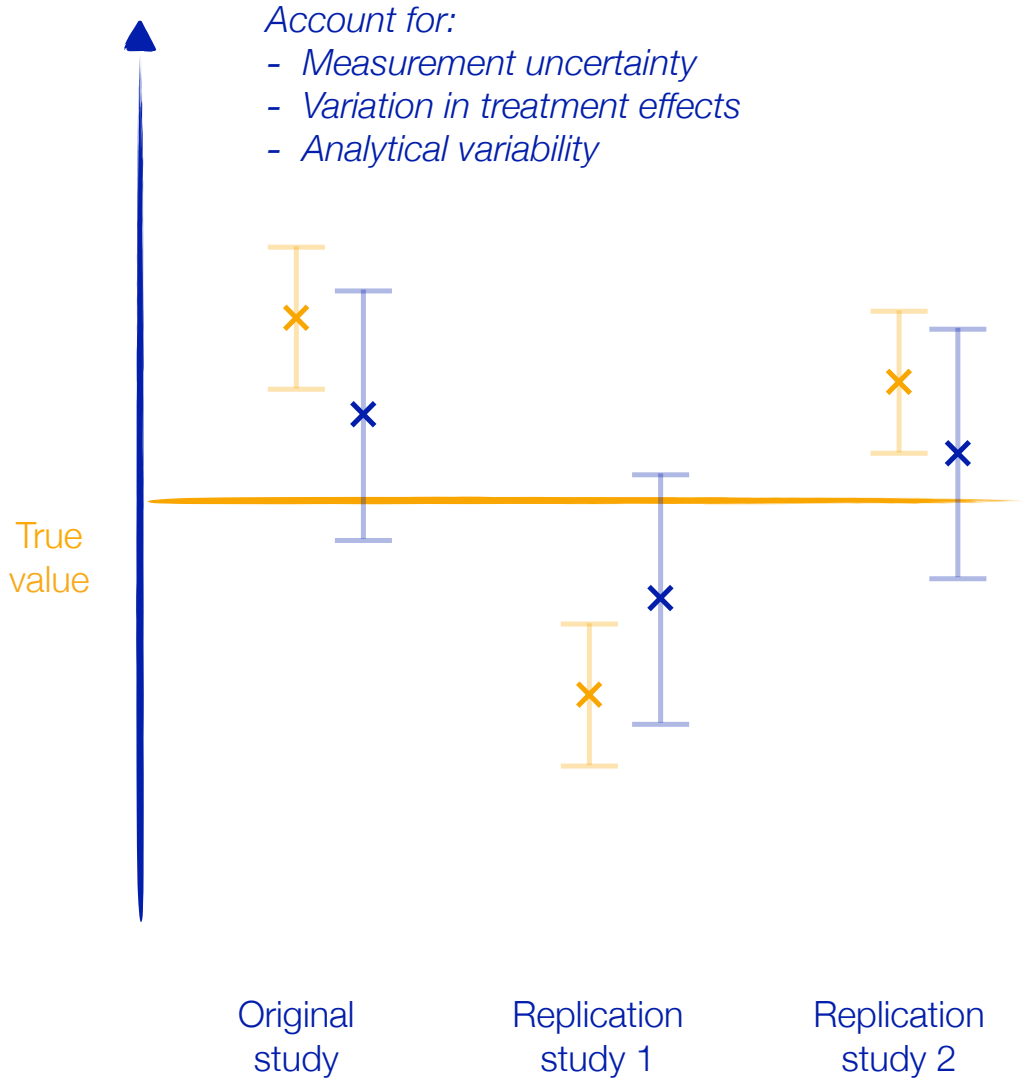
Unjustified auxiliary assumptions lead to bias and overoptimism



Auxiliary assumptions lead to non-replicable and incommensurable research findings

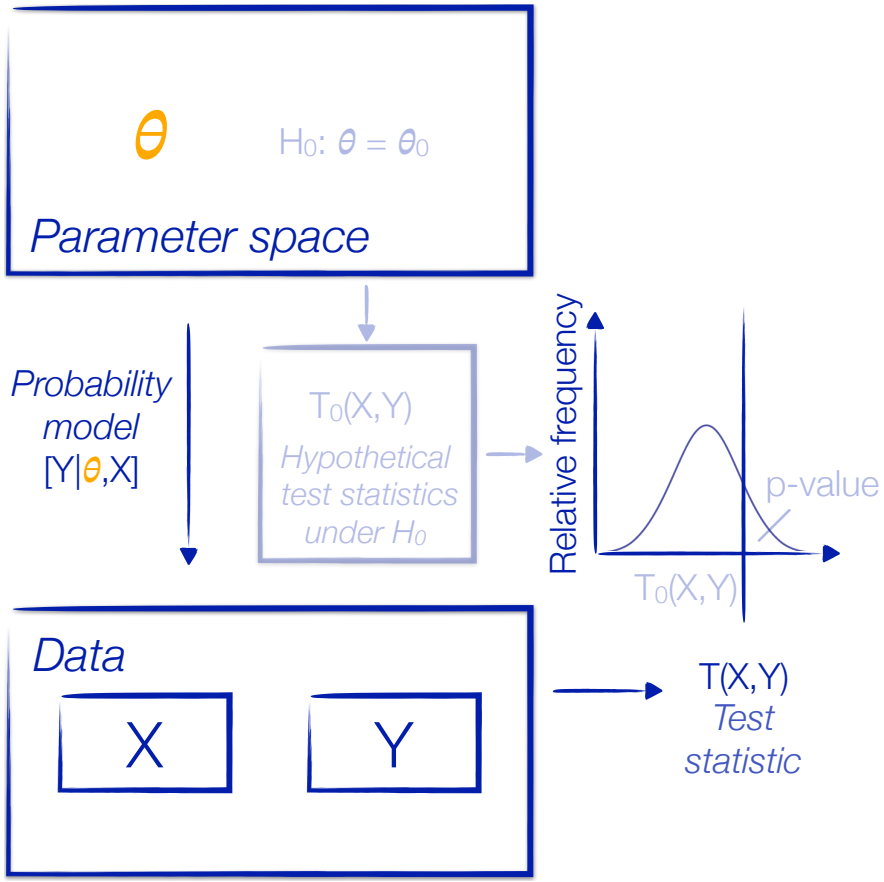


Aim: Relax auxiliary assumptions to reduce overconfidence and improve replicability

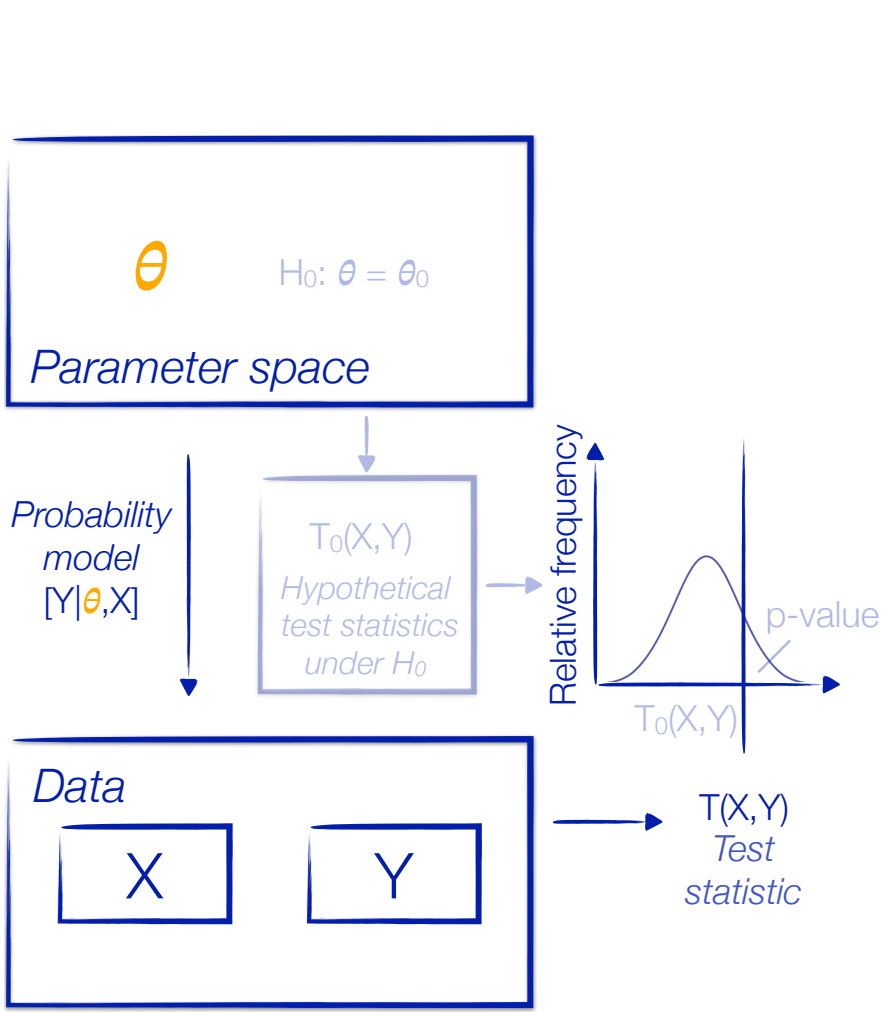


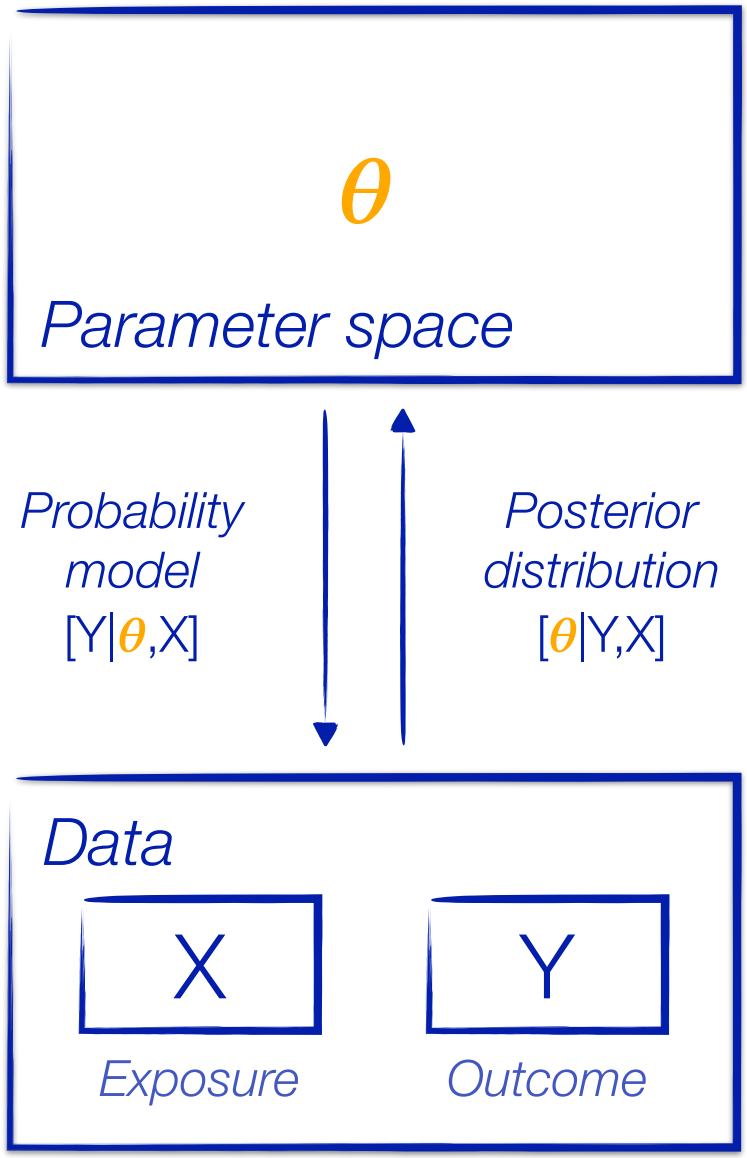
Bayesian hierarchical models to relax auxiliary assumptions

The inverse problem in statistical inference

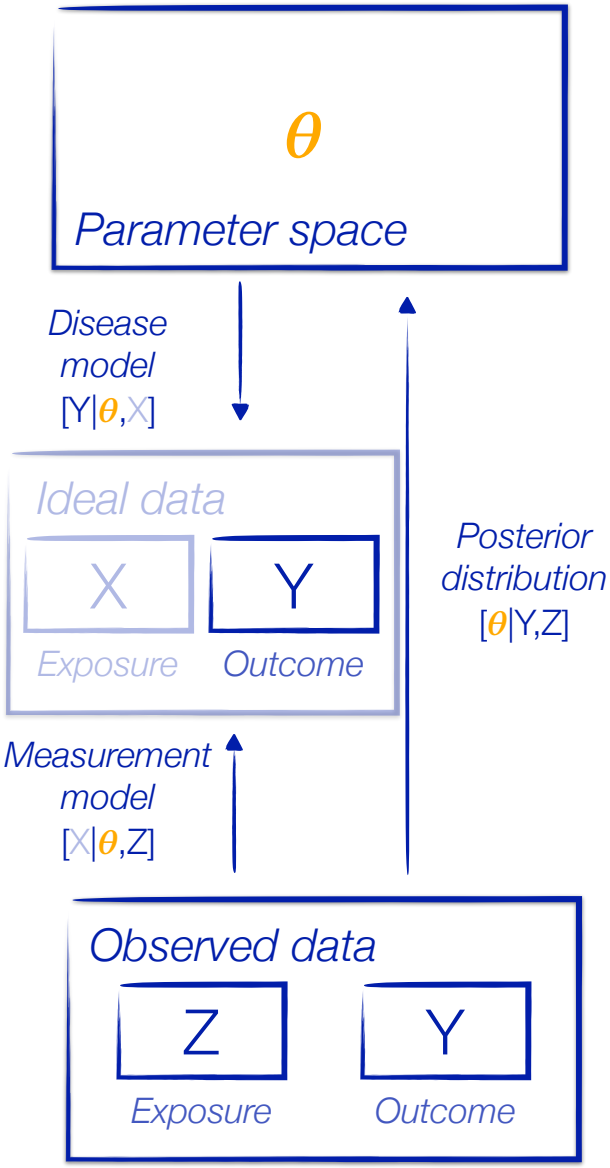


The inverse problem in statistical inference

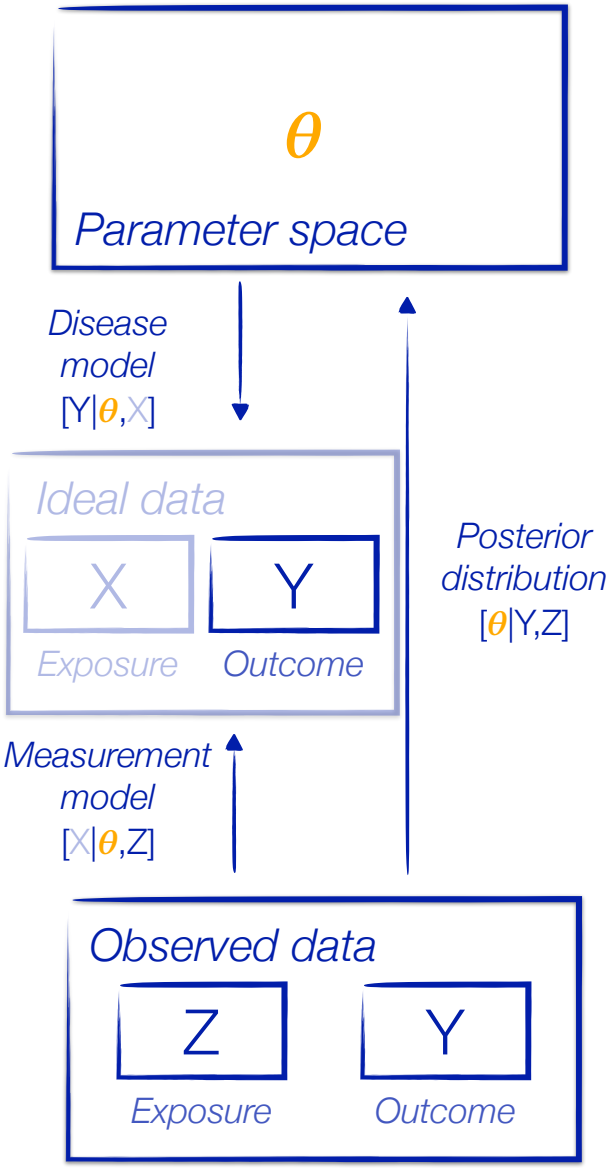




Bayesian hierarchical models

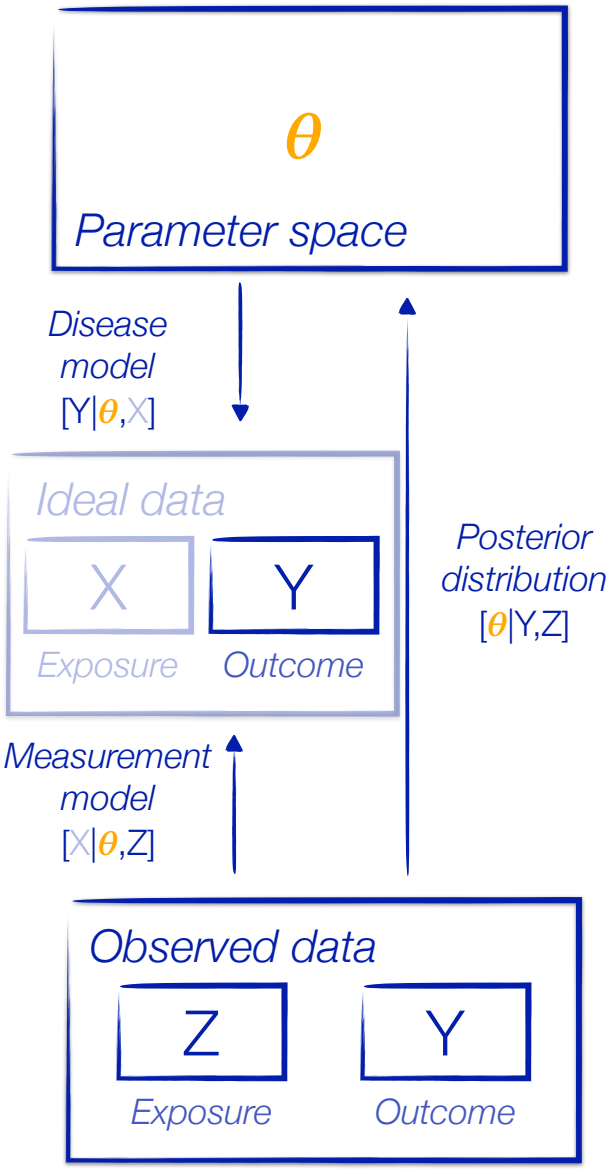


Bayesian hierarchical models



Suggestions on how to come up with statistical models

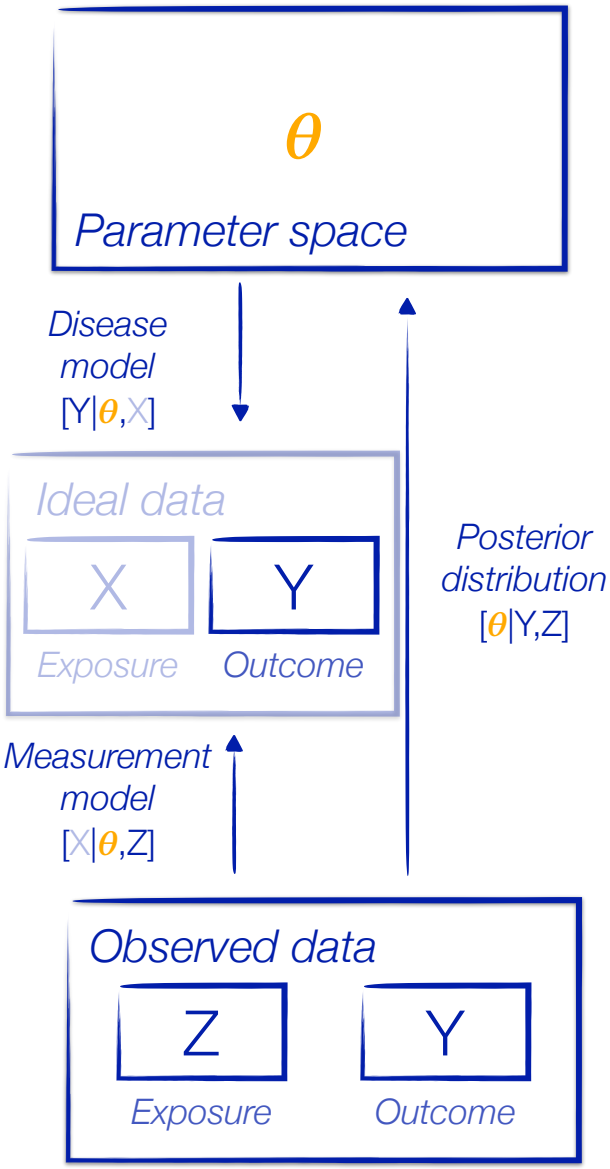
Bayesian hierarchical models



Suggestions on how to come up with statistical models

- Think (very hard) about your data generating mechanism

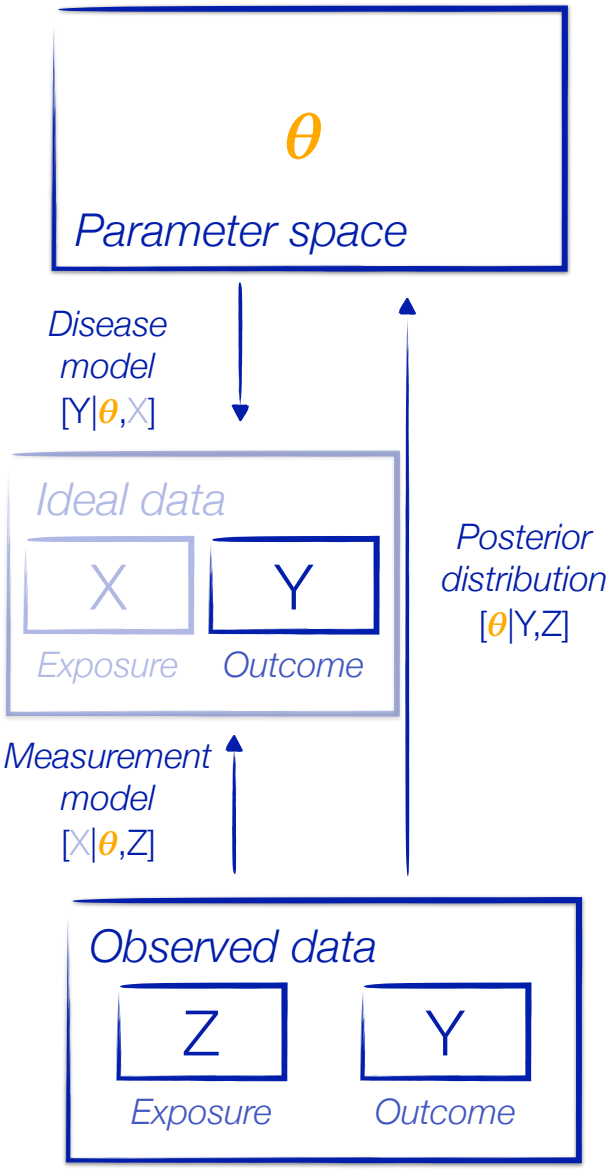
Bayesian hierarchical models



Suggestions on how to come up with statistical models

- Think (very hard) about your data generating mechanism
- Explore the design space to come up with reasonable auxiliary assumptions

Bayesian hierarchical models

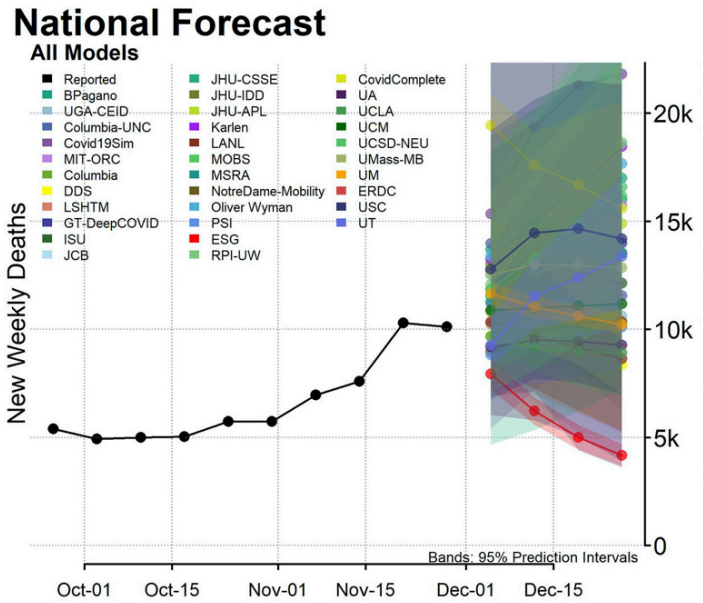


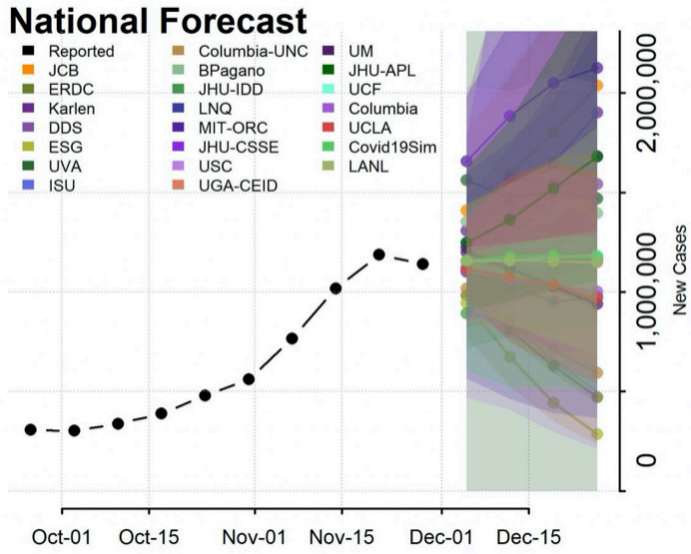
Suggestions on how to come up with statistical models

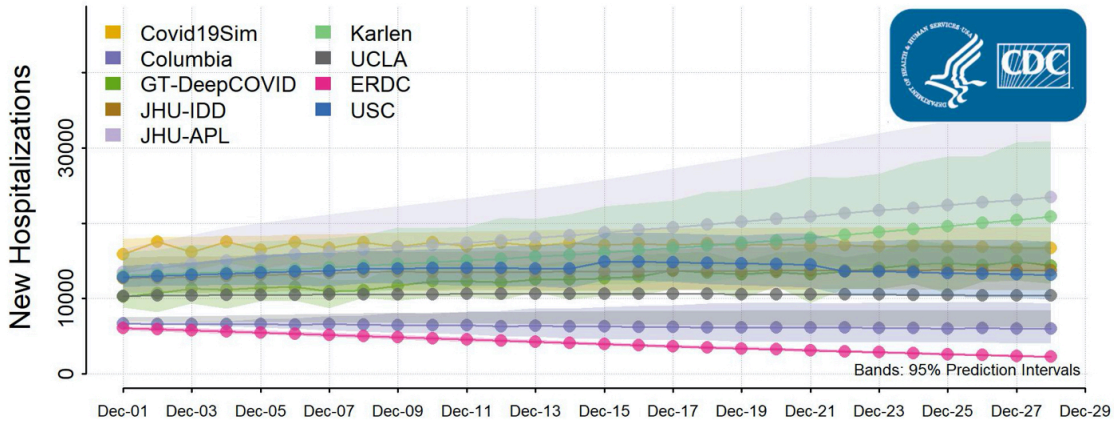
- Think (very hard) about your data generating mechanism
- Explore the design space to come up with reasonable auxiliary assumptions
- Account for all sources of evidence and all sources of uncertainty

Accounting for evidence and uncertainty in the modelling of infectious diseases

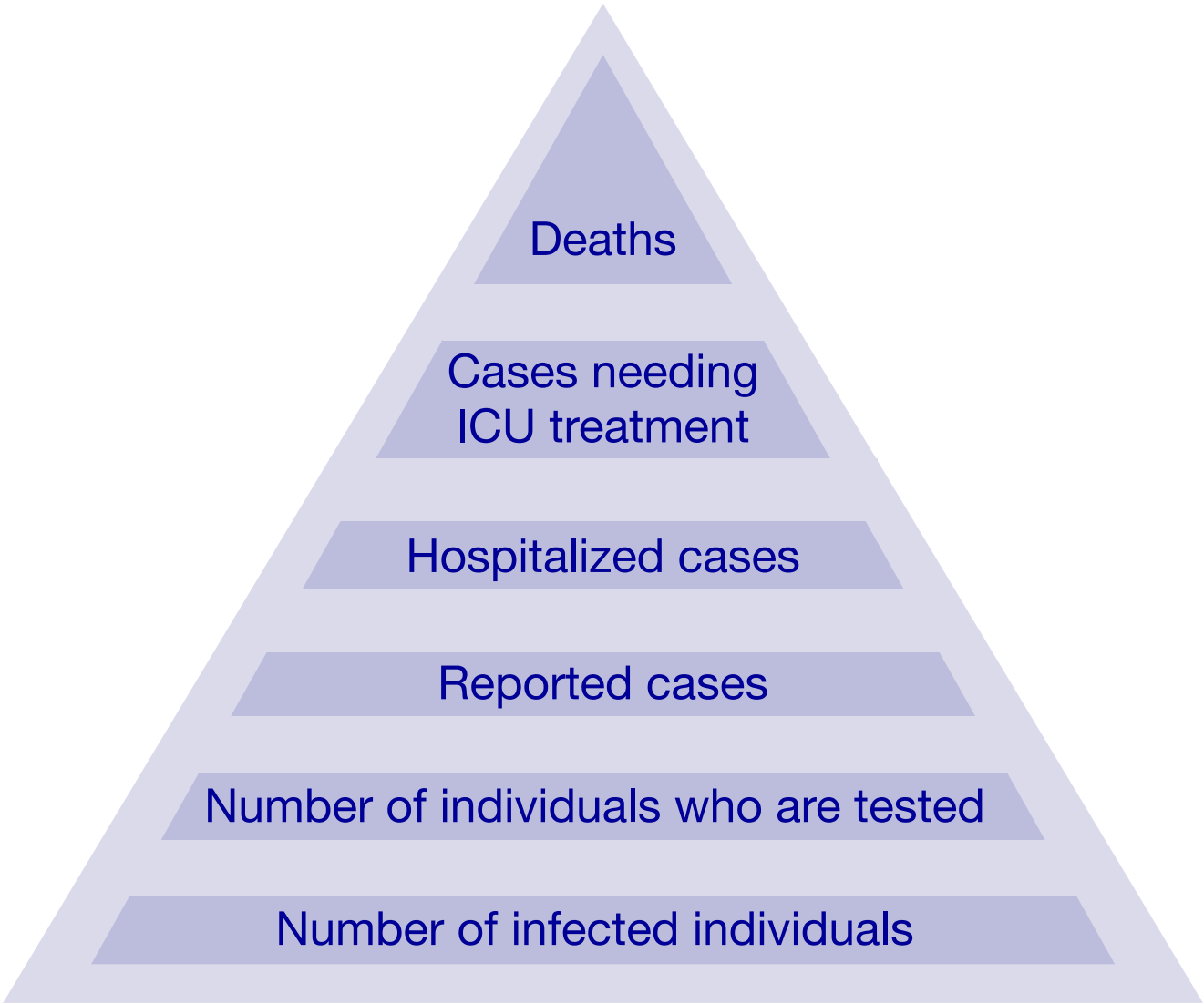
COVID-19 Forecasts from the CDC for the US



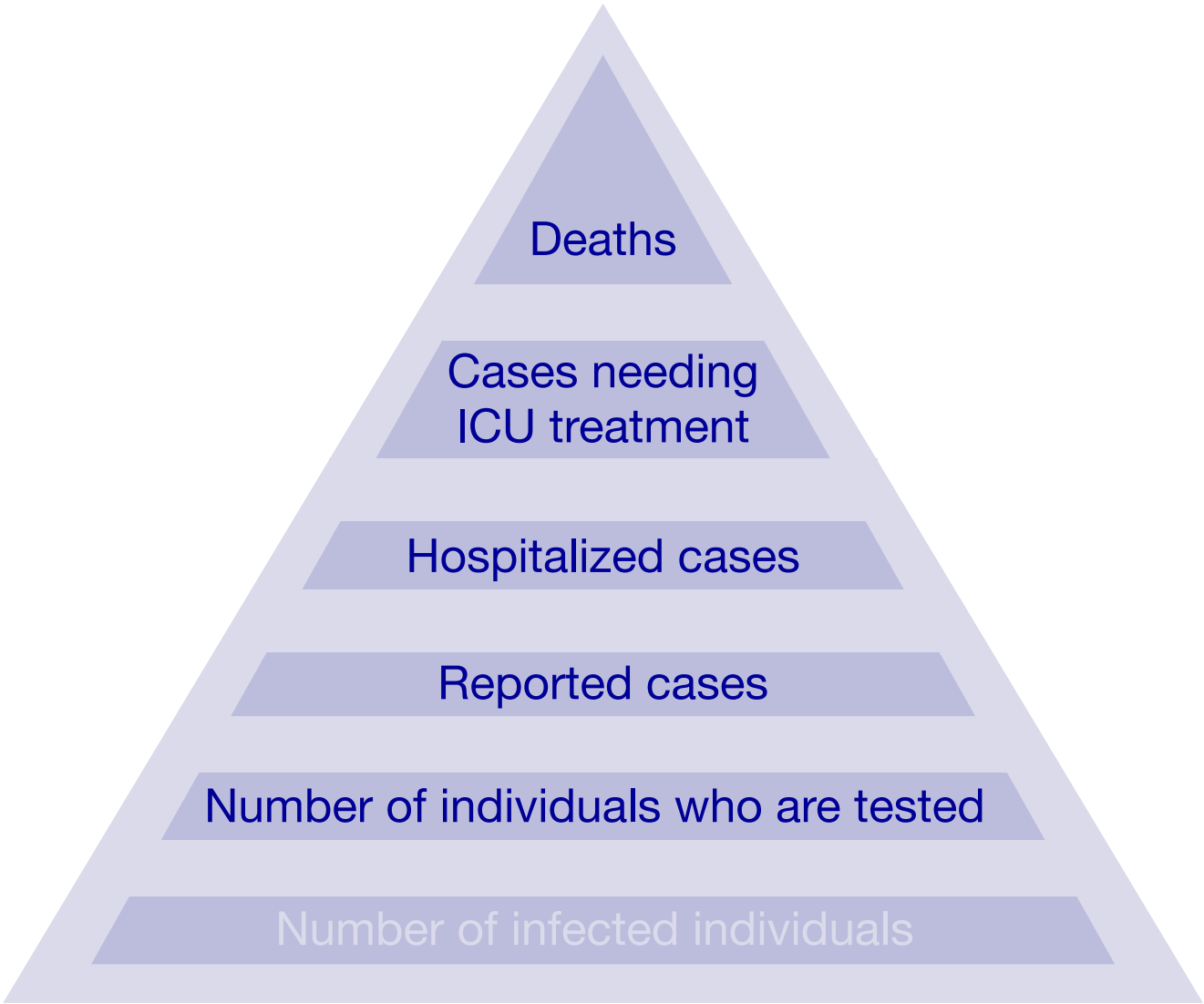




Modelling of infectious diseases

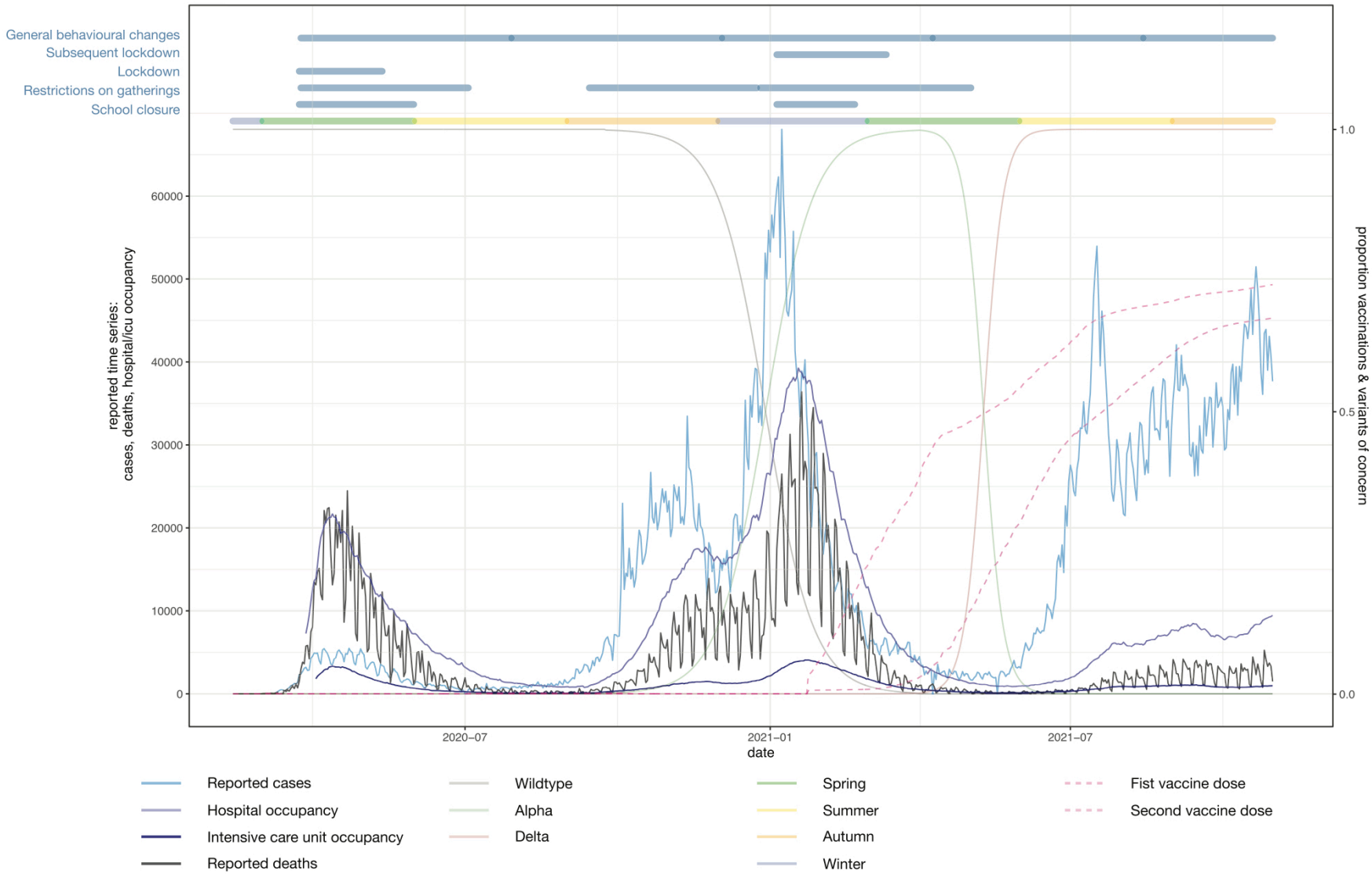


Modelling of infectious diseases



Auxiliary assumptions in the modelling of infectious diseases

- The number of reported cases, the number of hospitalized cases or the number of deaths is proportional to the number of infections

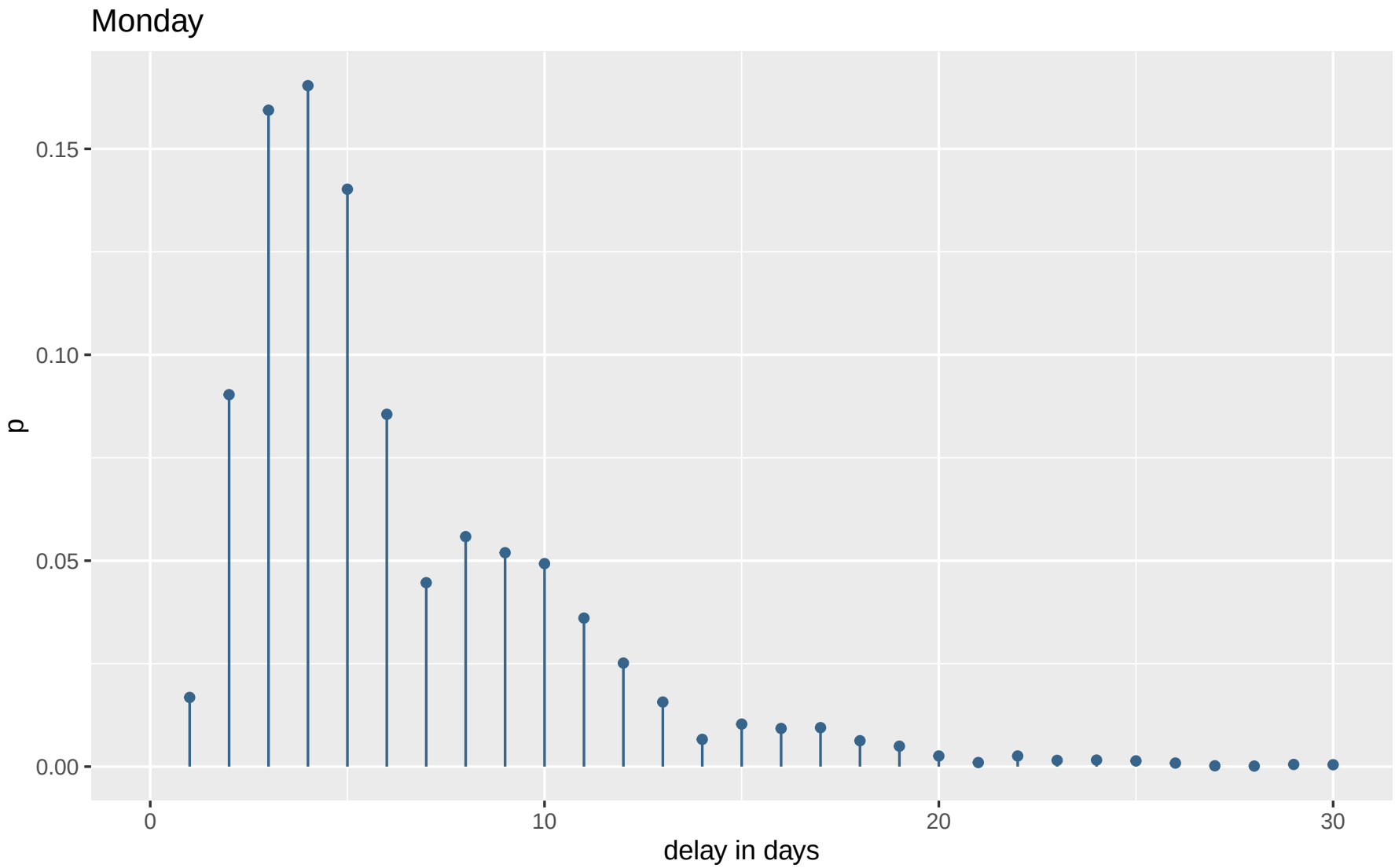


[Rehms et al., 2024, Khazaei et al., 2023]

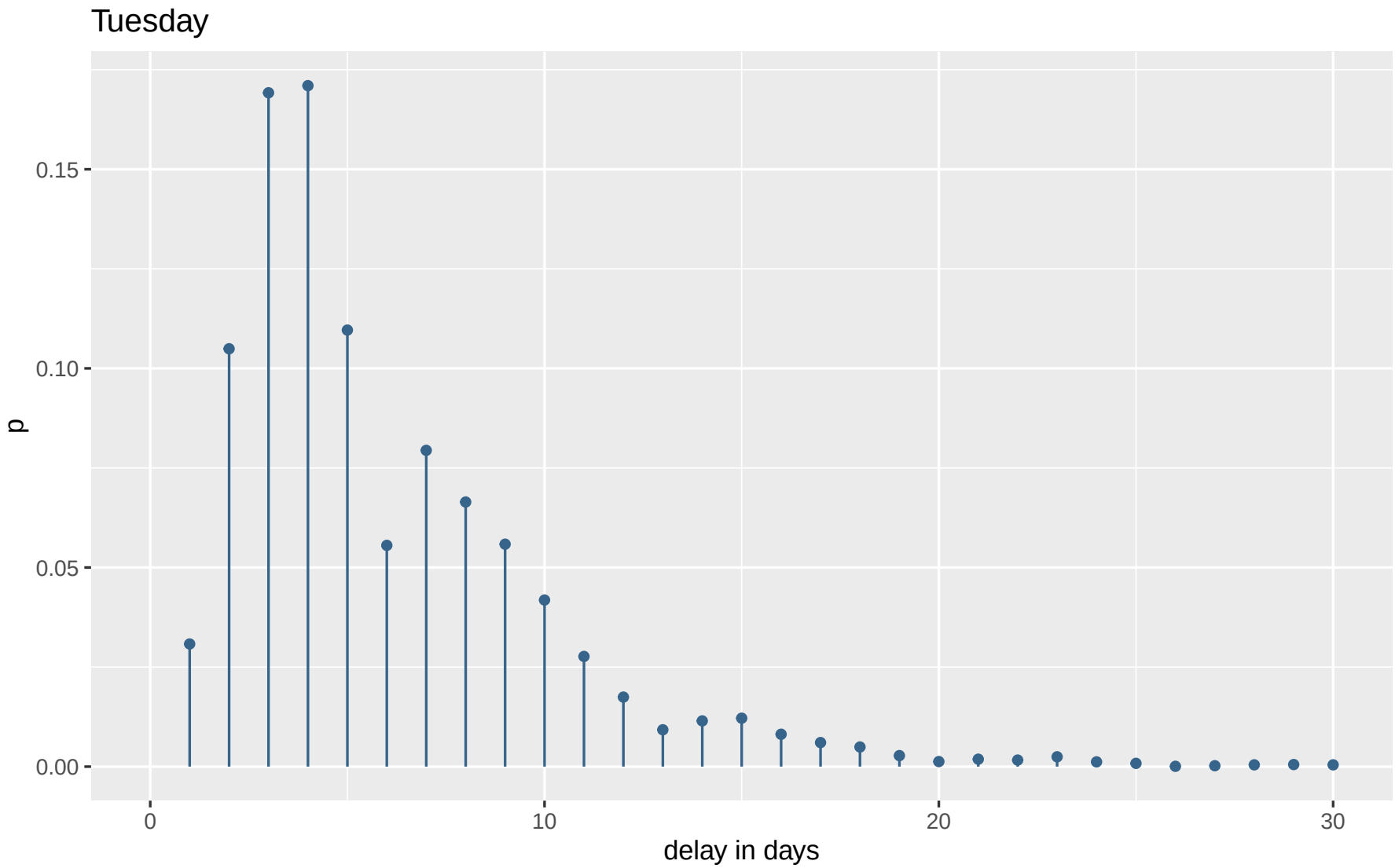
Auxiliary assumptions in the modelling of infectious diseases

- The number of reported cases, the number of hospitalized cases or the number of deaths is proportional to the number of infections
- The time between symptom onset and reporting, hospital admission and death are always the same and known perfectly

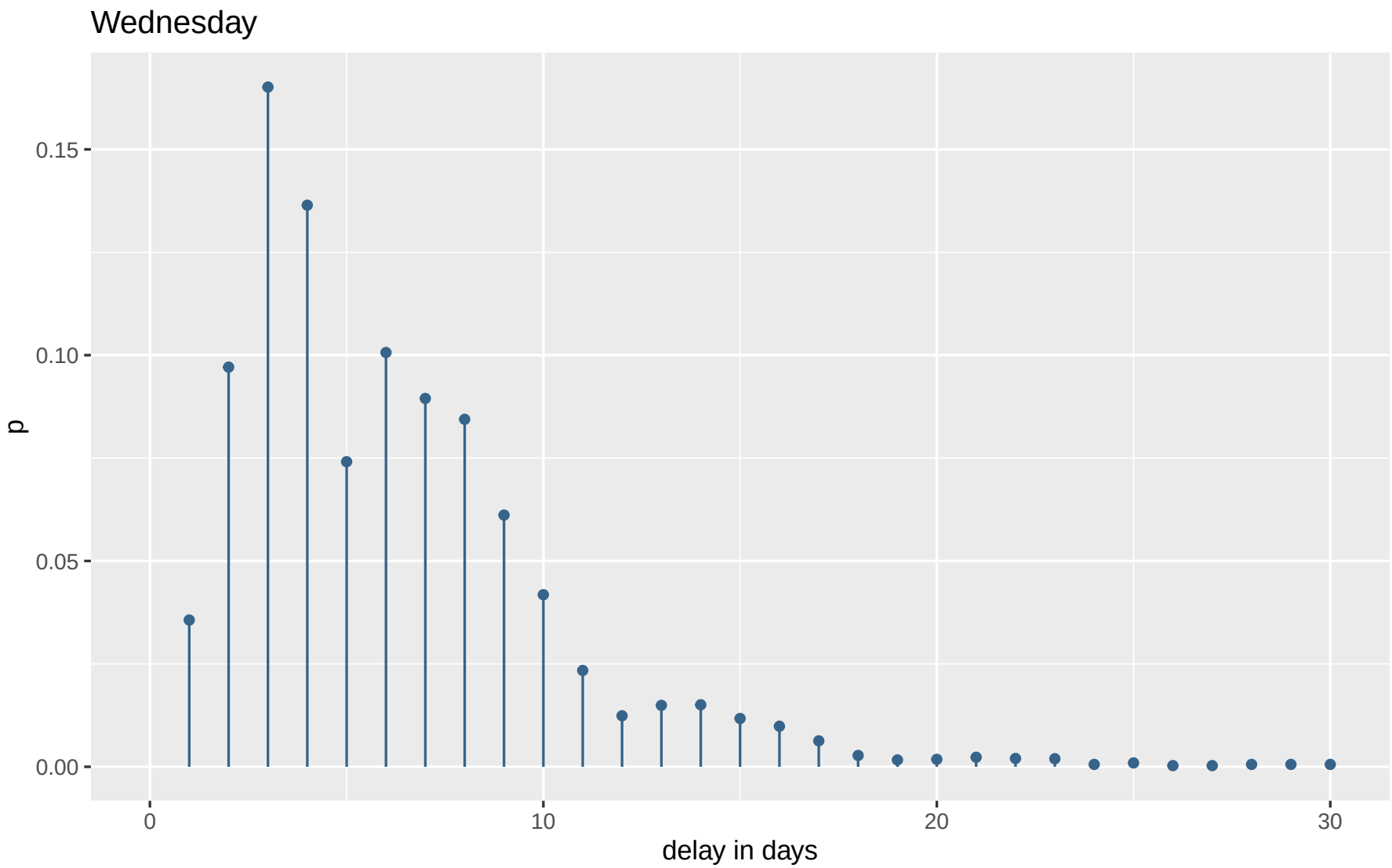
External information to relax auxiliary assumptions



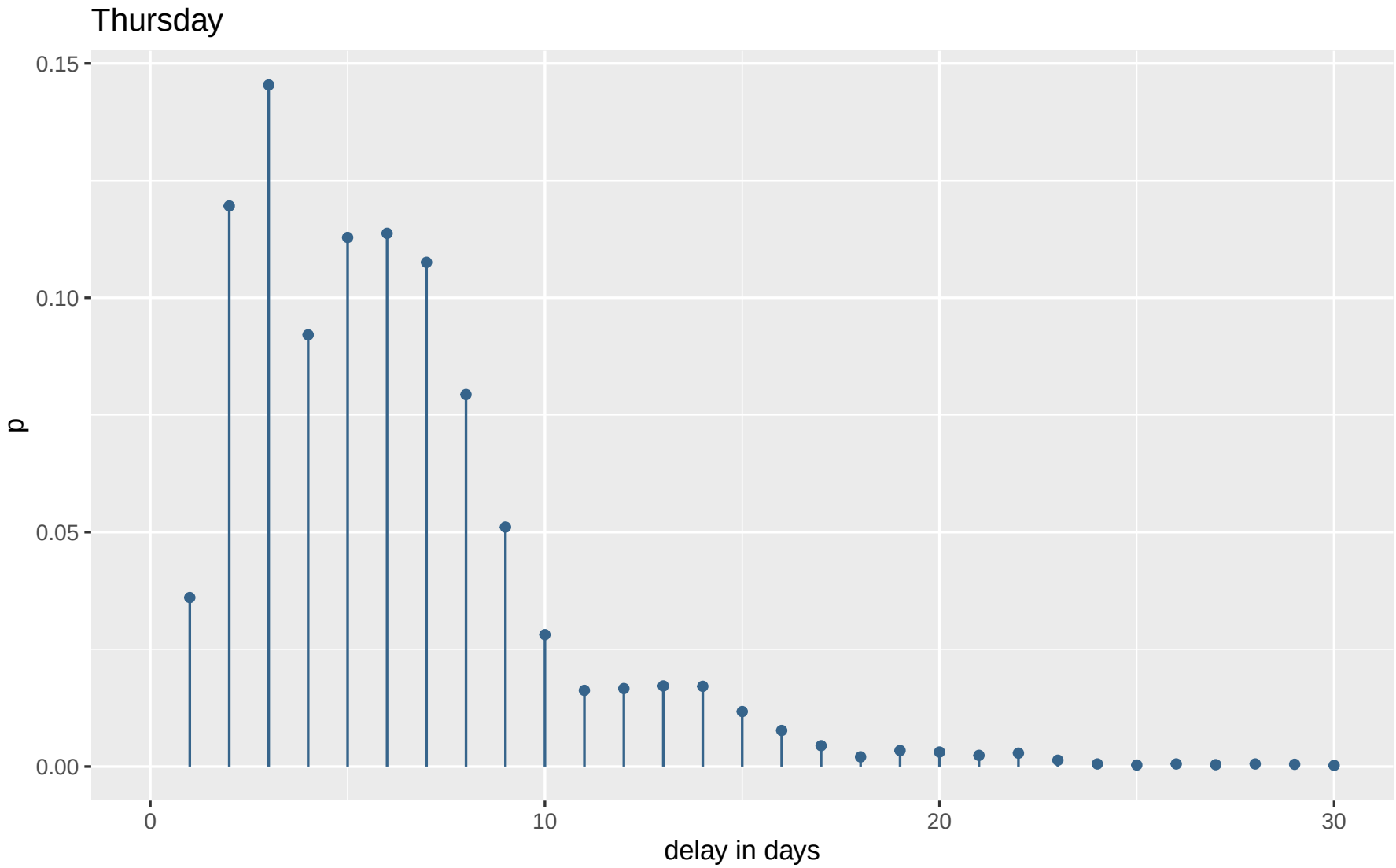
External information to relax auxiliary assumptions



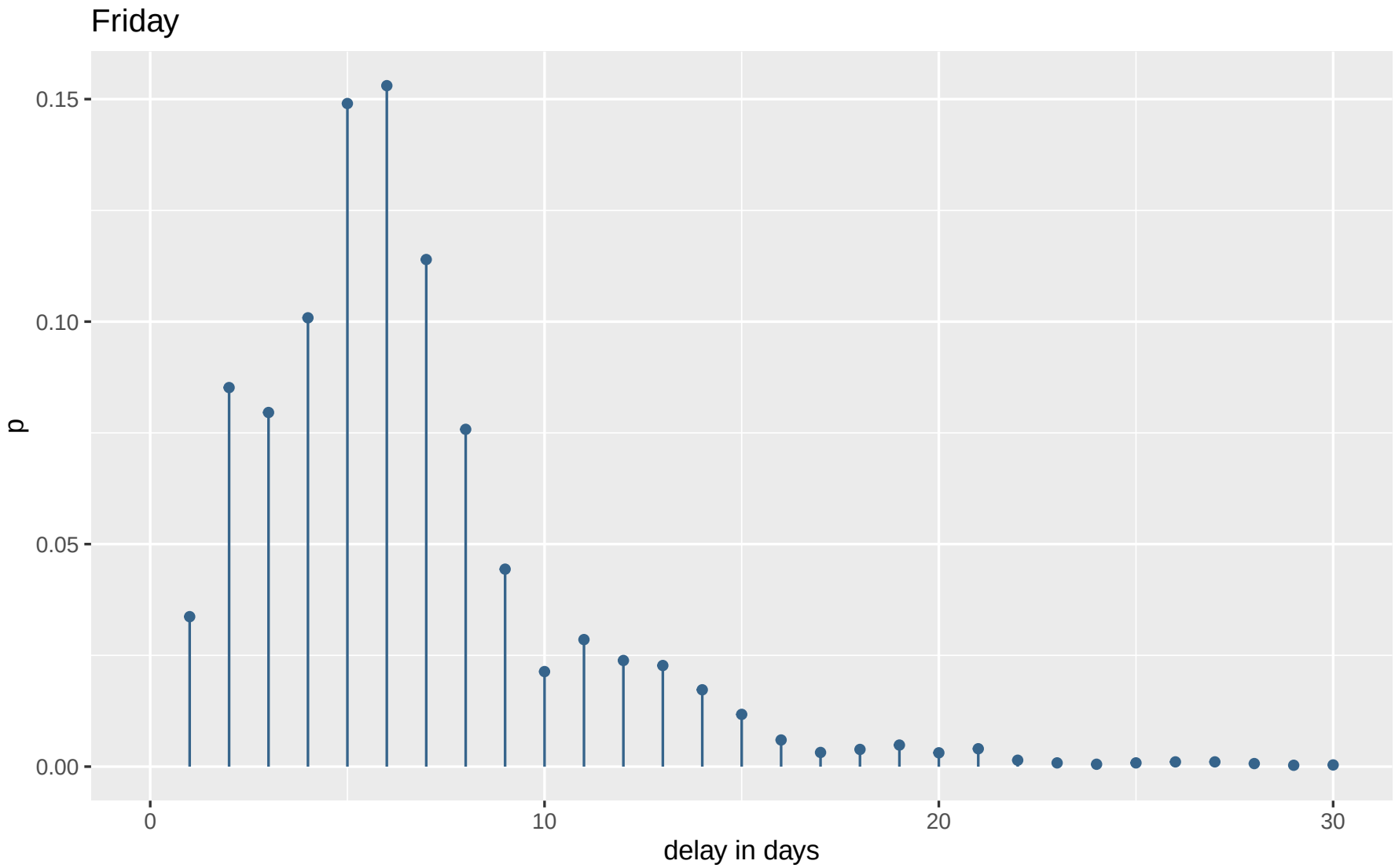
External information to relax auxiliary assumptions



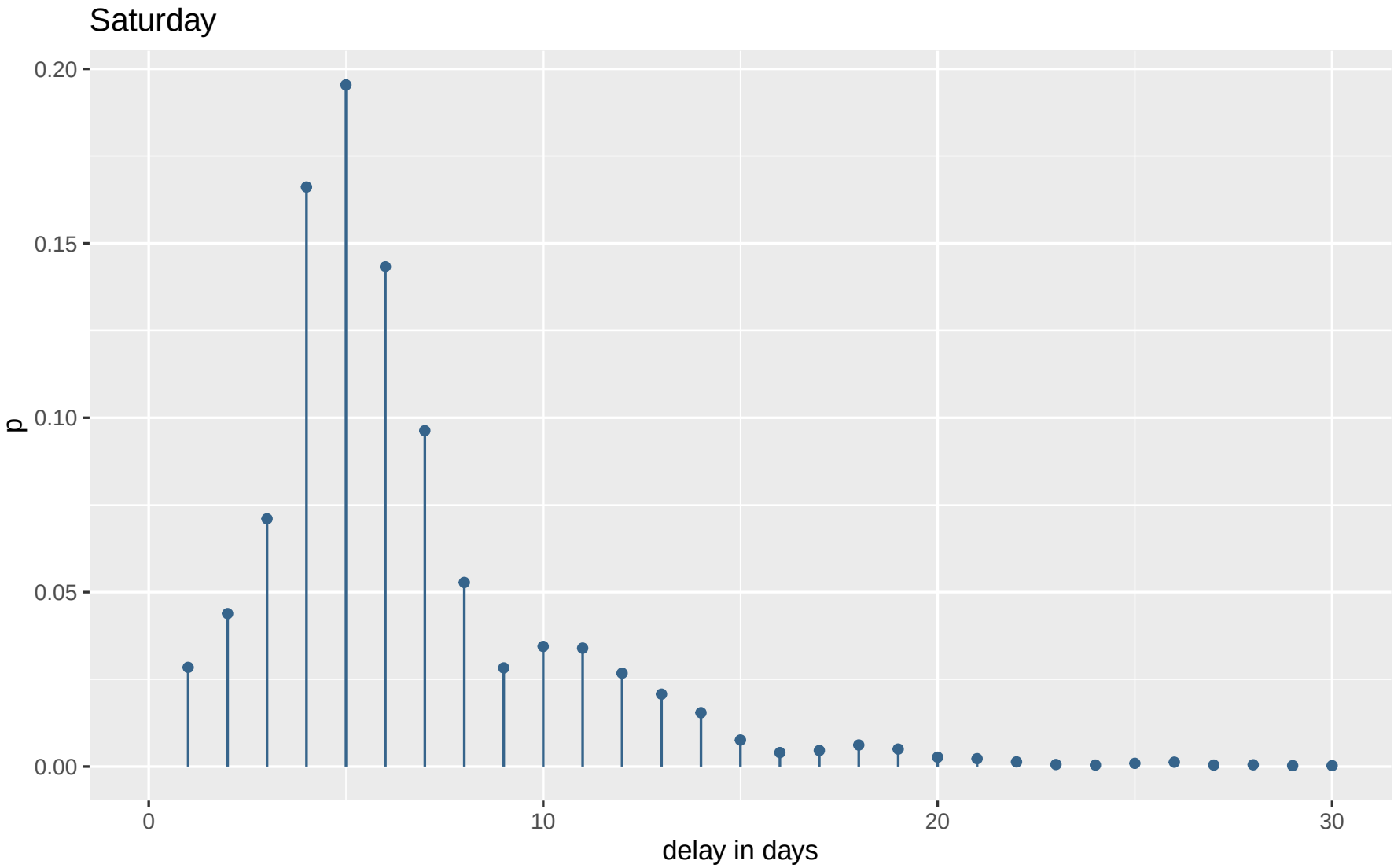
External information to relax auxiliary assumptions



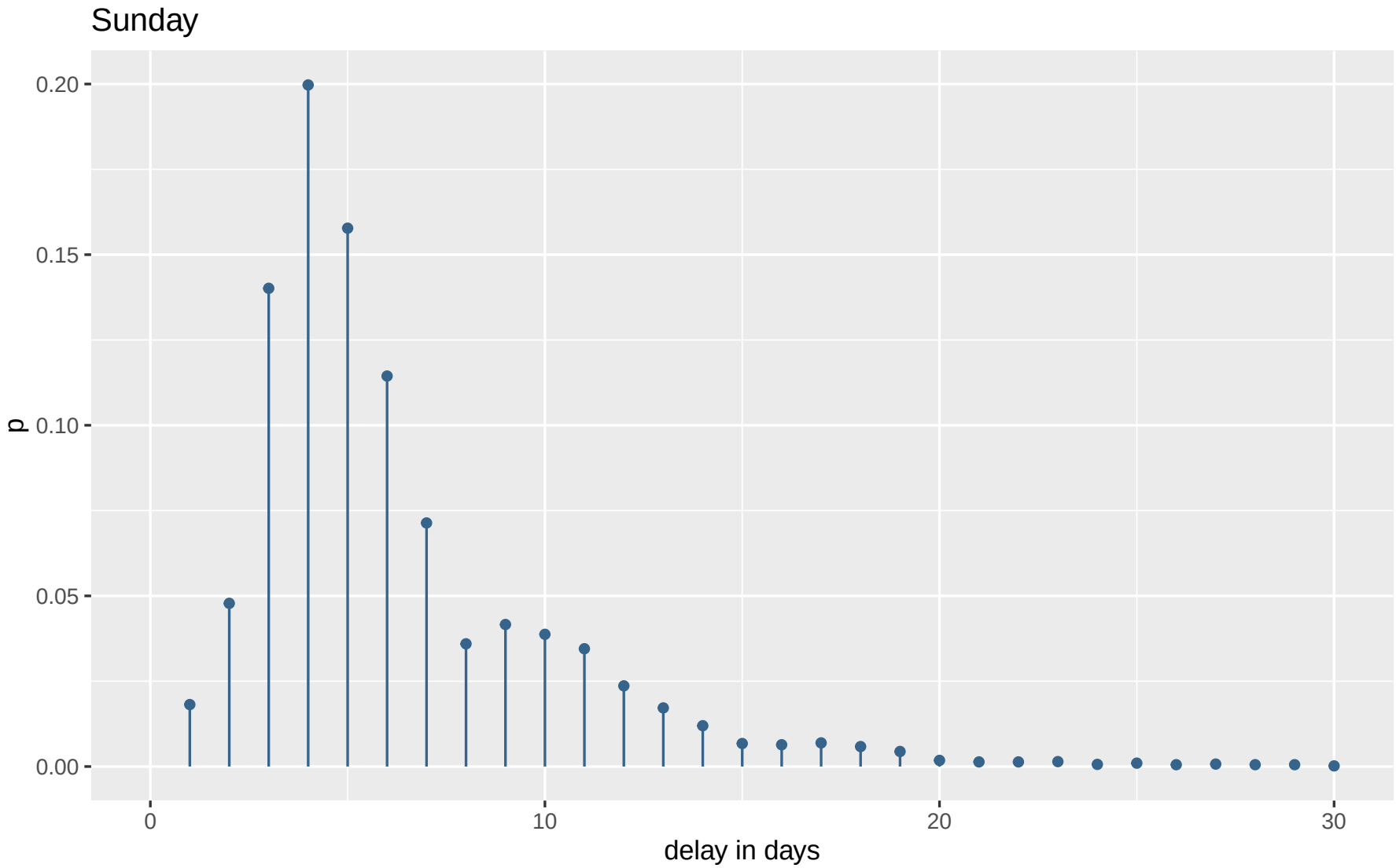
External information to relax auxiliary assumptions



External information to relax auxiliary assumptions



External information to relax auxiliary assumptions



Auxiliary assumptions in the modelling of infectious diseases

- The number of reported cases, the number of hospitalized cases or the number of deaths is proportional to the number of infections
- The time between symptom onset and reporting, hospital admission and death are always the same and known perfectly
- The incubation time and the generation time are always the same and known perfectly

Relaxing auxiliary assumptions

- Account for sources of uncertainty:
  - underreporting and weekday specific reporting delay in the number of cases
  - account for uncertainty in the estimates of infection fatality rates, in the incubation time distribution and the generation time distribution

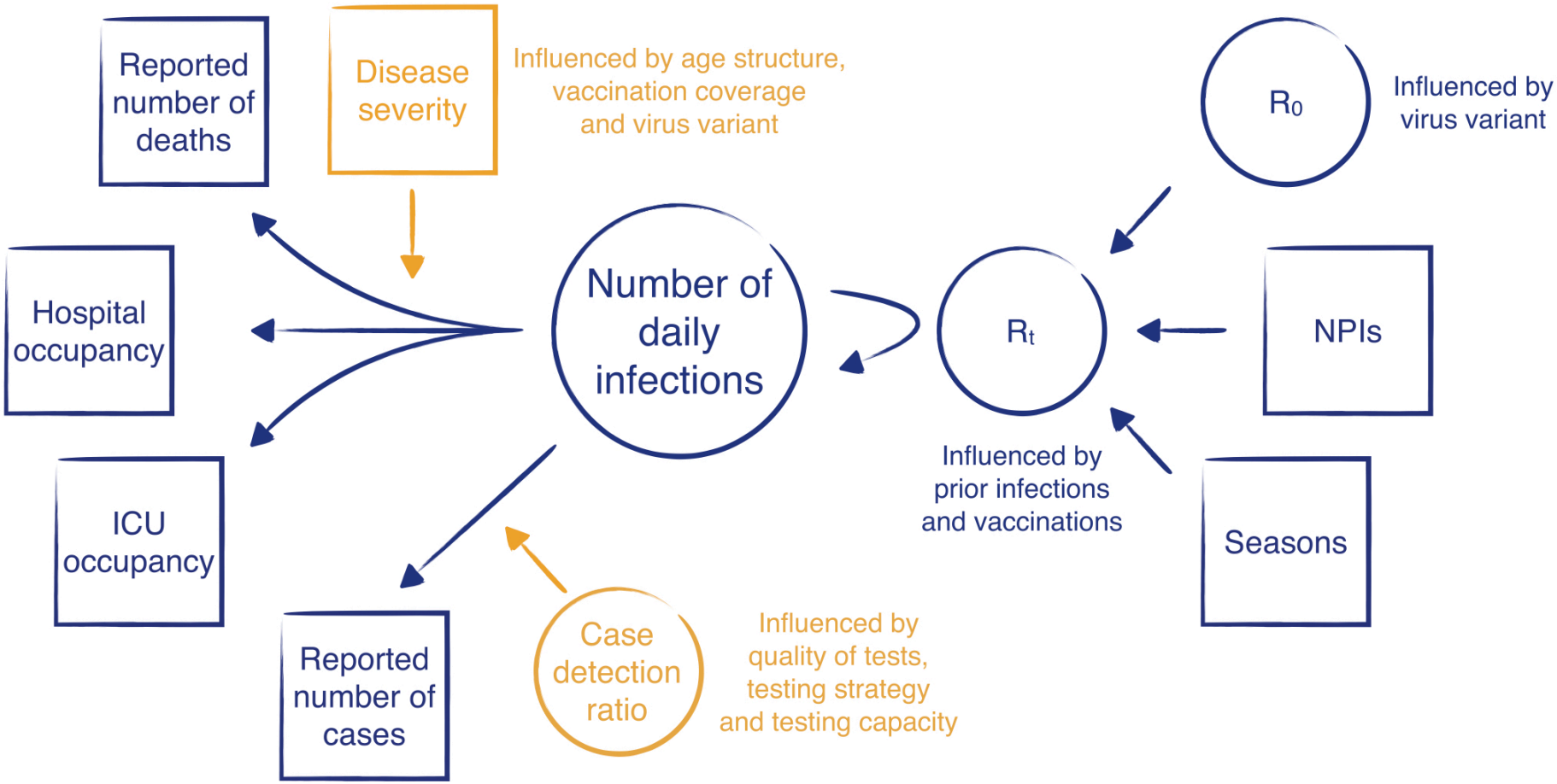
Relaxing auxiliary assumptions

- Account for sources of uncertainty:
  - underreporting and weekday specific reporting delay in the number of cases
  - account for uncertainty in the estimates of infection fatality rates, in the incubation time distribution and the generation time distribution
- Integrate information:
  - on reported number of cases, the number of deaths and hospital occupancy
  - on the prevalence of new variants and vaccination coverage
  - from different geographical regions

Relaxing auxiliary assumptions

- Account for sources of uncertainty:
  - underreporting and weekday specific reporting delay in the number of cases
  - account for uncertainty in the estimates of infection fatality rates, in the incubation time distribution and the generation time distribution
- Integrate information:
  - on reported number of cases, the number of deaths and hospital occupancy
  - on the prevalence of new variants and vaccination coverage
  - from different geographical regions

$\Rightarrow$ We do not need to focus on single countries and short time periods during which these factors remained unchanged



[Rehms et al., 2024, Khazaei et al., 2023]

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 reporting model:
  $C^R_{t,m} \sim$
  Negative Binomial $\left(\rho_{t,m}\pi^{NC}_t \sum_{u<t} C_{u,m}(F_{\xi^R_m}(t-u+1) - F_{\xi^R_m}(t-u)), \phi\right.$

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

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

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

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

Joint posterior

Target: Joint posterior distribution of $\theta =$
$\{\boldsymbol{\alpha}_k, \boldsymbol{\sigma}_{\alpha_k}, R_0, \sigma_R, \tau, R_{0,m}, \boldsymbol{\alpha}_{k,m}, \rho_{t,m}, \beta^{VOC}, \boldsymbol{\beta}^R_m, \boldsymbol{\beta}^D_m, I_{m,t}, \phi_I, \phi_R, \phi_D, \phi_H, \pi^H_m\}$

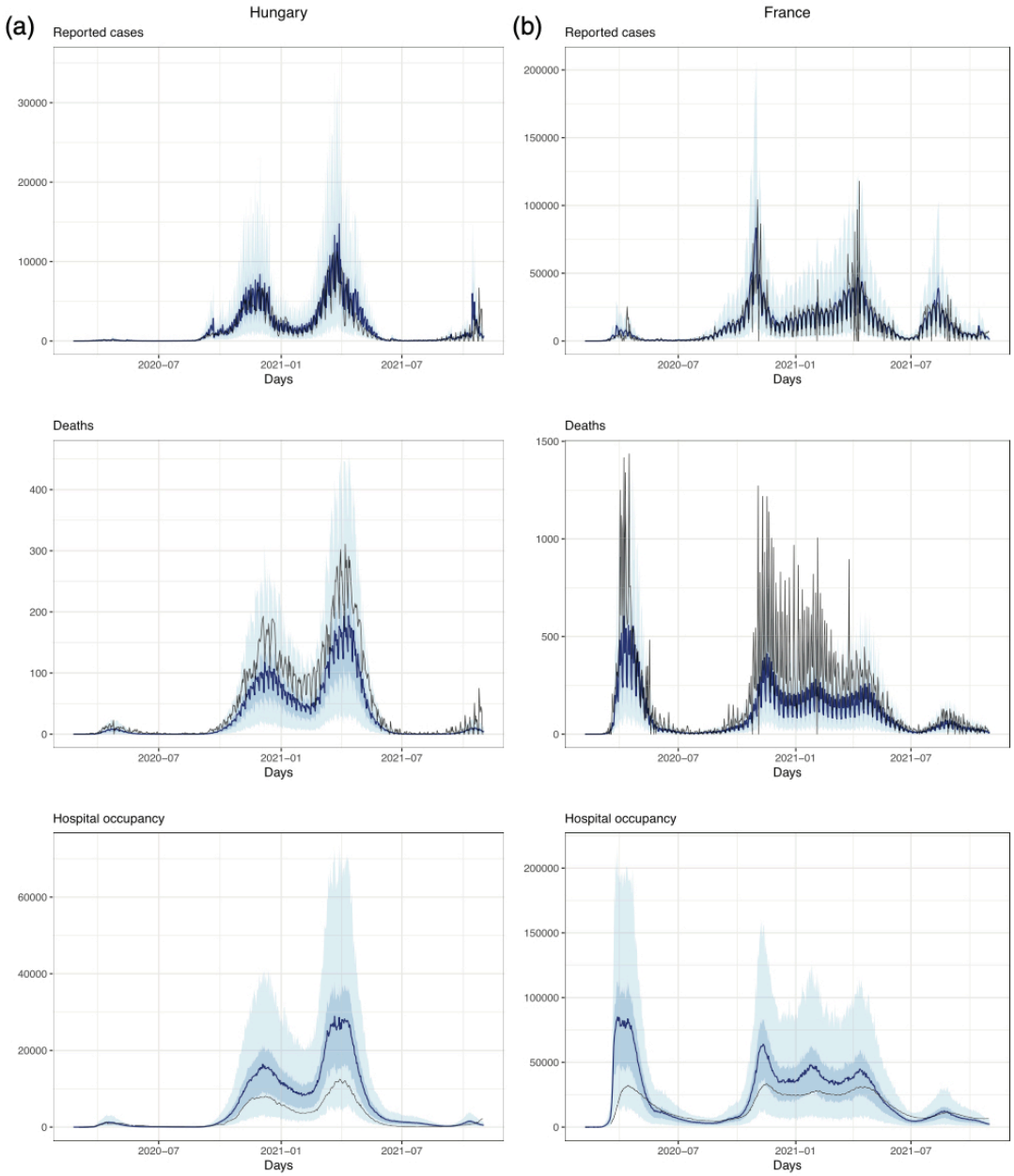
$$[\theta|\mathcal{D}] \propto \prod_{k=1}^{K}[\boldsymbol{\alpha}_k]\,[\sigma_{\alpha_k}] \prod_{m=1}^{M}\prod_{k=1}^{K}[\alpha_{k,m}|\boldsymbol{\alpha}_k, \sigma_{\alpha_k}] \times [R_0][\sigma_R][\beta^{VOC}] \prod_{m=1}^{M}[R_{0,m}|R_0, \sigma_R, \beta^{VOC}]$$

$$\times [\tau] \prod_{m=1}^{M}\prod_{t=1}^{T}[I_{t,m}|R_{0,m}, \boldsymbol{\alpha}, \ln_m(t), I^{hist}_m(t), \beta^{VOC}, \gamma]$$

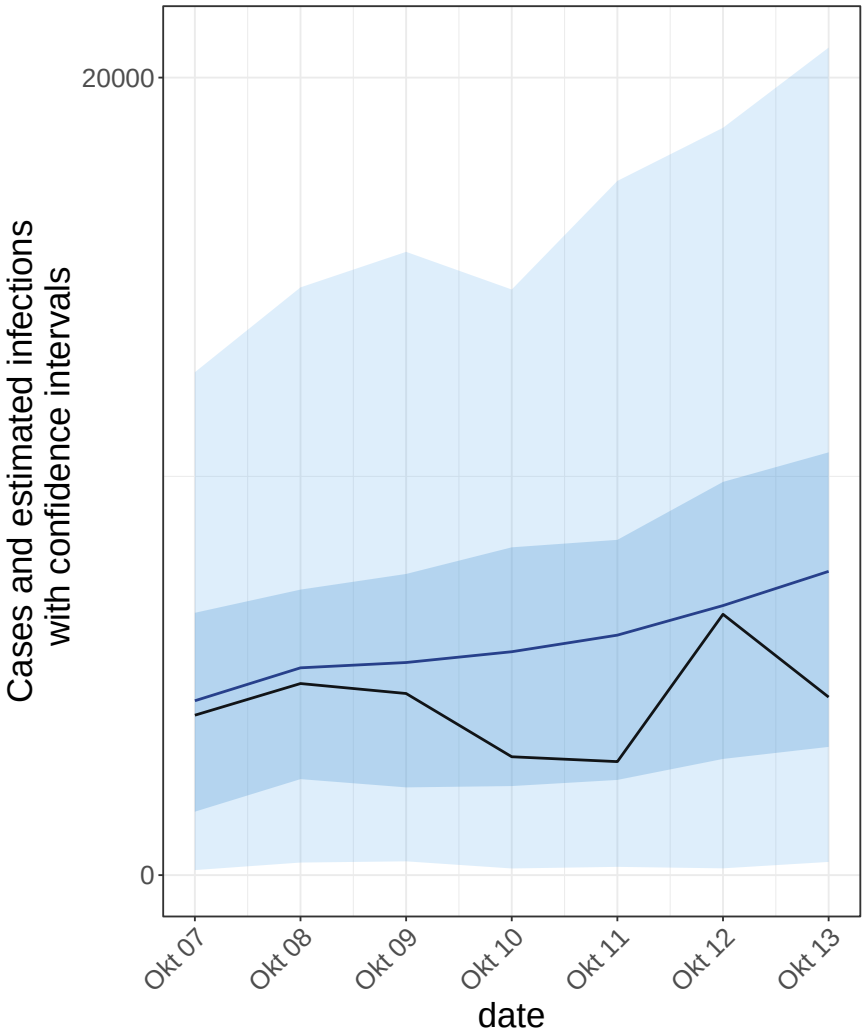
$$\times \prod_{m=1}^{M}\prod_{t=1}^{T}[C_{t,m}|I^{hist}_{t,m}, \xi^C] \times \prod_{m=1}^{M}[\boldsymbol{\beta}^R_m] \prod_{m=1}^{M}\prod_{t=1}^{T}[C^R_{t,m}|C_{t,m}, \rho_{t,m}|\boldsymbol{\beta}^R_m]$$

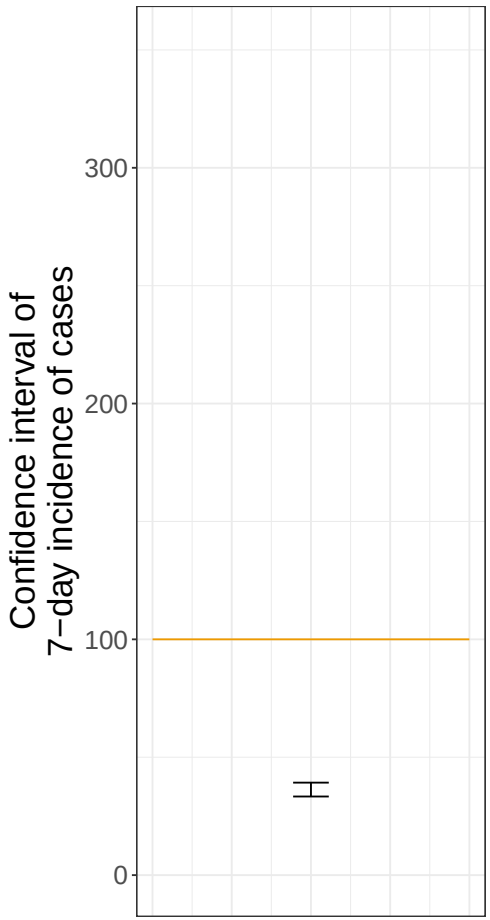
$$\times \prod_{m=1}^{M}[\boldsymbol{\beta}^D_m] \prod_{m=1}^{M}\prod_{t=1}^{T}[D_{t,m}|C^{hist}_{t,m}, C_{t,m}, \rho_{t,m}, \boldsymbol{\beta}^D_m]$$

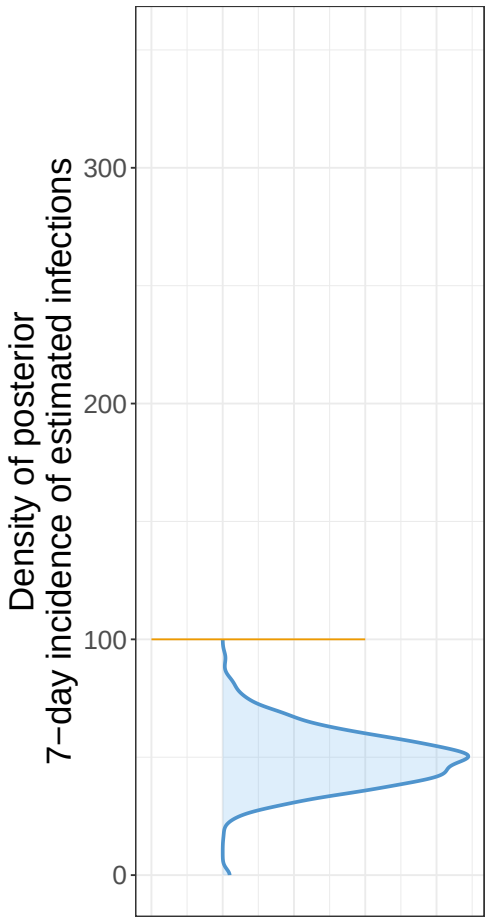
$$\times \prod_{m=1}^{M}\prod_{t=1}^{T}[H_{t,m}|C_{t,m}, C^{hist}_{t,m}, \xi^H, \pi^H]$$


(a)
Hungary
Reported cases
30000
20000
10000
0
2020–07
2021–01
2021–07
Days
Deaths
400
300
200
100
0
2020–07
2021–01
2021–07
Days
Hospital occupancy
60000
40000
20000
0
2020–07
2021–01
2021–07
Days
(b)
France
Reported cases
200000
150000
100000
50000
0
2020–07
2021–01
2021–07
Days
Deaths
1500
1000
500
0
2020–07
2021–01
2021–07
Days
Hospital occupancy
200000
150000
100000
50000
0
2020–07
2021–01
2021–07
Days

Improving the communication of evidence and uncertainty

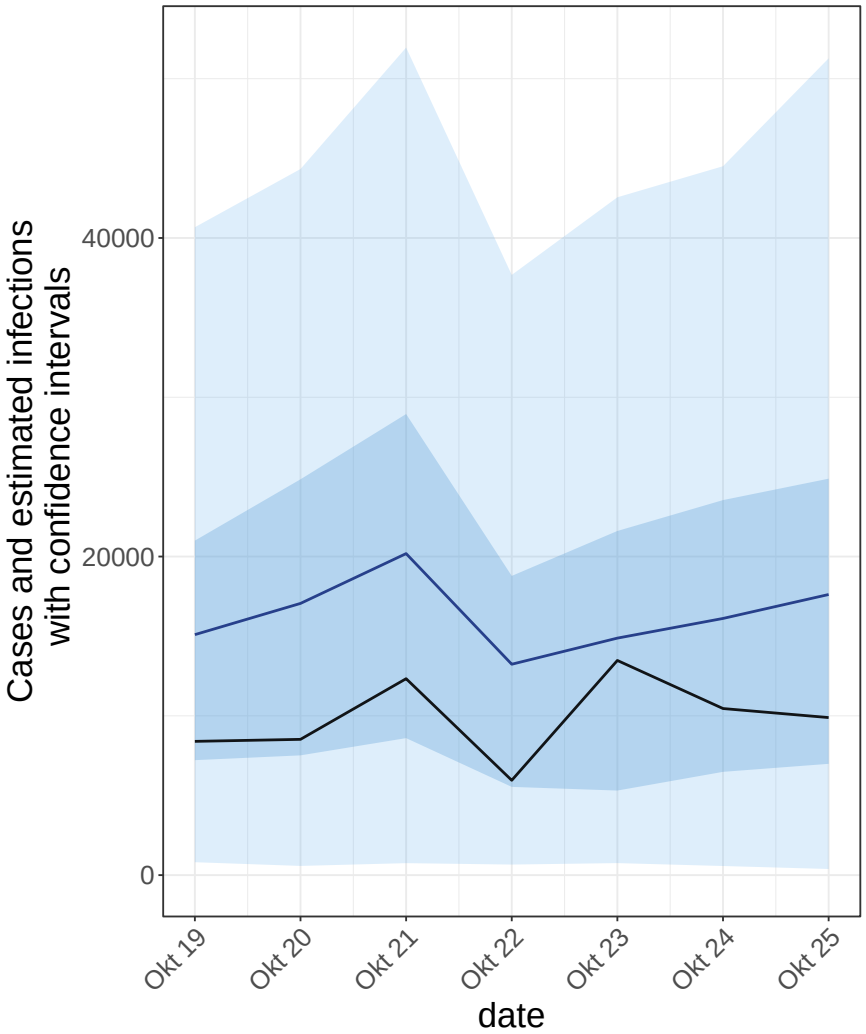


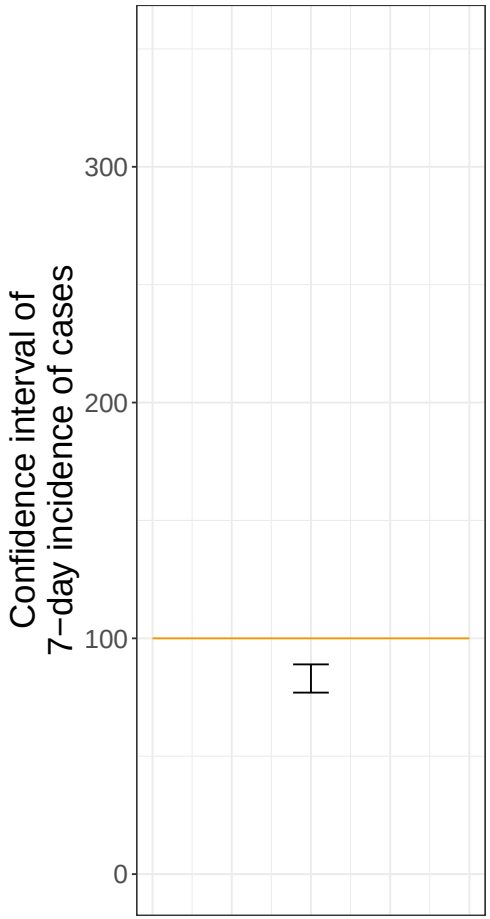


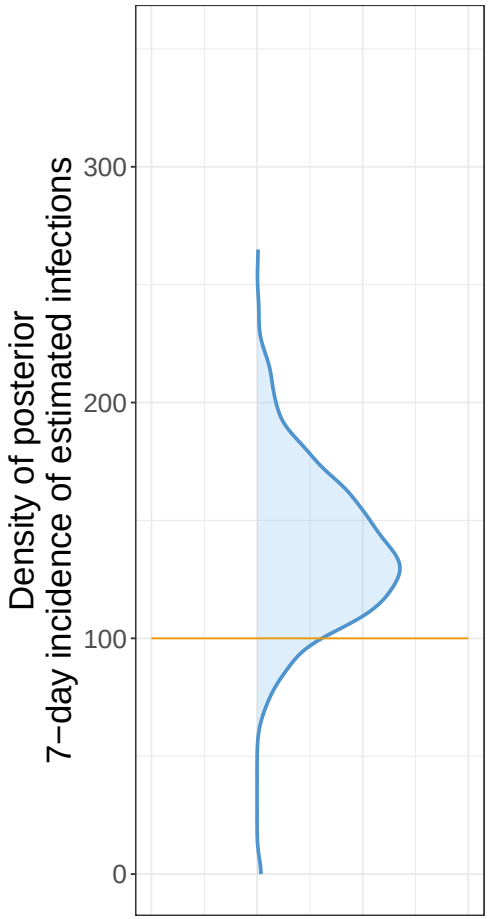


[Rehms et al., 2024, Khazaei et al., 2023]

Improving the communication of evidence and uncertainty

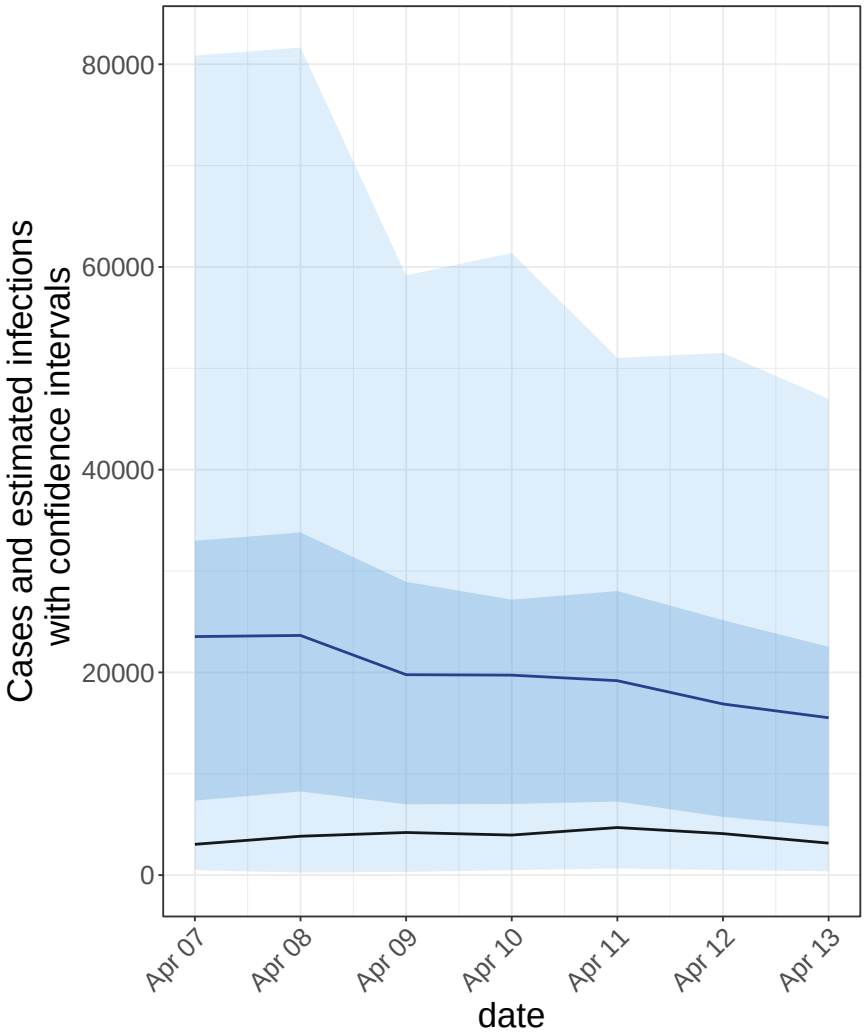


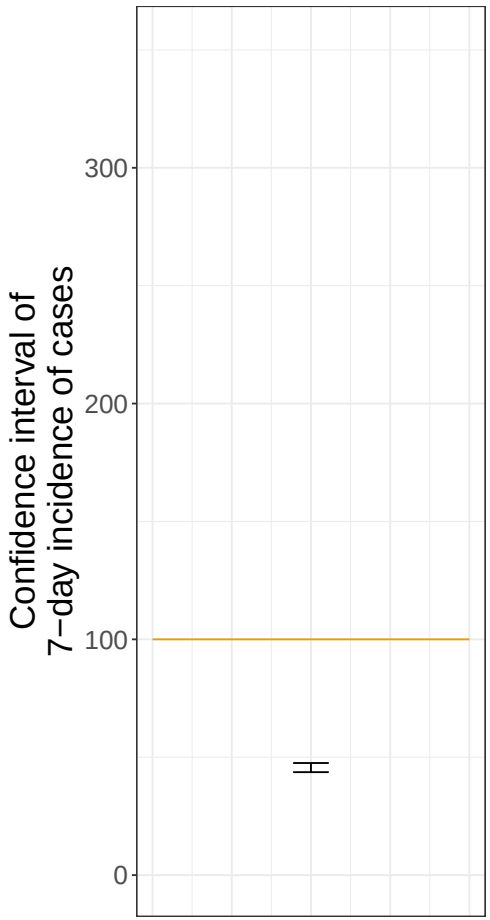


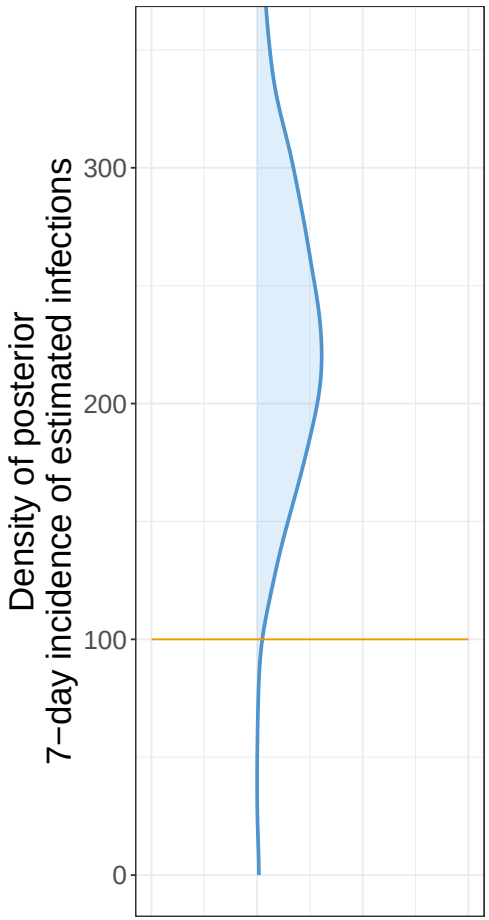


[Rehms et al., 2024, Khazaei et al., 2023]

Improving the communication of evidence and uncertainty



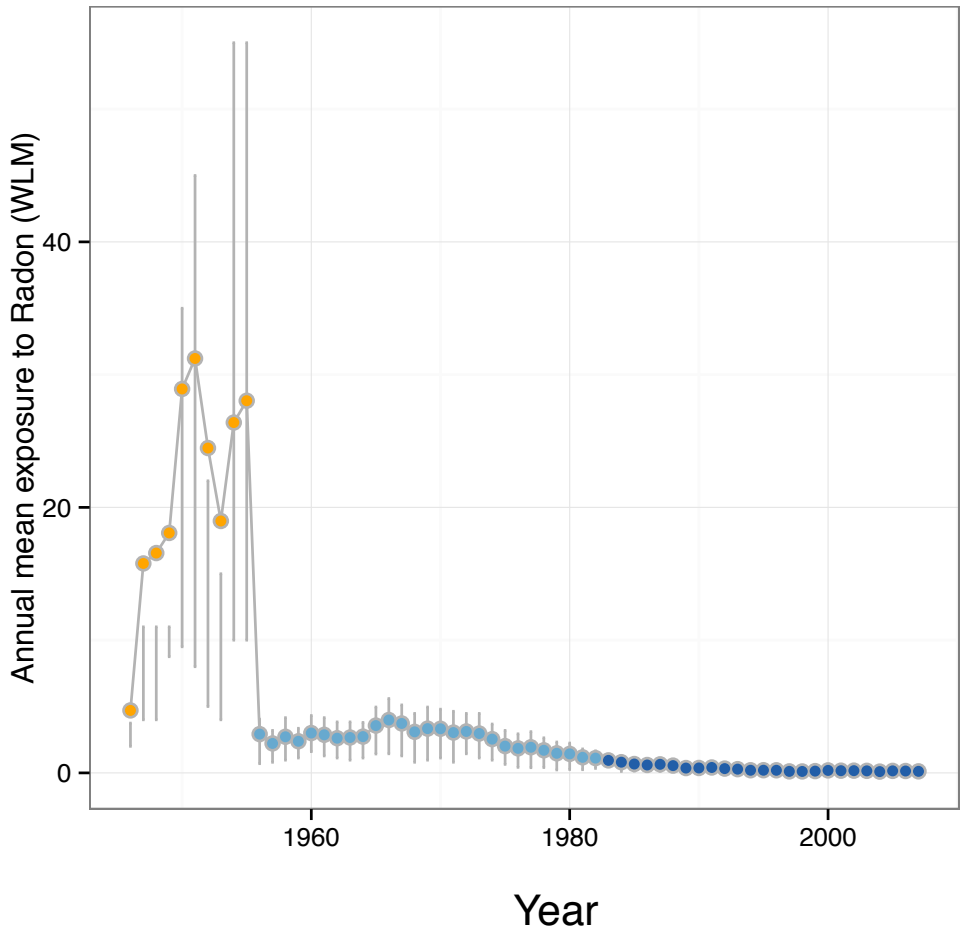




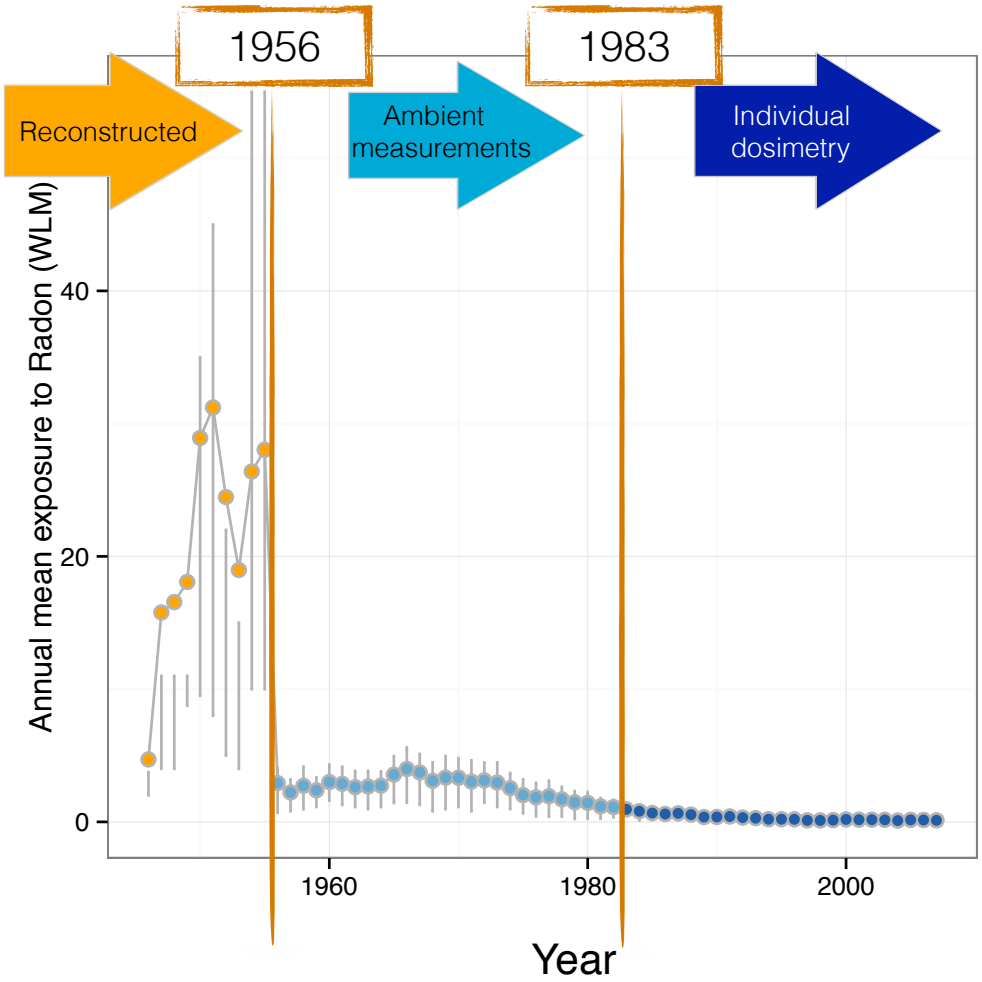
[Rehms et al., 2024, Khazaei et al., 2023]

Accounting for exposure measurement error in occupational cohorts

Exposure measurement error in occupational cohorts

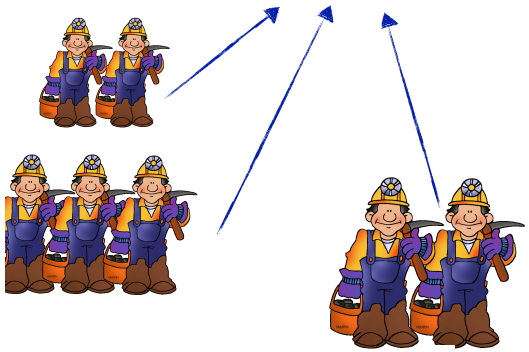


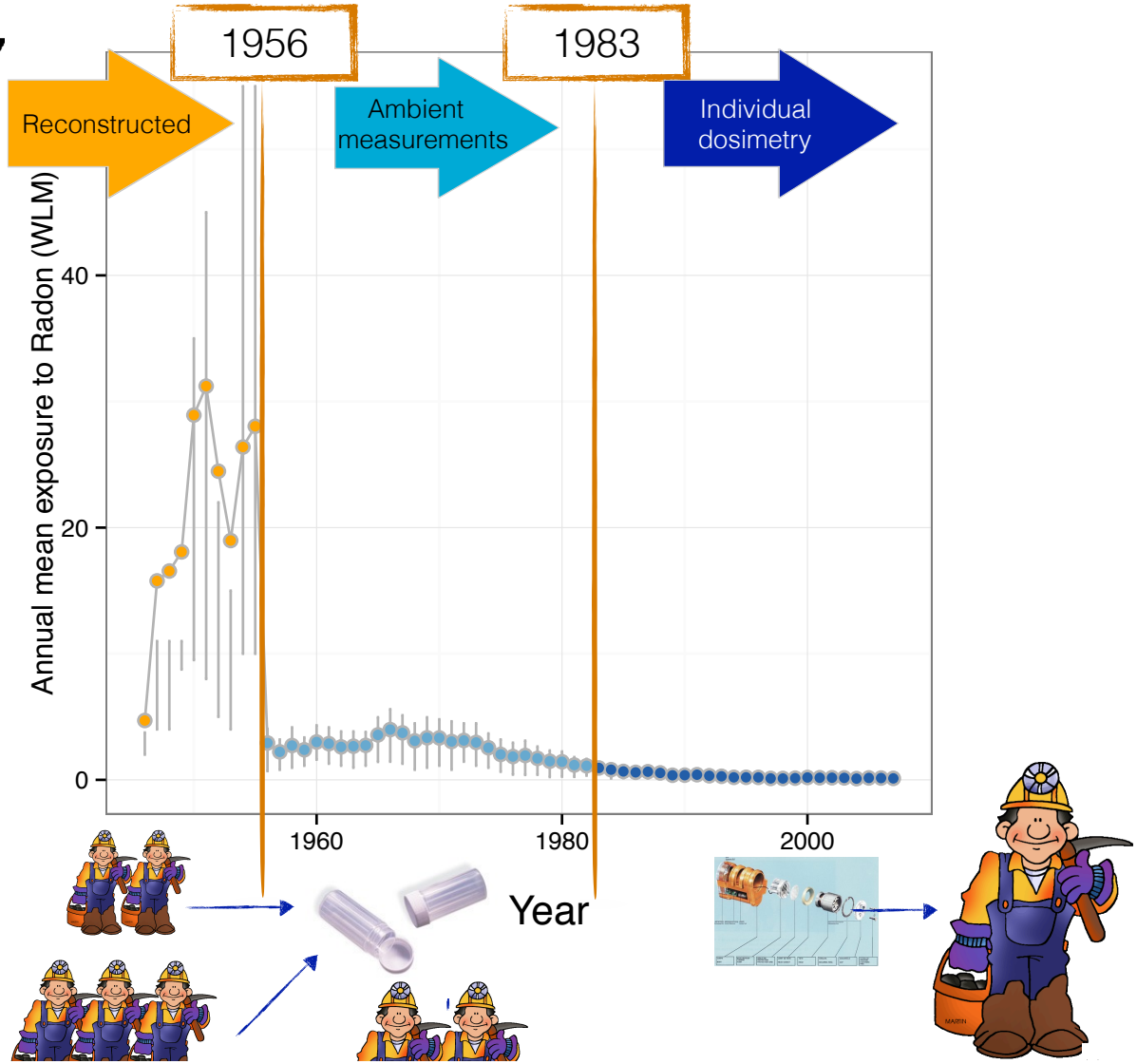
Exposure measurement error in occupational cohorts



Exposure measurement error in occupational cohorts

MINE	Henriette	Brugeaud	Margnac	Fanay-Sagnes	Ecarpière	Chapelle
De 1947 à 1950	10					
1951	10	2		2		
1952	10	2		5		
1953	10	5	5	5	5	5
1954	10	5	5	5	5	5
1955	10	5	5	5	5	5
1956	1					





Exposure reconstruction in the German cohort of uranium miners

In the German cohort, the exposure to radon was reconstructed through the following formula

$$
\begin{aligned}
E(t,o,j) &= \overline{\overline{\overline{C}}}_{Rn}(t,o) \cdot 12 \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j) \\
X_i(t) &= E(t,o,j) \cdot U_i(t) \\
\hat{\overline{\overline{\overline{C}}}}_{Rn}(t,o) &= \overline{\overline{\overline{C}}}_{Rn}(t,o) \cdot U_c(t,o)
\end{aligned}
$$

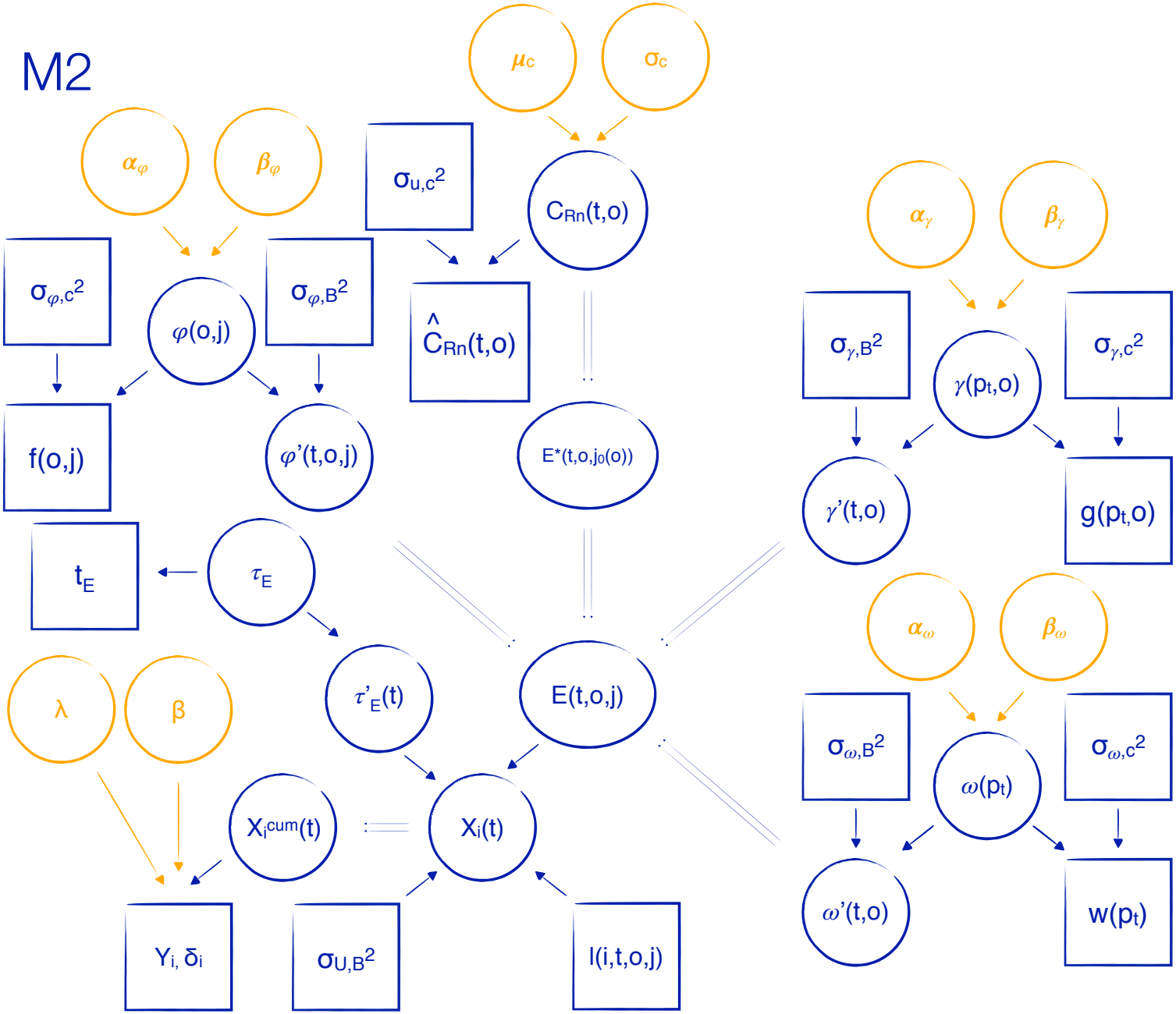
with $U_i(t) \sim \mathcal{LN}\left(\frac{-\sigma^2_{u,B}}{2}, \sigma^2_{u,B}\right)$ and $U_c(t,o) \sim \mathcal{LN}\left(\frac{-\sigma^2_{u,c}}{2}, \sigma^2_{u,c}\right)$

Exposure reconstruction in the German cohort of uranium miners

$$f(o,j) = \varphi(o,j) \cdot U_{\varphi,c}(o,j)$$
$$\varphi'(t,o,j) = \varphi(o,j) \cdot U_{\varphi,B}(t,o,j)$$

$$w(p_t) = \omega(p_t) \cdot U_{\omega,c}(p_t)$$
$$\omega'(t,o) = \omega(p_t) \cdot U_{\omega,B}(t,o)$$

$$g(p_t,o) = \gamma(p_t,o) \cdot U_{\gamma,c}(p_t,o)$$
$$\gamma'(t,o) = \gamma(p_t,o) \cdot U_{\gamma,B}(t,o)$$


M2
μ_C
σ_C
α_φ
β_φ
$\sigma_{u,c}^2$
$C_{Rn}(t,o)$
$\sigma_{\varphi,c}^2$
$\varphi(o,j)$
$\sigma_{\varphi,B}^2$
$\hat{C}_{Rn}(t,o)$
f(o,j)
$\varphi'(t,o,j)$
$E^*(t,o,j_0(o))$
t_E
τ_E
$\tau'_E(t)$
E(t,o,j)
λ
β
$X_i^{cum}(t)$
$X_i(t)$
Y_i, δ_i
$\sigma_{U,B}^2$
I(i,t,o,j)
α_γ
β_γ
$\sigma_{\gamma,B}^2$
$\gamma(p_t,o)$
$\sigma_{\gamma,c}^2$
$\gamma'(t,o)$
$g(p_t,o)$
α_ω
β_ω
$\sigma_{\omega,B}^2$
$\omega(p_t)$
$\sigma_{\omega,c}^2$
$\omega'(t,o)$
$w(p_t)$

Joint posterior

$$
\begin{aligned}
[\theta|Y,X] = {} & [\beta][\lambda][\alpha_\omega][\beta_\omega][\alpha_\gamma][\beta_\gamma][\alpha_\varphi][\beta_\varphi][\mu_C][\sigma_C]\times \\
& \prod_{i,t}[Y_i|\lambda,\beta,X_i^{cum}(t)]\times \\
& \prod_{i,t}[X_i(t)|\sigma^2_{U,B},I(i,t,o,j),\varphi'(t,o,j),\gamma'(t,o),\omega'(t,o),\tau'_E(t)]\times \\
& \prod_{t,o}[\omega'(t,o)|\sigma^2_{U,B},\omega(p_t)]\prod_{p_t}[w(p_t)|\sigma^2_{\omega,c},\omega(p_t)]\prod_{p_t}[\omega(p_t)|\alpha_\omega,\beta_\omega]\times \\
& \prod_{t,o}[\gamma'(t,o)|\sigma^2_{\gamma,B},\gamma(p_t,o)]\prod_{p_t,o}[g(p_t,o)|\sigma^2_{\gamma,c},\gamma(p_t,o)]\prod_{p_t,o}[\gamma(p_t,o)|\alpha_\gamma,\beta_\gamma]\times \\
& \prod_{t,o,j}[\varphi'(t,o,j)|\sigma^2_{\varphi,B},\varphi(o,j)]\prod_{o,j}[f(o,j)|\sigma^2_{\varphi,c},\varphi(o,j)]\prod_{o,j}[\varphi(o,j)|\alpha_\varphi,\beta_\varphi]\times \\
& \prod_{t,o}[C_{R_n(t,o)}|\sigma^2_{u,c},\mathcal{C}_{R_n(t,o)}]\prod_{t,o}[\mathcal{C}_{R_n(t,o)}|\mu_C,\sigma_C]
\end{aligned}
$$

Disease model	Correction	Mean	Median	HDI (95%)
Proportional hazards	Corrected	0.16	0.15	[0.11, 0.20]
	Uncorrected	0.12	0.12	[0.11, 0.13]
EHR	Corrected	0.40	0.39	[0.26, 0.54]
	Uncorrected	0.30	0.30	[0.26, 0.35]

Accounting for evidence and uncertainty in the prevalence of drug use

Assumptions in the estimation of the prevalence of drug use

- Drug surveys include a random sample of the target population of interest

Assumptions in the estimation of the prevalence of drug use

- Drug surveys include a random sample of the target population of interest
- Survey participants know their drug consumption perfectly and they give honest answers

Assumptions in the estimation of the prevalence of drug use

- Drug surveys include a random sample of the target population of interest
- Survey participants know their drug consumption perfectly and they give honest answers
- Missing data are missing completely at random

Assumptions in the estimation of the prevalence of drug use

- Drug surveys include a random sample of the target population of interest
- Survey participants know their drug consumption perfectly and they give honest answers
- Missing data are missing completely at random

$\Rightarrow$ Drug surveys capture less than half of alcohol sales and only 30% to 60% of illicit substance use

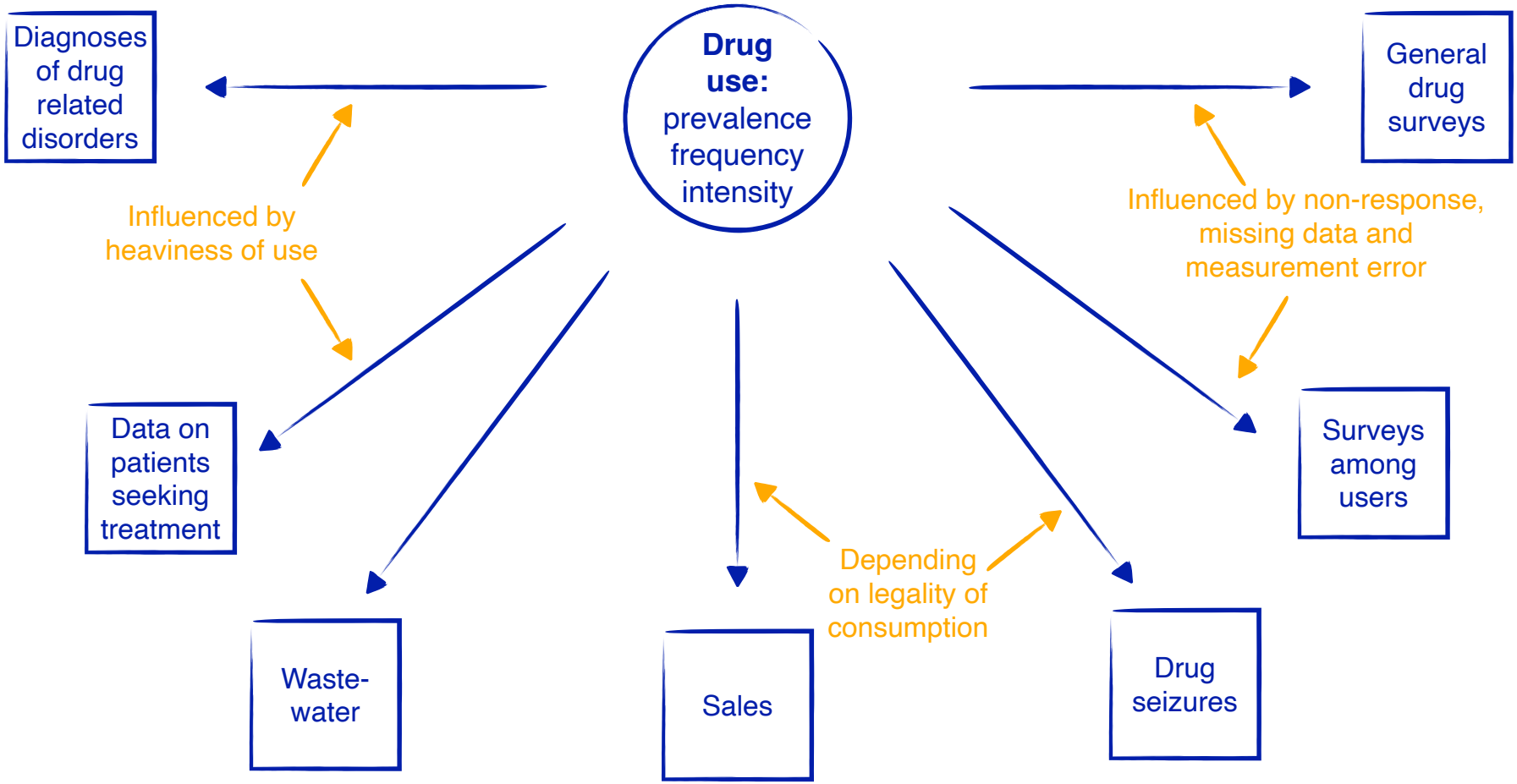
Assumptions in the estimation of the prevalence of drug use

- Drug surveys include a random sample of the target population of interest
- Survey participants know their drug consumption perfectly and they give honest answers
- Missing data are missing completely at random

$\Rightarrow$ Drug surveys capture less than half of alcohol sales and only 30% to 60% of illicit substance use

$\Rightarrow$ Reported cannabis consumption in Germany has gradually increased over the past 15 years. The social acceptability of cannabis consumption has also increased. We do not know whether people have started to consume more cannabis or whether they have just stopped lying about their consumption.

Accounting for evidence and uncertainty in drug surveys



Addressing the incommensurability in Individual Participant Data (IPD) meta-analysis of randomised controlled trials on treatments for depression

Auxiliary assumptions in IPD meta-analysis

- Different scales for the measurement of depression severity measure the same latent construct

Auxiliary assumptions in IPD meta-analysis

- Different scales for the measurement of depression severity measure the same latent construct
- The same scale measures the same latent construct before and after treatment

Auxiliary assumptions in IPD meta-analysis

- Different scales for the measurement of depression severity measure the same latent construct
- The same scale measures the same latent construct before and after treatment
- The measurements from different scales can be standardised using the cutoff value for mild depression

Auxiliary assumptions in IPD meta-analysis

- Different scales for the measurement of depression severity measure the same latent construct
- The same scale measures the same latent construct before and after treatment
- The measurements from different scales can be standardised using the cutoff value for mild depression
- All follow-up measurements are irrelevant to the estimation of the treatment effect, except for the measurement at six months

Auxiliary assumptions in IPD meta-analysis

- Different scales for the measurement of depression severity measure the same latent construct
- The same scale measures the same latent construct before and after treatment
- The measurements from different scales can be standardised using the cutoff value for mild depression
- All follow-up measurements are irrelevant to the estimation of the treatment effect, except for the measurement at six months
- All measurements of anxiety, quality of life and functioning are irrelevant to the estimation of the treatment effect

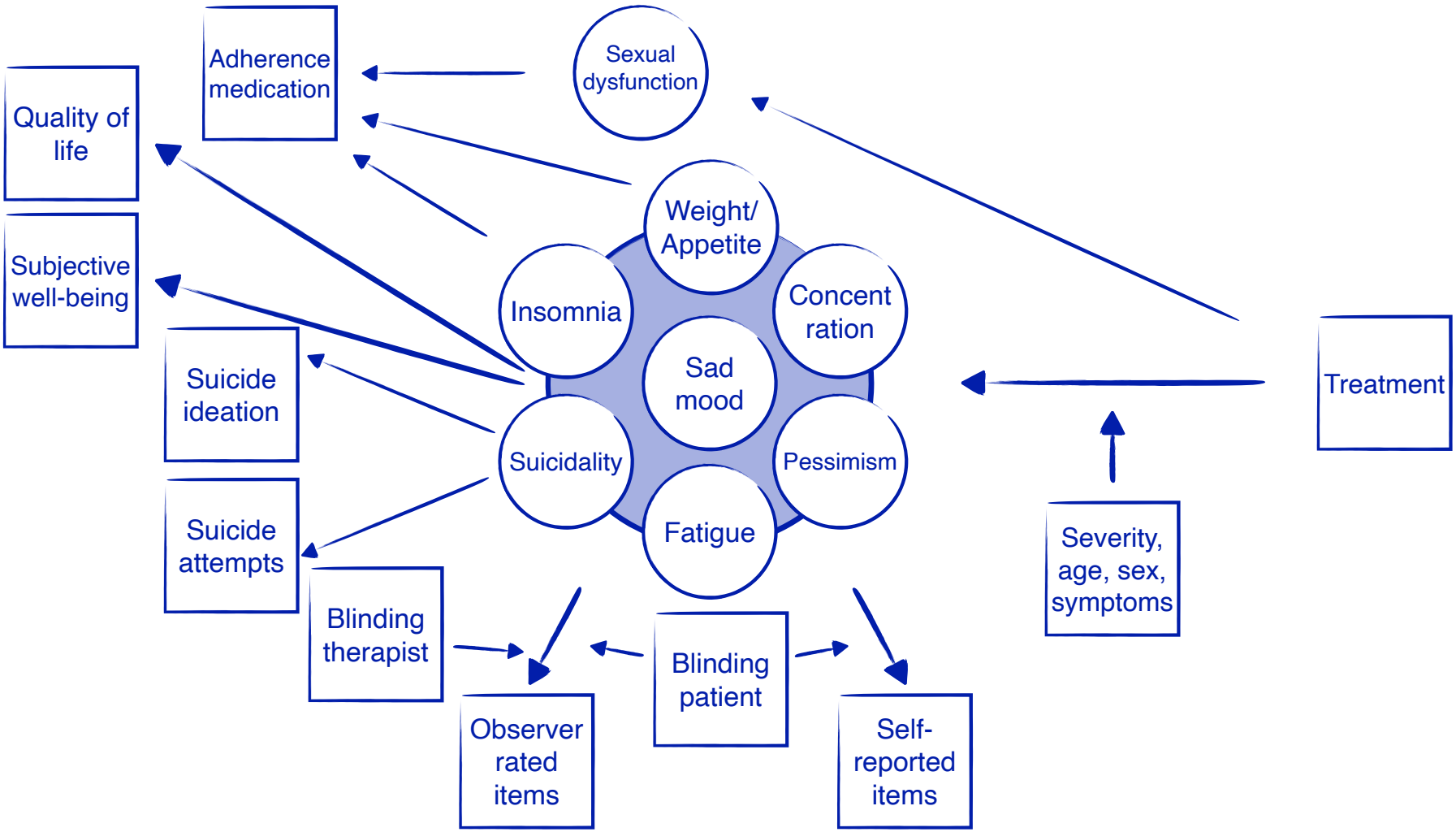
Auxiliary assumptions in IPD meta-analysis

- Different scales for the measurement of depression severity measure the same latent construct
- The same scale measures the same latent construct before and after treatment
- The measurements from different scales can be standardised using the cutoff value for mild depression
- All follow-up measurements are irrelevant to the estimation of the treatment effect, except for the measurement at six months
- All measurements of anxiety, quality of life and functioning are irrelevant to the estimation of the treatment effect
- The treatment effect is the same for all patients, irrespective of their disease severity

Auxiliary assumptions in IPD meta-analysis

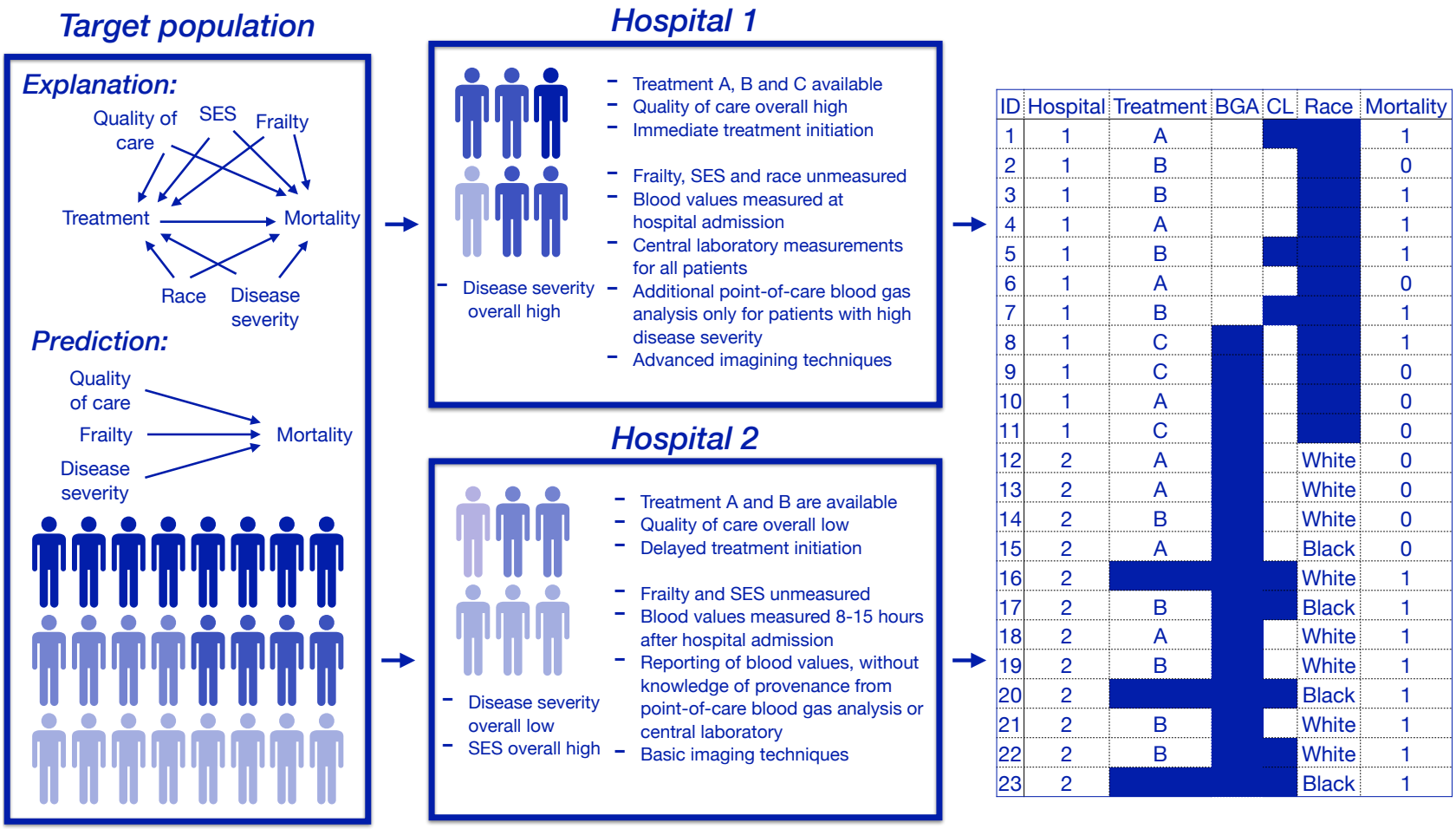
- Different scales for the measurement of depression severity measure the same latent construct
- The same scale measures the same latent construct before and after treatment
- The measurements from different scales can be standardised using the cutoff value for mild depression
- All follow-up measurements are irrelevant to the estimation of the treatment effect, except for the measurement at six months
- All measurements of anxiety, quality of life and functioning are irrelevant to the estimation of the treatment effect
- The treatment effect is the same for all patients, irrespective of their disease severity
- The treatment effect is the same for patients who continue the treatment and patients who discontinue the treatment

Addressing incommensurability in IPD meta-analysis on RCTs for depression



Outlook: Big data paradoxes in routinely collected data

Big data paradoxes in routinely collected data



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

Hoffmann et al. (2026, *provisional acceptance in the BMJ*)

Summary

- In my experience, a good model is not necessarily one that is simple

Summary

- In my experience, a good model is not necessarily one that is simple, simply because data generating mechanisms in the social, behavioural and medical sciences are rarely ever simple.

Summary

- In my experience, a good model is not necessarily one that is simple, simply because data generating mechanisms in the social, behavioural and medical sciences are rarely ever simple.
- Models can still be parsimonious by reflecting everything else we know about the data generating mechanism

Summary

- In my experience, a good model is not necessarily one that is simple
- Models can still be parsimonious by reflecting everything else we know about the data generating mechanism
- Bayesian hierarchical models invite researchers to explore their design space and to verify the plausibility of their auxiliary assumption

Summary

- In my experience, a good model is not necessarily one that is simple
- Models can still be parsimonious by reflecting everything else we know about the data generating mechanism
- Bayesian hierarchical models invite researchers to explore their design space and to verify the plausibility of their auxiliary assumption
- Rather than being uncertain, frequentists often prefer to be wrong, insisting that a donkey is indeed a horse

Summary

- In my experience, a good model is not necessarily one that is simple
- Models can still be parsimonious by reflecting everything else we know about the data generating mechanism
- Bayesian hierarchical models invite researchers to explore their design space and to verify the plausibility of their auxiliary assumption
- Rather than being uncertain, frequentists often prefer to be wrong, insisting that a donkey is indeed a horse

...and I am not actually a Bayesian in the sense of Andrew Gelman:
"Every statistician, from R.A. Fisher on, uses Bayesian inference when it's appropriate. What makes a Bayesian a Bayesian is that he or she uses Bayesian inference when it's inappropriate as well.
(And, yes, I'm a Bayesian.'")

Thank you for your attention!


Good, I. J. (1976).
The bayesian influence, or how to sweep subjectivism under the carpet.
In *Foundations of probability theory, statistical inference, and statistical theories of science*, pages 125–174. Springer.

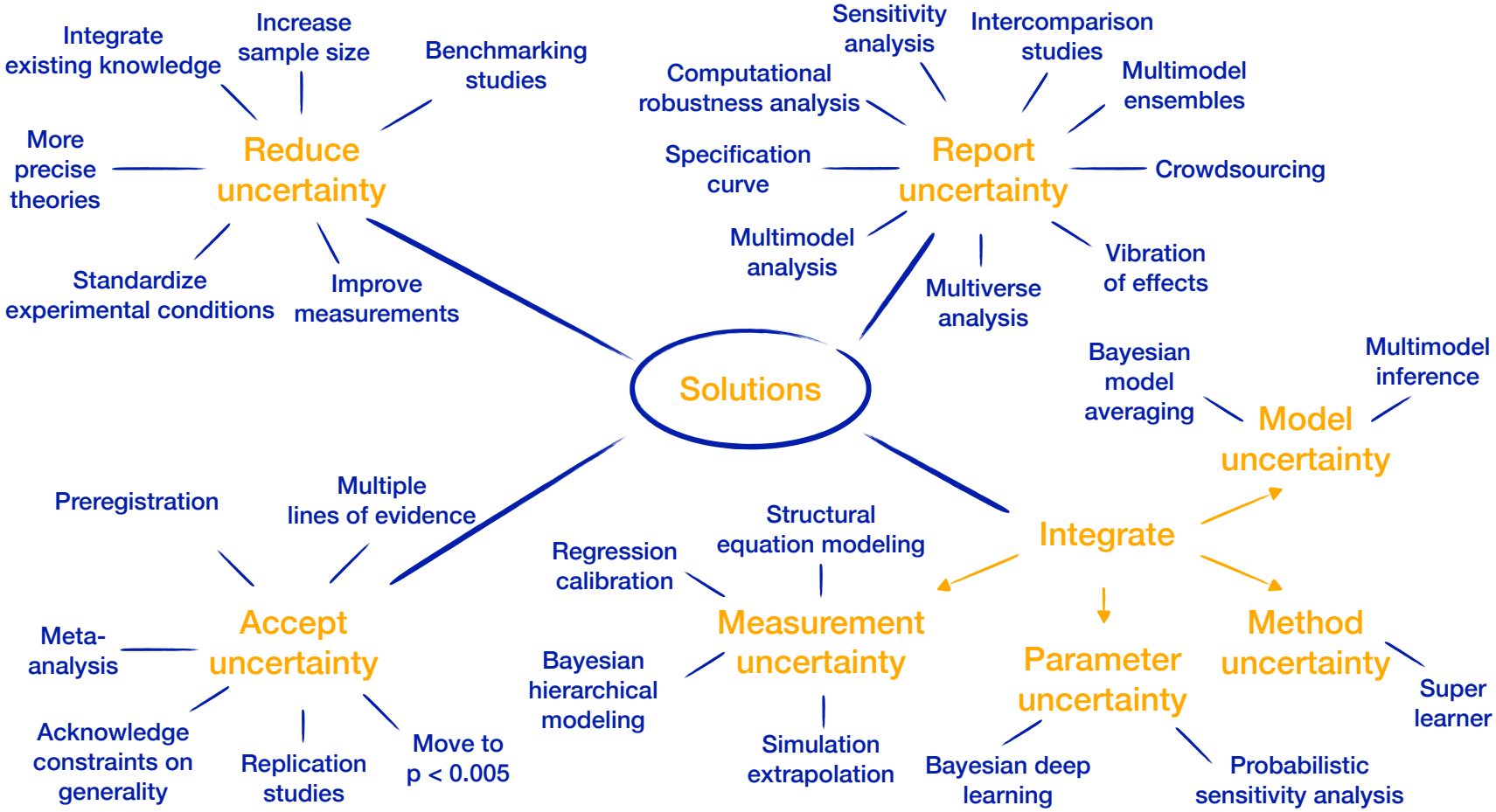

Khazaei, Y., Küchenhoff, H., Hoffmann, S., Syliqi, D., and Rehms, R. (2023).
Using a bayesian hierarchical approach to study the association between non-pharmaceutical interventions and the spread of covid-19 in germany.
Scientific Reports, 13(1):18900.


Meng, X.-L. (2018).
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.


Rehms, R., Ellenbach, N., Rehfuess, E., Burns, J., Mansmann, U., and Hoffmann, S. (2024).

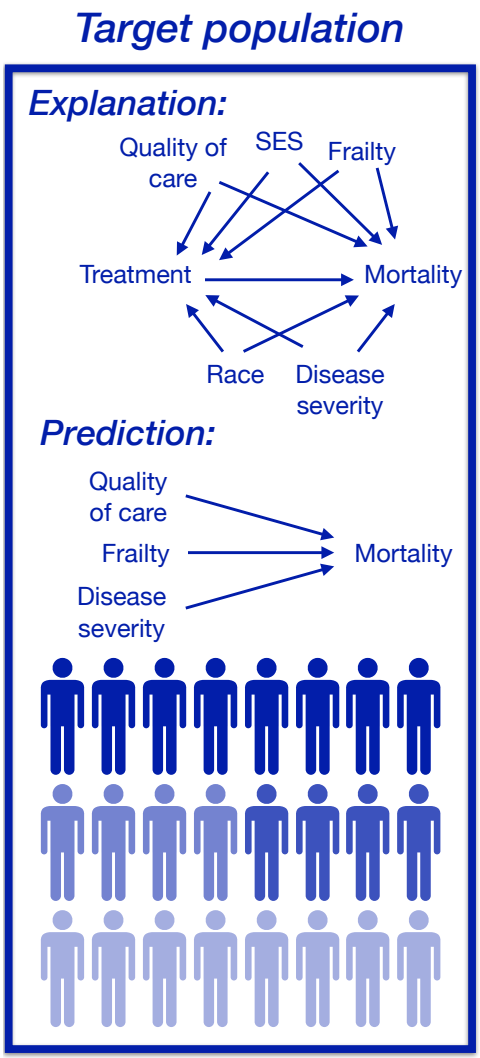
A bayesian hierarchical approach to account for evidence and uncertainty in the modeling of infectious diseases: An application to covid-19.
Biometrical Journal, 66(1):2200341.

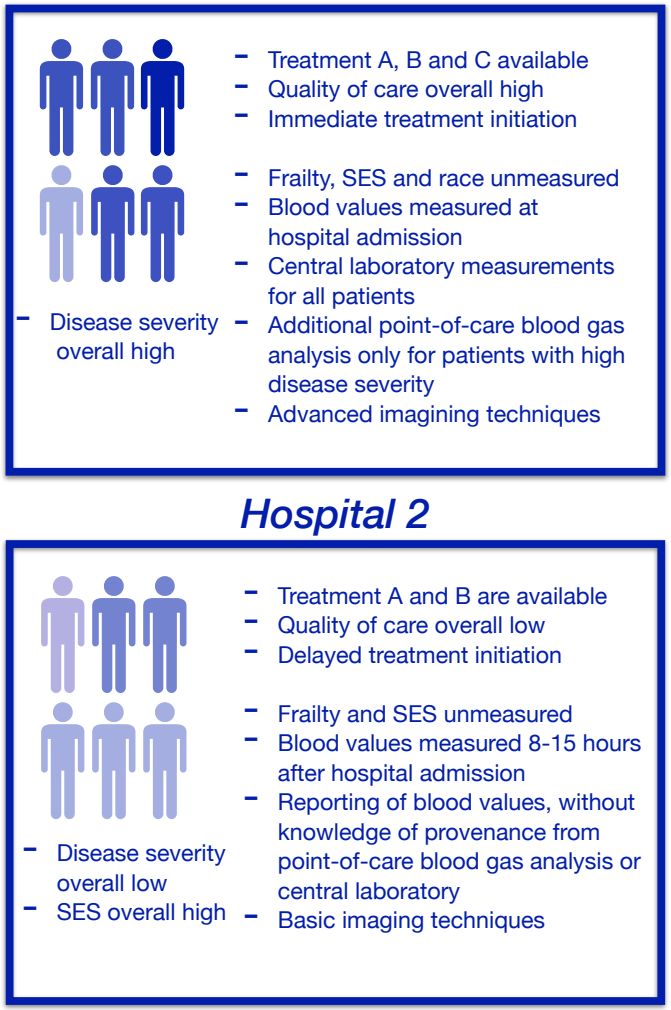
Silberzahn, R. and Uhlmann, E. L. (2015).
Crowdsourced research: Many hands make tight work.
Nature, 526(7572):189–91.



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

Big data paradoxes in routinely collected data





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

Big data paradoxes in routinely collected data

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

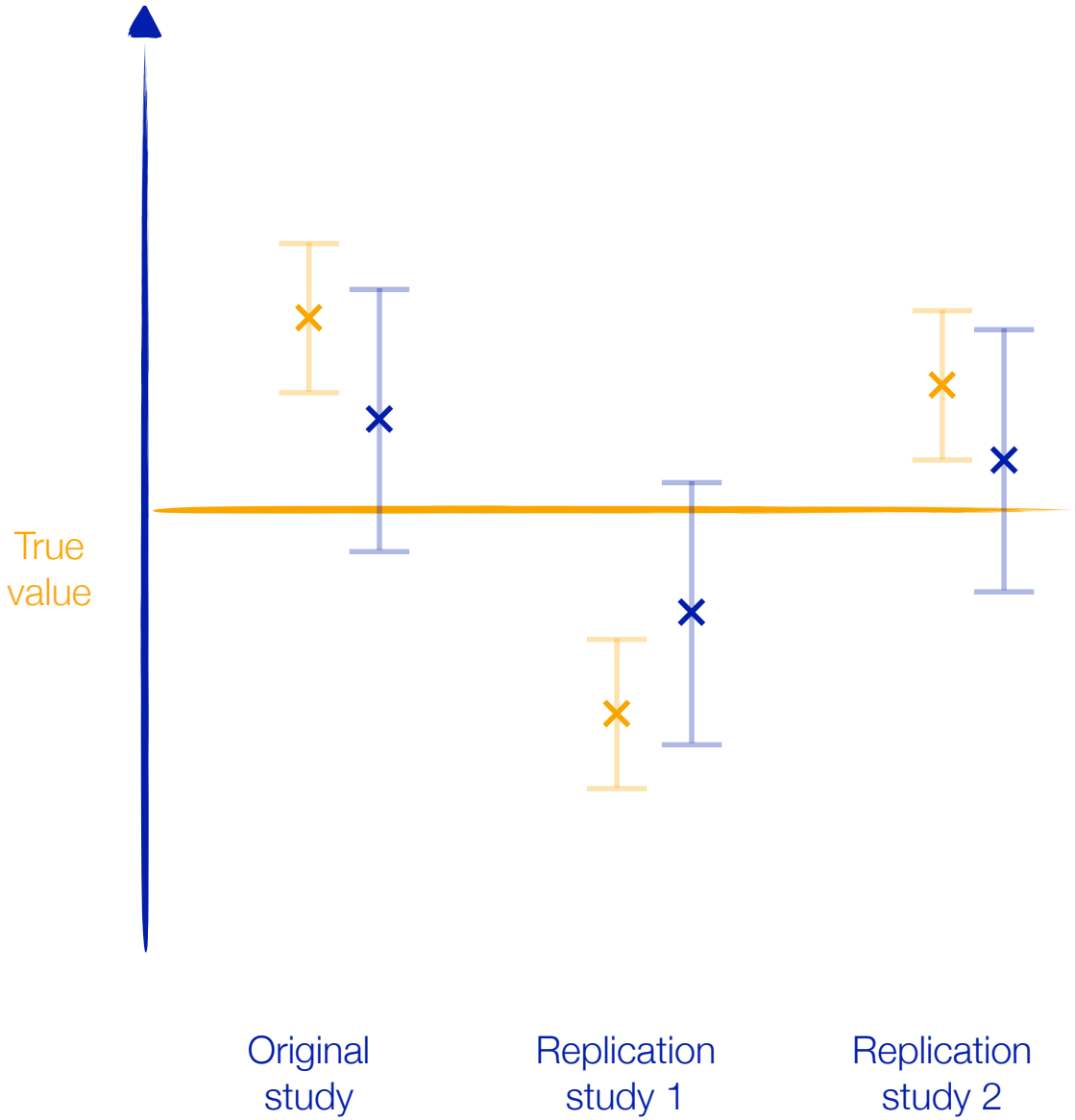
Scott A. Turner, MD*, Hee Soo Jung, MD, FACS, John E. Scarborough, MD, FACS

adison, WI

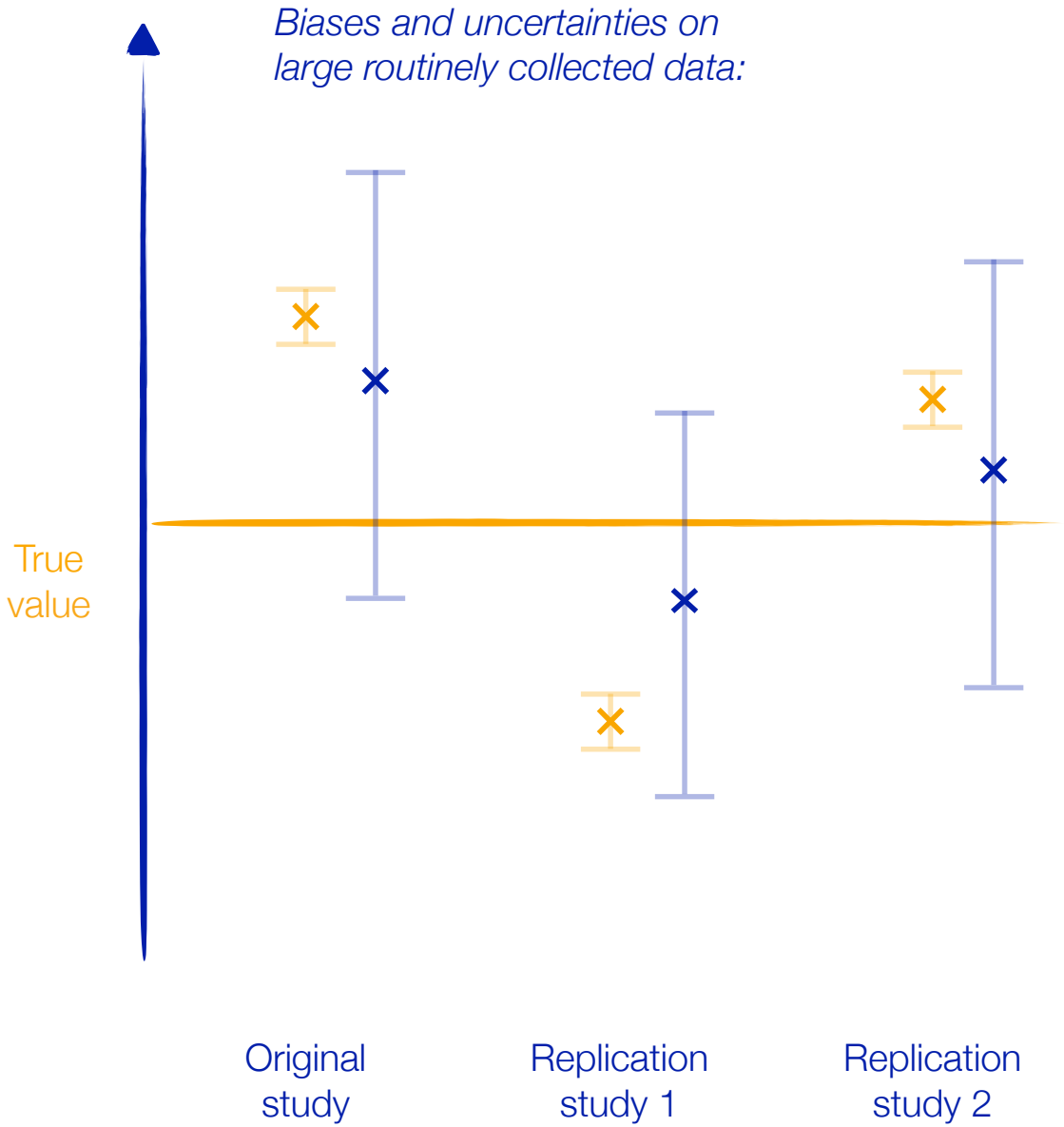
Table. Comparison of 2 Studies on the Association of a Specimen Retrieval Bag With Surgical Site Infection Rates in Laparoscopic Appendectomy

	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 [Meng, 2018]



Big data paradoxes [Meng, 2018]



The multiplicity of analysis strategies

[Silberzahn and Uhlmann, 2015]: Are football referees more likely to give red cards to players with dark skin than to players with light skin?

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

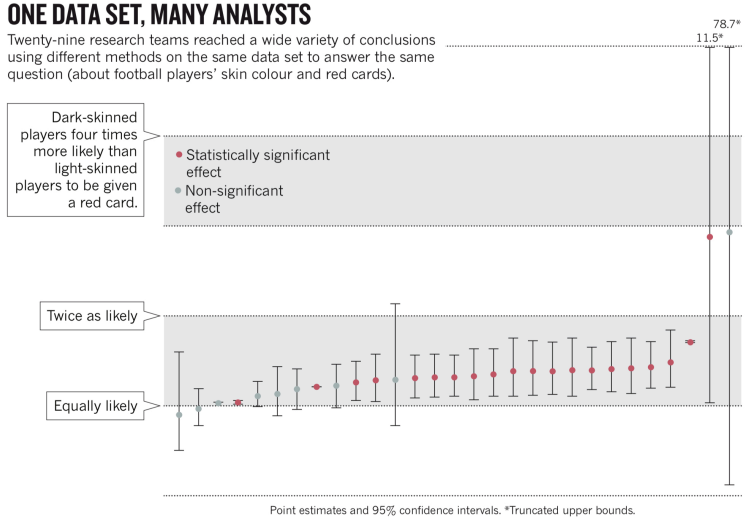


Table 4. Covariates Included by Each Team

	Team																										Percentage
Covariate	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	21	23	24	25	26	27	28	30	31	32	of teams
Player position	X	X	X			X			X	X	X	X	X	X					X	X	X	X	X	X		X	X	X	X	69%
Player's height	X	X		X		X								X						X		X	X	X			X	X		38%
Player's weight	X	X		X		X						X		X								X	X	X			X	X		38%
Player's league country[a]	X										X				X						X			X		X		X		24%
Player's age	X					X					X			X		X				X				X						24%
Goals scored by player		X									X					X							X		X			X		21%
Player's club	X					X								X														X		14%
Referee's country		X	X	X					X																					14%
Referee	X					X								X																10%
Player's number of victories		X								X														X						10%
Number of cards received by player																	X	X												7%
Player																		X												3%
Number of cards awarded by referee																	X													3%
Number of draws										X																				3%
Number of covariates	7	6	2	3	0	6	0	0	2	3	4	2	1	6	1	2	2	2	1	3	2	3	4	6	1	2	3	6	1	

Note: The covariates are listed in order of the frequency with which they were included in teams' analytic approaches. The number of games a player had played was essential to the analysis, was used by all teams, and is thus not listed here as a separate covariate.
[a]Team 9 mistakenly labeled referee's country as league country.

The multiplicity of possible analysis strategies

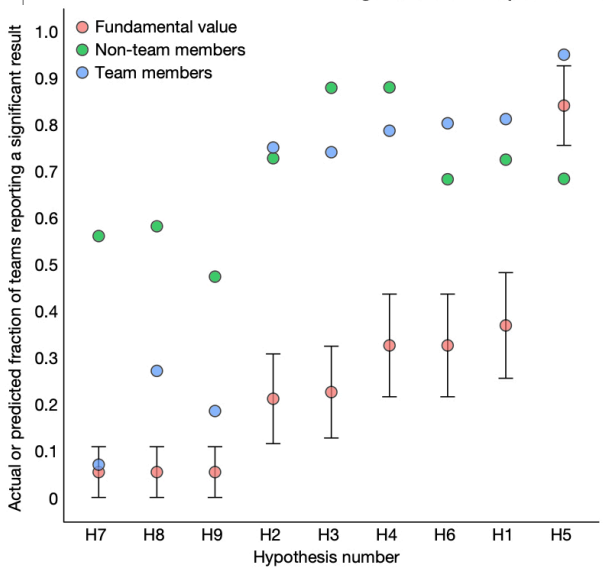
Article

Variability in the analysis of a single neuroimaging dataset by many teams

https://doi.org/10.1038/s41586-020-2314-9
Received: 14 November 2019
Accepted: 7 April 2020
Published online: 20 May 2020
Check for updates

A list of authors and affiliations appears in the online version of the pap

Data analysis workflows in many scientific domains have becom
complex and flexible. Here we assess the effect of this flexibility
functional magnetic resonance imaging by asking 70 independe
the same dataset, testing the same 9 ex-ante hypotheses[1]. The fle
approaches is exemplified by the fact that no two teams chose id
analyse the data. This flexibility resulted in sizeable variation in t
hypothesis tests, even for teams whose statistical maps were hig
intermediate stages of the analysis pipeline. Variation in reporte
to several aspects of analysis methodology. Notably, a meta-ana
aggregated information across teams yielded a significant conse
regions. Furthermore, prediction markets of researchers in the f
gnificant findings, even by
ır findings show that analyt
conclusions, and identify fa
functional magnetic resor
validating and sharing com
l for performing and repor
es that could be used to mit



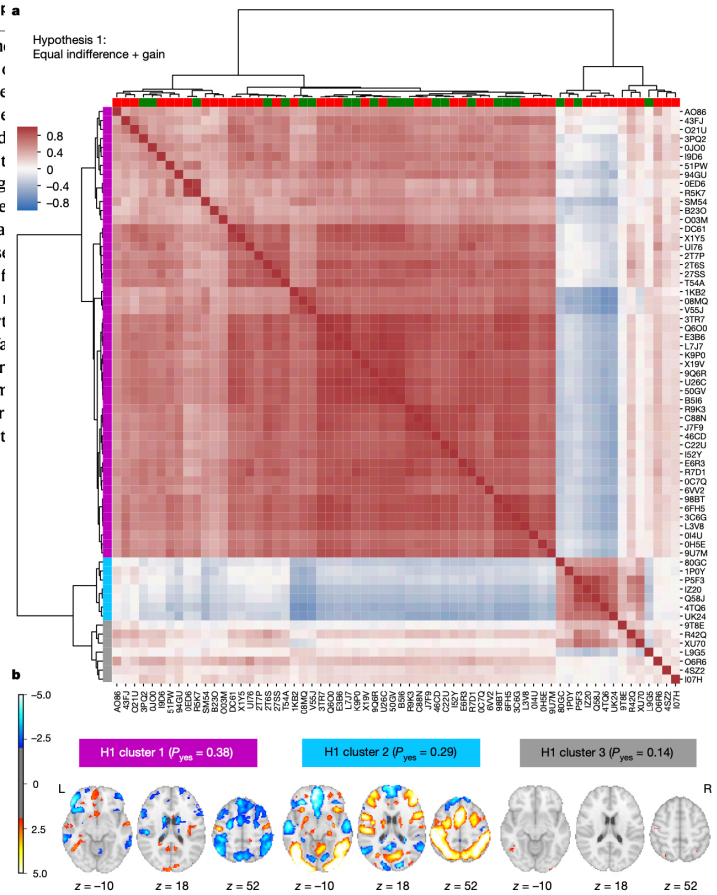


Fig. 1 | Fraction of teams reporting a significant result and prediction market beliefs. The observed fraction of teams reporting significant results

The multiplicity of possible analysis strategies

Received: 5 June 2020 | Accepted: 14 February 2021

DOI: 10.1111/ecin.12992

Economic Inquiry

ORIGINAL ARTICLE

The influence of hidden researcher decisions in applied microeconomics

Nick Huntington-Klein, Assistant Professor[1] | Andreu Arenas, Assistant Professor[2] | Emily Beam, Assistant Professor[3] | Marco Bertoni, Associate Professor[4] | Jeffrey R. Bloem, Research Economist[5] | Pralhad Burli, Economist[6] | Naibin Chen, Graduate Student[7] | Paul Grieco, Associate Professor[8] | Godwin Ekpe, Graduate Student[9] | Todd Pugatch, Associate Professor[10] | Martin Saavedra, Associate Professor[11] | Yaniv Stopnitzky, Assistant Professor[12]

[1]Seattle University, Seattle, Washington, USA

[2]University of Barcelona & IEB, Barcelona, Spain

[3]University of Vermont, Burlington, Vermont, USA

[4]Department of Economics and Management "M. Fanno", Padova University, Padova, Italy

[5]USDA Economic Research Service, Kansas City, Missouri, USA

[6]Idaho National Laboratory, Idaho Falls, Idaho, USA

[7]Pennsylvania State University, 303 Kern Building, University Park, Pennsylvania, USA

[8]Pennsylvania State University, 508 Kern Graduate Building, University Park, Pennsylvania, USA

[9]Northern Illinois University, Dekalb,

Abstract

Researchers make hundreds of decisions about data collection, preparation, and analysis in their research. We use a many-analysts approach to measure the extent and impact of these decisions. Two published causal empirical results are replicated by seven replicators each. We find large differences in data preparation and analysis decisions, many of which would not likely be reported in a publication. No two replicators reported the same sample size. Statistical significance varied across replications, and for one of the studies the effect's sign varied as well. The standard deviation of estimates across replications was 3–4 times the mean reported standard error.

KEYWORDS

metascience, replication, research

JEL CLASSIFICATION

C81; C10; B41