IDM Annual Symposium 2023
Session 3E

Estimating the population-level impact of vaccines using counterfactual prediction with LASSO regression

Anabelle Wong [1,2], Sarah Kramer [1], Marco Piccininni [2,3], Jessica L. Rohmann [2,3], Tobias Kurth [2], Sylvie Escolano [4], Ulrike Grittner [5,6] and Matthieu Domenech de Cellès [1]

[1] Max Planck Institute for Infection Biology, Charitéplatz 1, 10117 Berlin, Germany.
[2] Institute of Public Health, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.
[3] Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.
[4] Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Disease, Inserm U1181 (B2PHI), UVSQ, University Paris Saclay, Institut Pasteur, Paris, France.
[5] Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.
[6] Berlin Institute of Health, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.

The pneumococcal conjugate vaccines (PCVs)

- *Streptococcus pneumoniae* causes pneumonia and invasive diseases

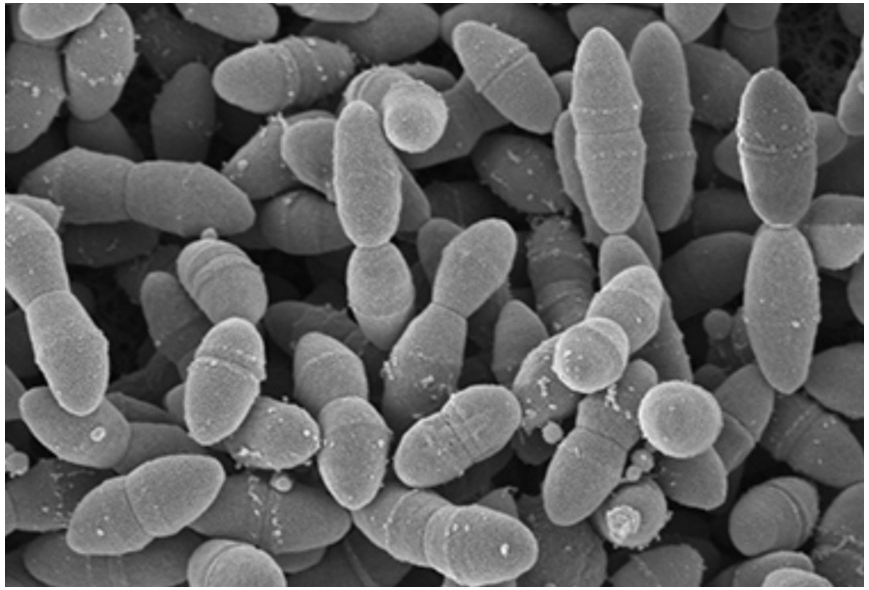
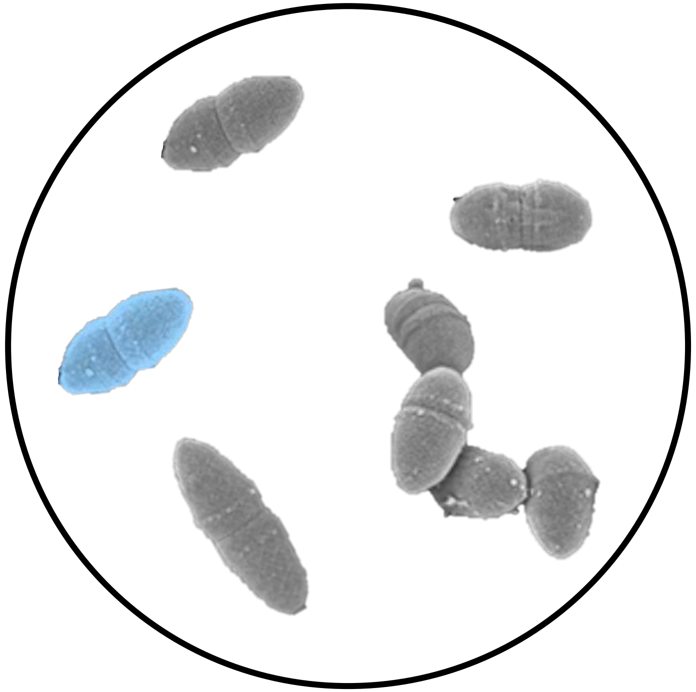


Image: Robert Koch Institut

The pneumococcal conjugate vaccines (PCVs)

- *Streptococcus pneumoniae* causes pneumonia and invasive diseases
- PCVs cover up to **20** out of 100 serotypes[1]
- Serotype replacement may erode vaccine impact

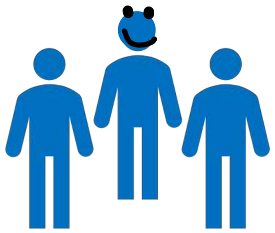


1. Ganaie et al. (2020) *mBio*

The challenges in estimating PCV impact

- Population impact: direct effect + indirect effect[1]

Mixed vaccinated
and unvaccinated



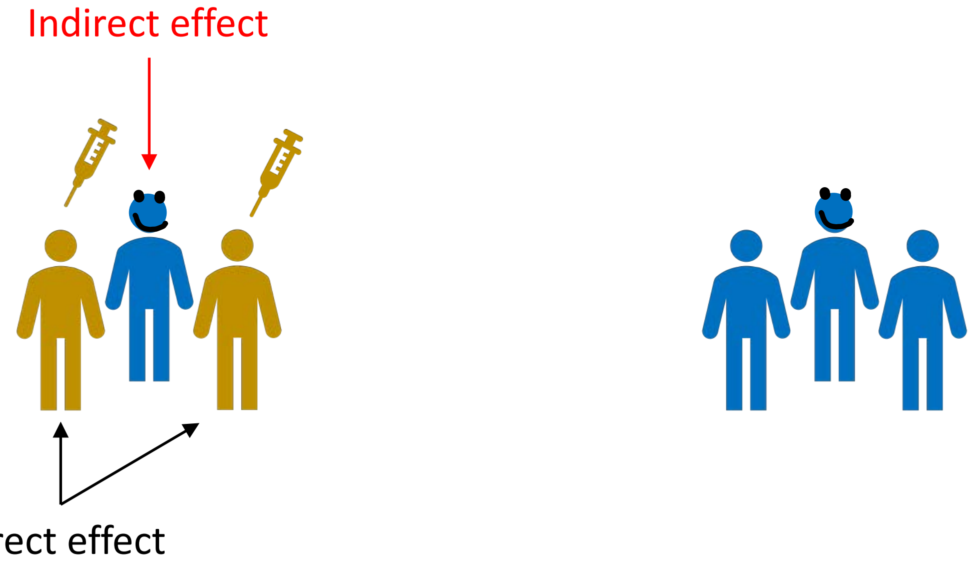
Fully unvaccinated

1. Halloran et al. *Design and Analysis of Vaccine Studies*. New York: Springer US; 2010.

The challenges in estimating PCV impact

- Population impact: direct effect + indirect effect[1]
- Indirect effect cannot be easily estimated in RCTs

Mixed vaccinated and unvaccinated



Fully unvaccinated

1. Halloran et al. *Design and Analysis of Vaccine Studies*. New York: Springer US; 2010.

The challenges in estimating PCV impact

- Population impact: direct effect + indirect effect[1]
- Indirect effect cannot be easily estimated in RCTs

- Vaccine impact can be estimated from observational studies

1. Halloran et al. *Design and Analysis of Vaccine Studies*. New York: Springer US; 2010.

The challenges in estimating PCV impact

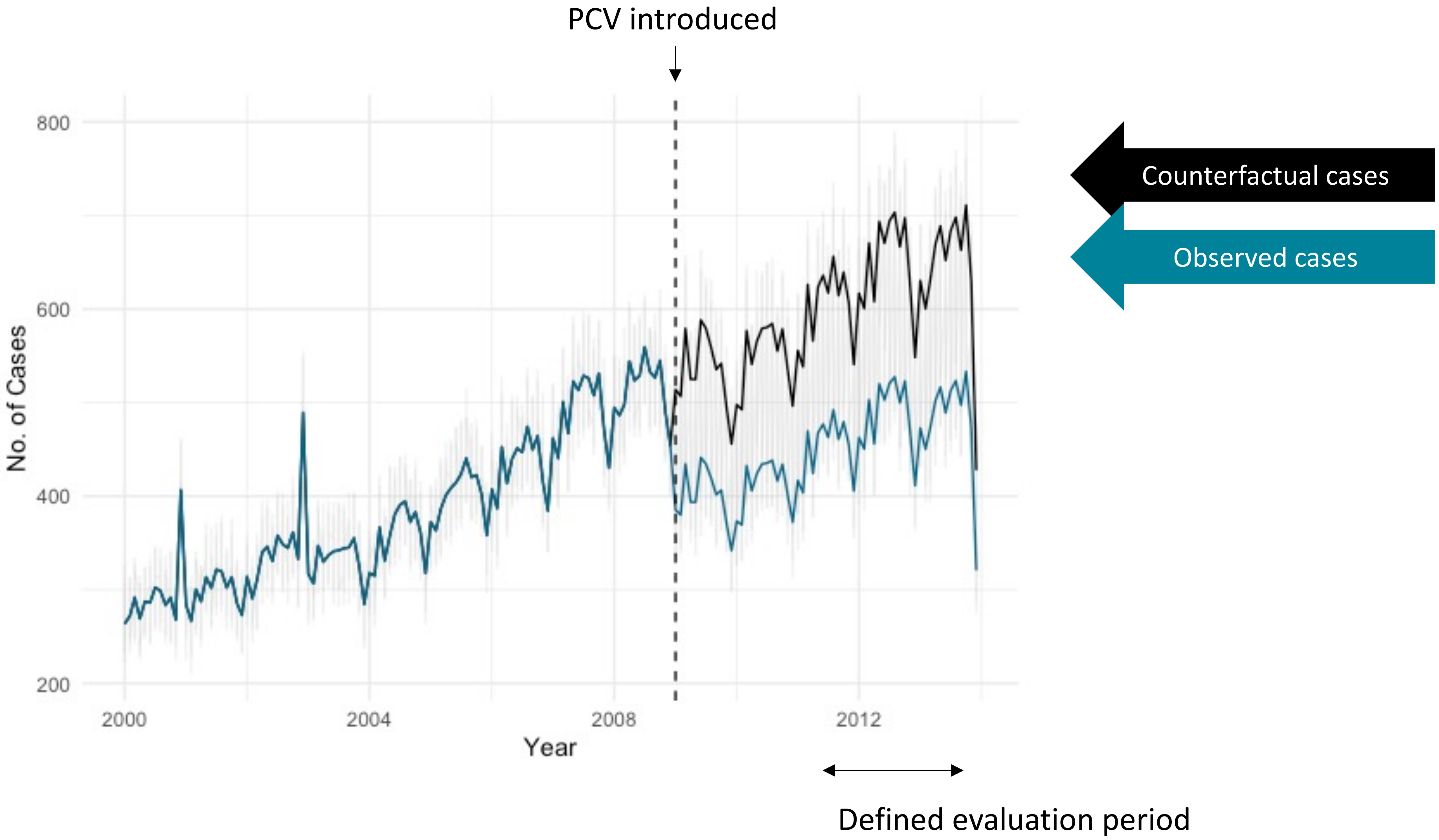
- Population impact: direct effect + indirect effect[1]
- Indirect effect cannot be easily estimated in RCTs

- Vaccine impact can be estimated from observational studies
- But confounding bias may occur

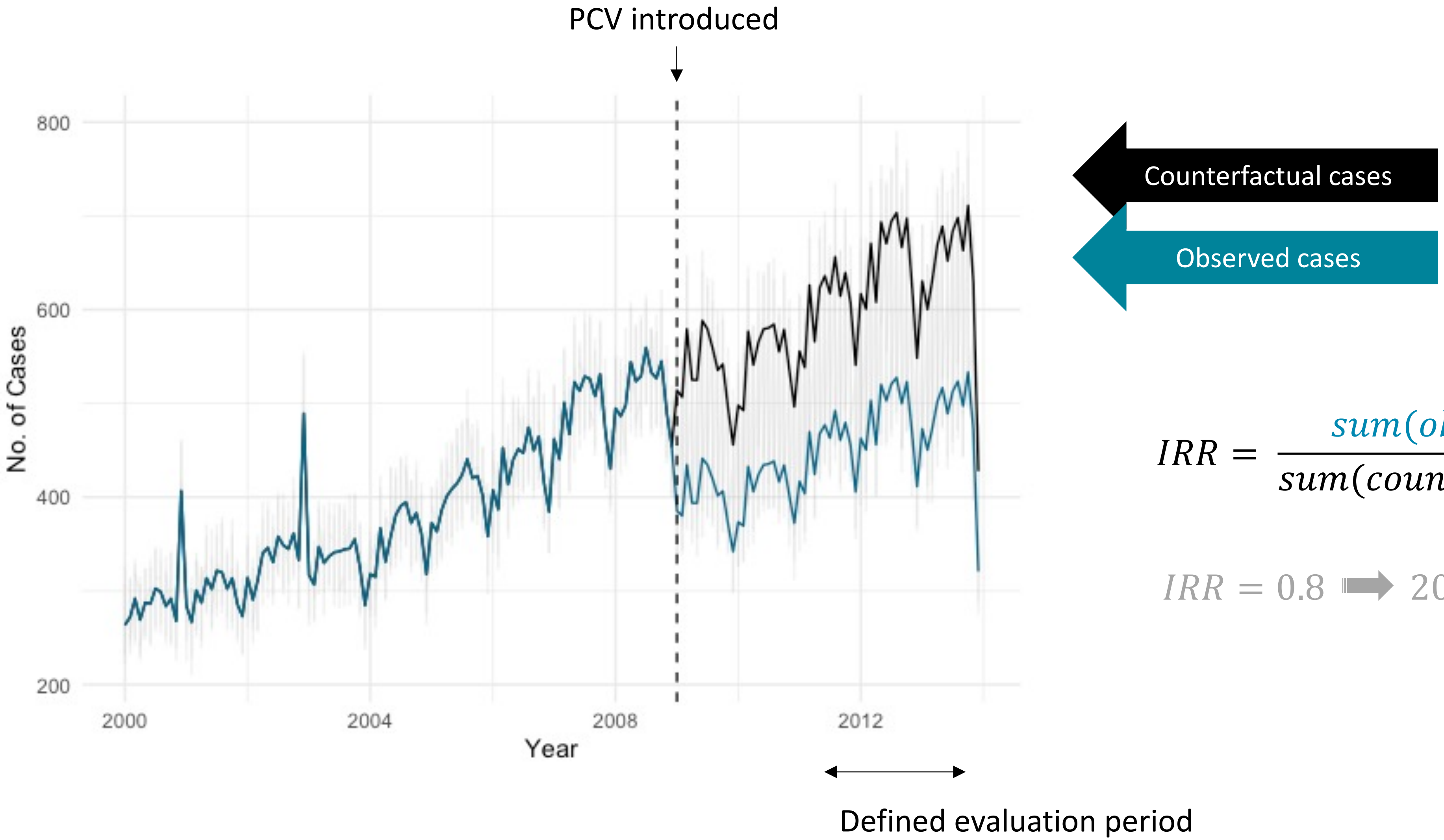
  - overestimation: improved living condition and infection prevention
  - underestimation: increased surveillance and diagnosis

1. Halloran et al. *Design and Analysis of Vaccine Studies*. New York: Springer US; 2010.

How to estimate vaccine impact?



How to estimate vaccine impact?



$$IRR = \frac{sum(observed)}{sum(counterfactual)}$$

$IRR = 0.8$ ➡ 20% *reduction*

How to predict counterfactual outcome?

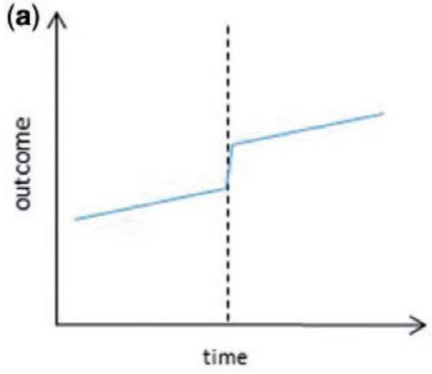
<table>
<tr><th colspan="2">Rely on outcome of interest</th><th colspan="2">Synthetic control</th></tr>
<tr><th>Interrupted Time Series (ITS)</th><th>ITS + offset</th><th>Hand-picked:
Select unaffected controls[1]</th><th>Data-driven:
Bayesian variable selection[2]</th></tr>
<tr>
<td>

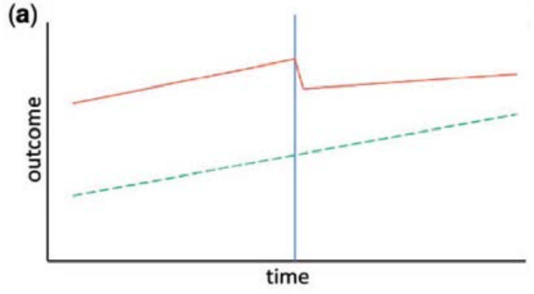
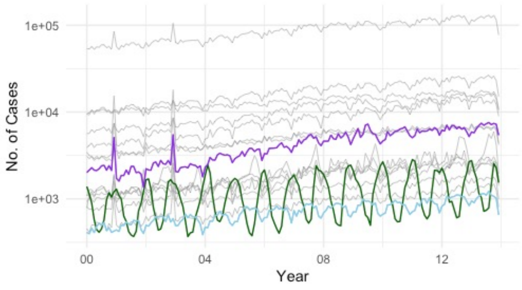
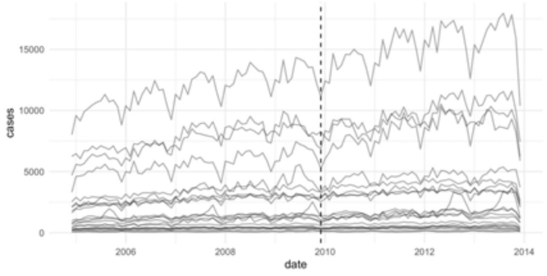
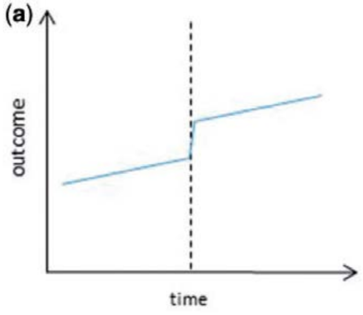
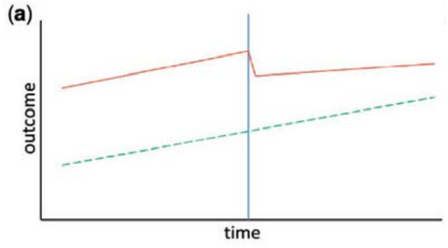
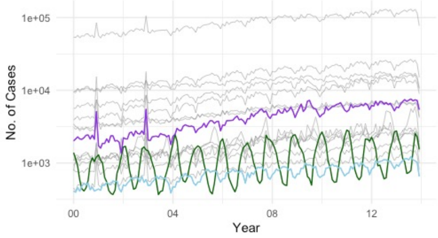
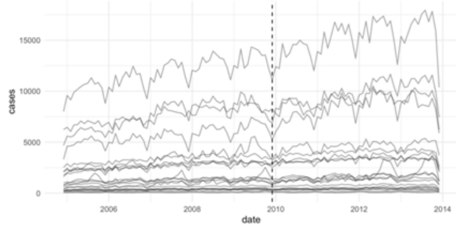
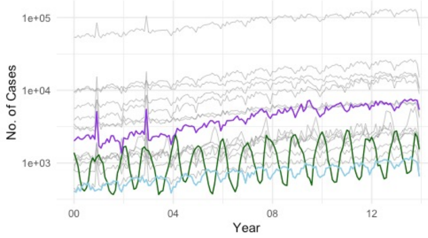
Image: Bernal et al. (2017) Int I Epidemiol</td>
<td>

Image: Bernal et al. (2018) Int I Epidemiol</td>
<td>
</td>
<td>
</td>
</tr>
</table>

1. Thorrington et al. (2018) *BMC Medicine*
2. Bruhn et al. (2017) *PNAS*

How to predict counterfactual outcome?

Rely on outcome of interest		Synthetic control		
			Data driven	
Interrupted Time Series (ITS)	ITS + offset	Hand-picked controls	Bayesian variable selection	LASSO regression[1]
 Image: Bernal et al. (2017) *Int I Epidemiol*	 Image: Bernal et al. (2018) *Int I Epidemiol*			

1. Tibshirani et al. (1996) *J R Statist Soc B*

How to predict counterfactual outcome?

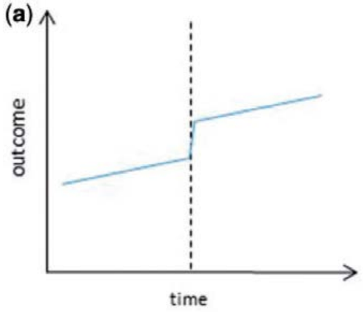
<table>
<tr><td colspan="2">Rely on outcome of interest</td><td colspan="3">Synthetic control</td></tr>
<tr><td rowspan="2">Interrupted Time Series (ITS)</td><td rowspan="2">ITS + offset
ITS</td><td rowspan="2">Hand-picked controls</td><td colspan="2">Data driven</td></tr>
<tr><td>Bayesian variable selection
SC</td><td>LASSO regression[1]
LASSO</td></tr>
<tr>
<td>

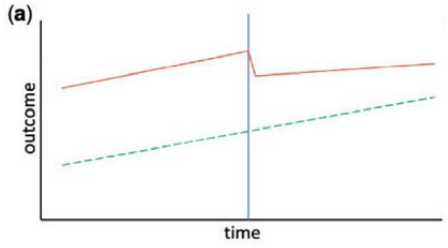
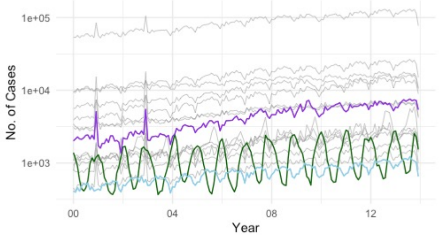
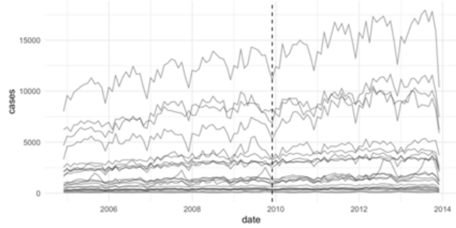
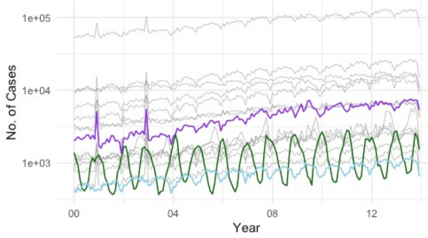
Image:
Bernal et al. (2017) *Int J Epidemiol*</td>
<td>

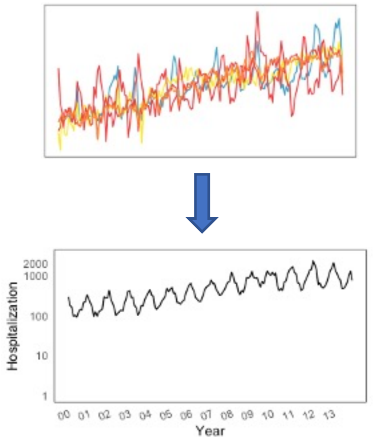
Image:
Bernal et al. (2018) *Int J Epidemiol*</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>

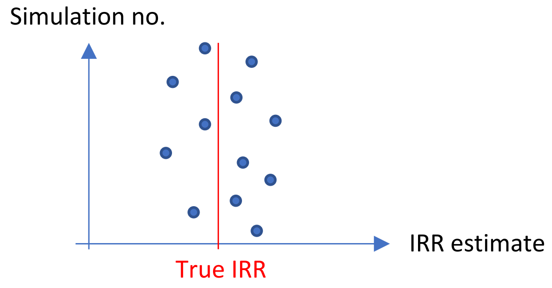
1. Tibshirani et al. (1996) *J R Statist Soc B*

Study design

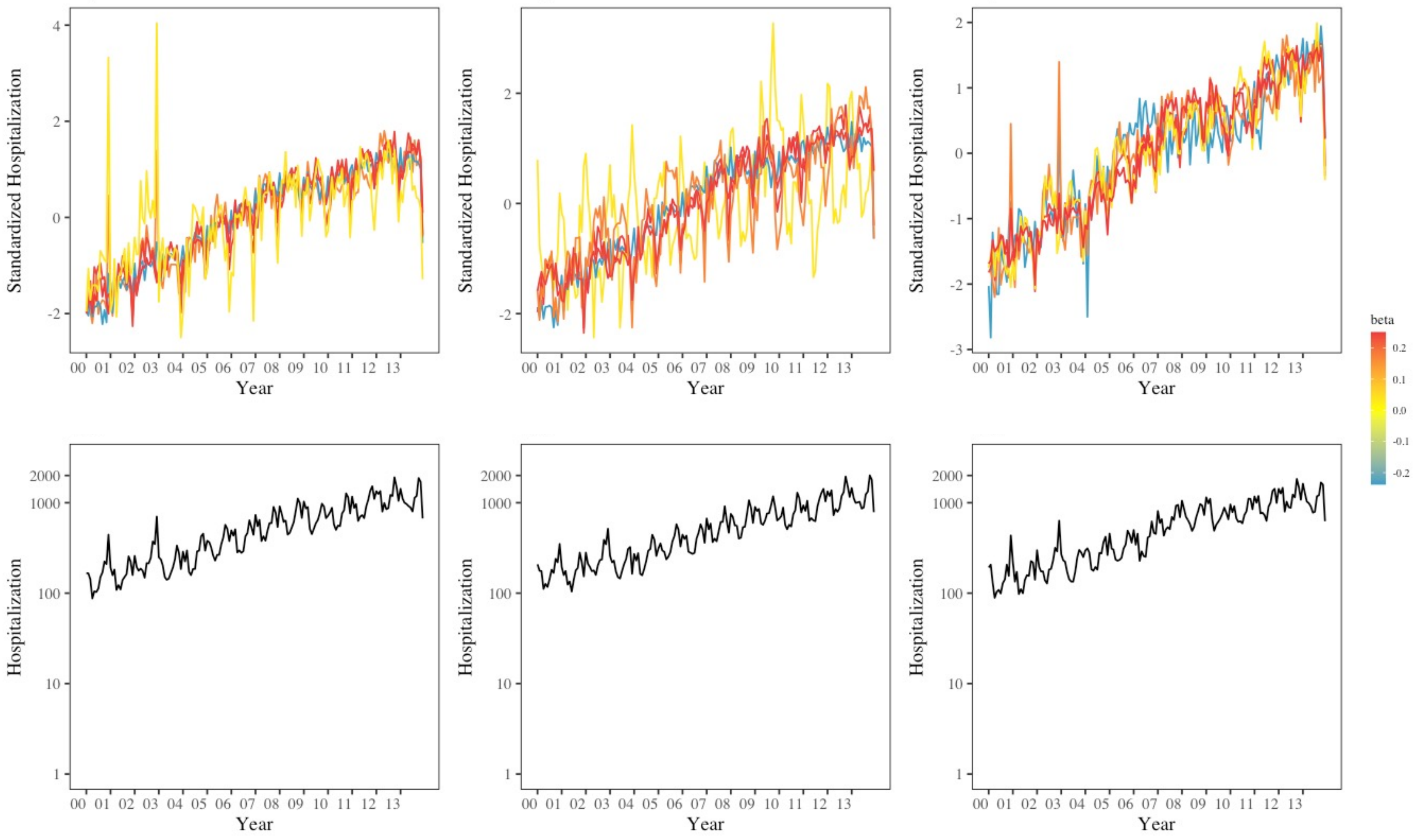
Simulate data

Test methods on simulated data

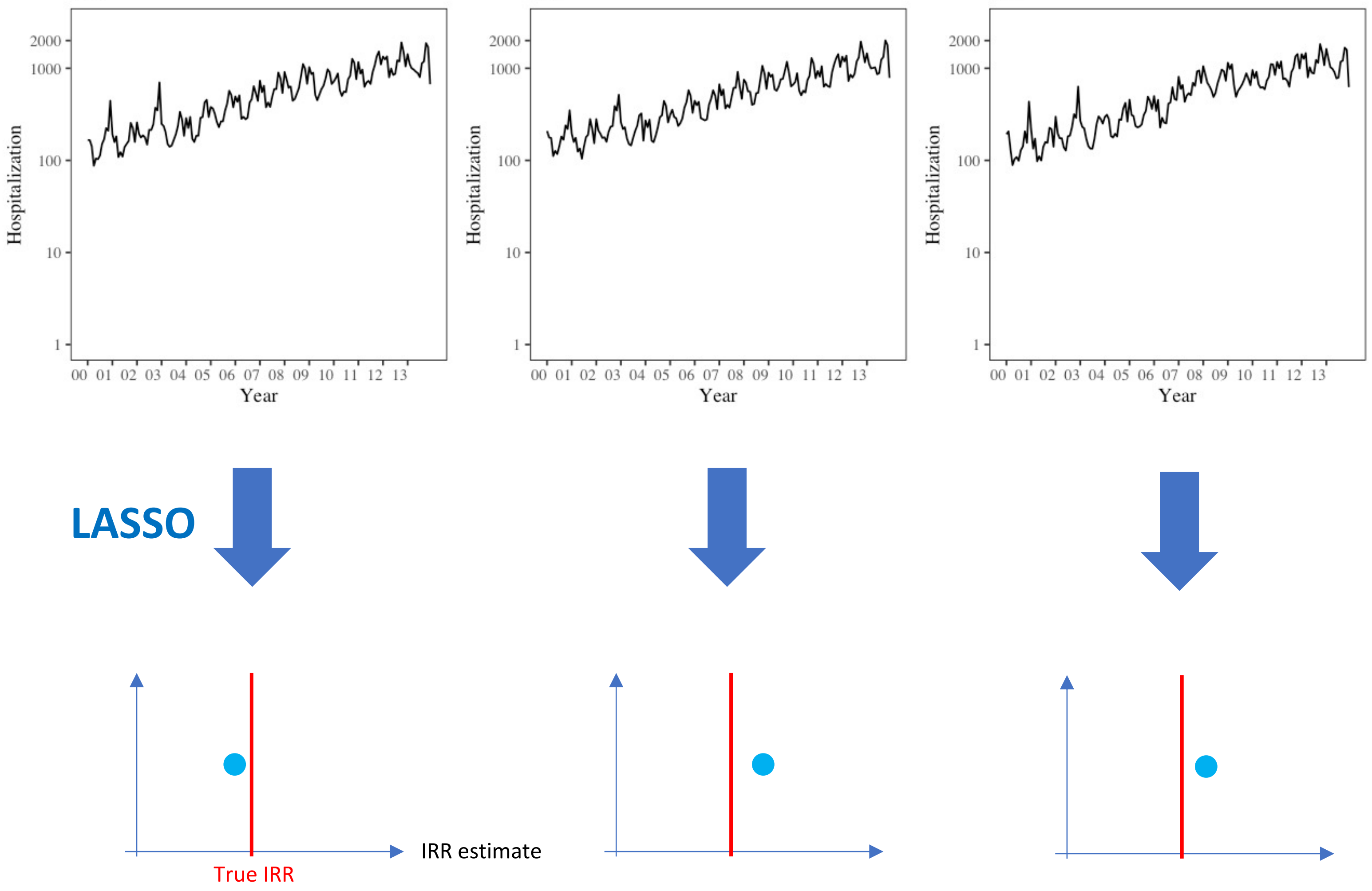




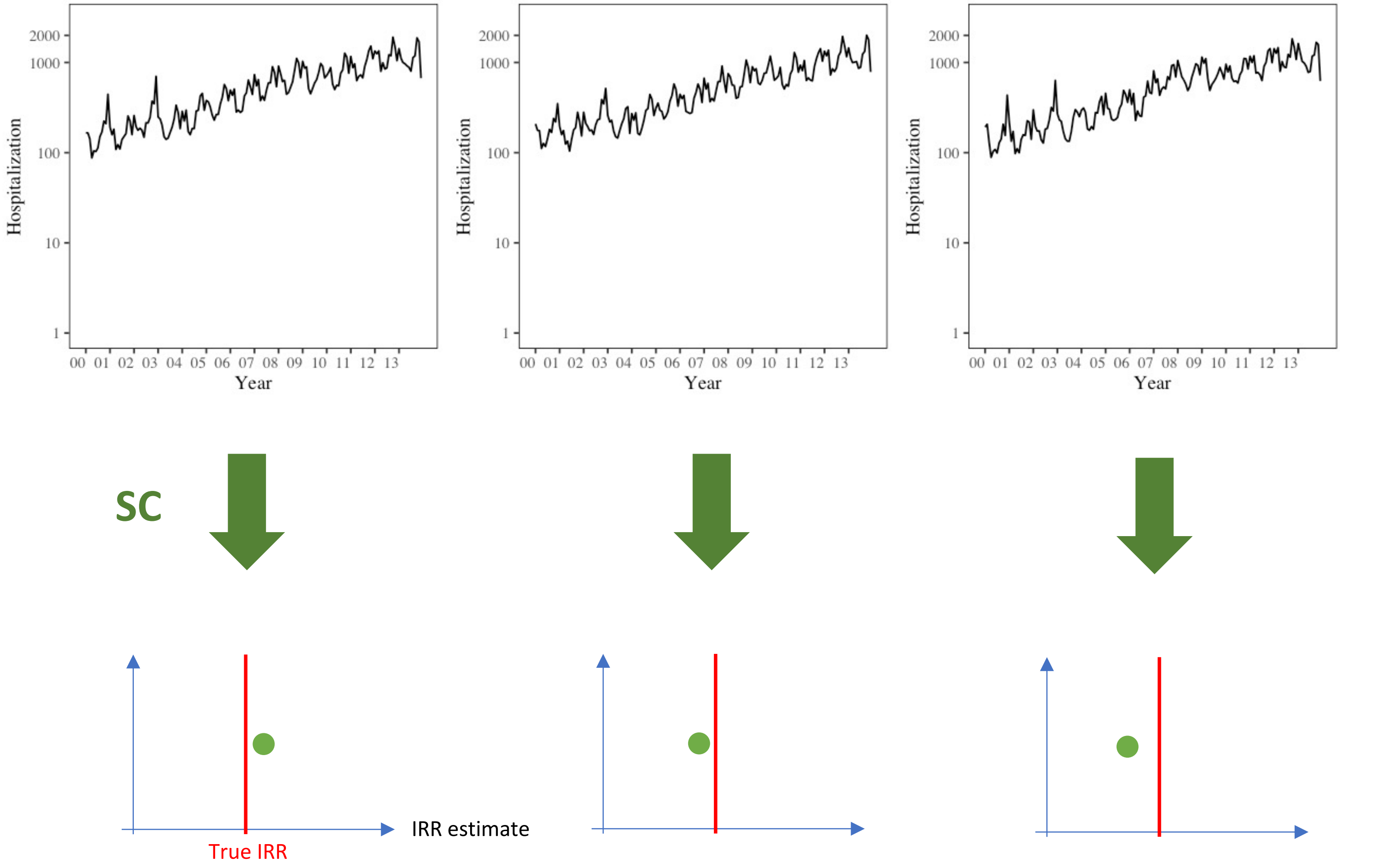
We simulated outcome based on real data



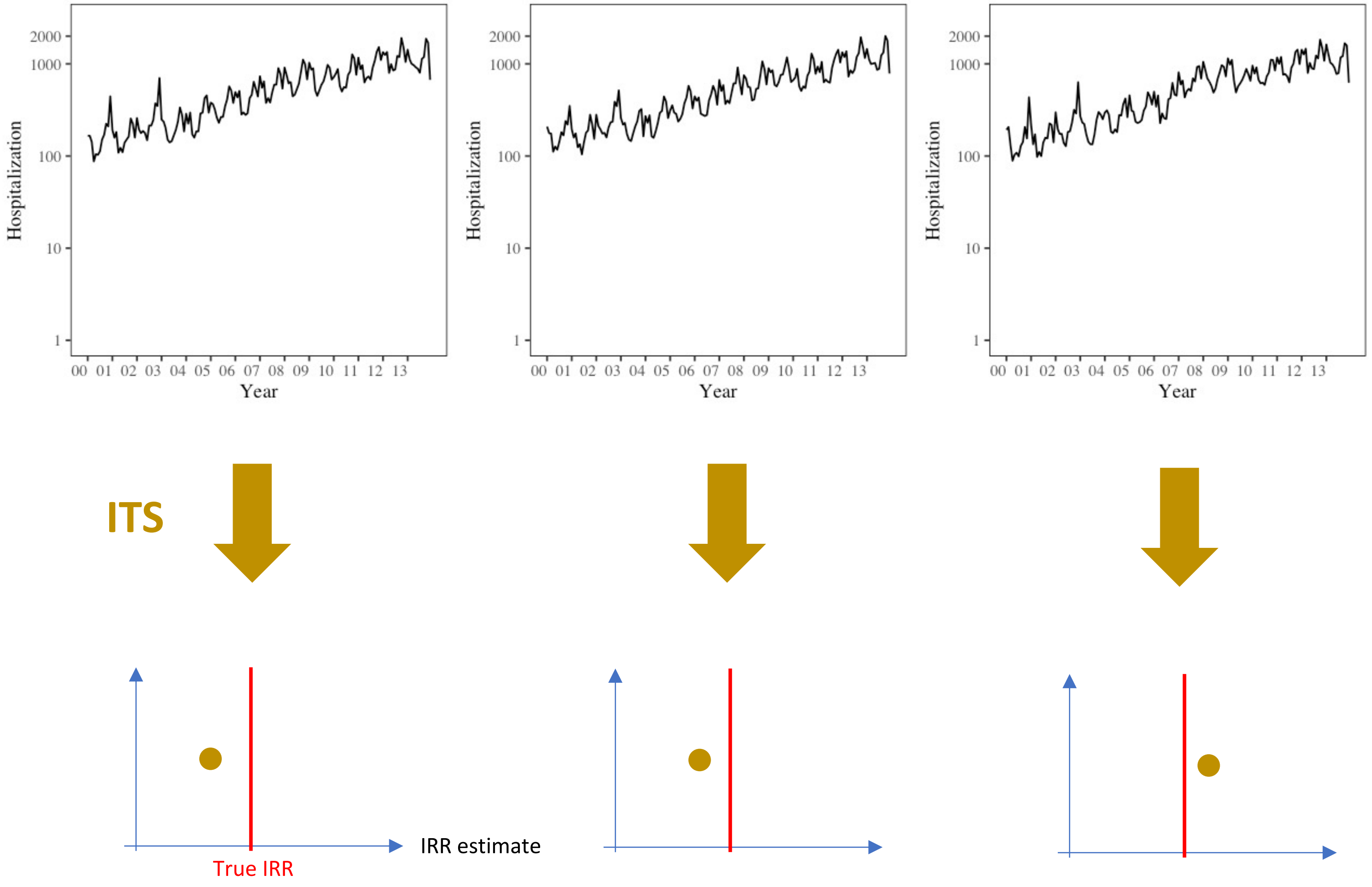
We estimated IRR in each simulated data set



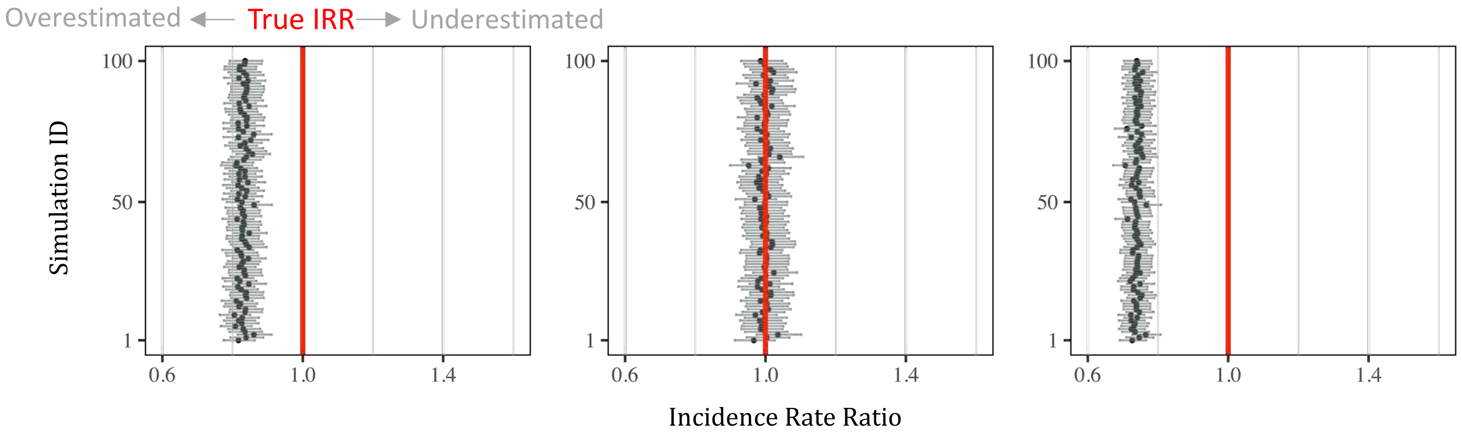
We estimated IRR in each simulated data set



We estimated IRR in each simulated data set

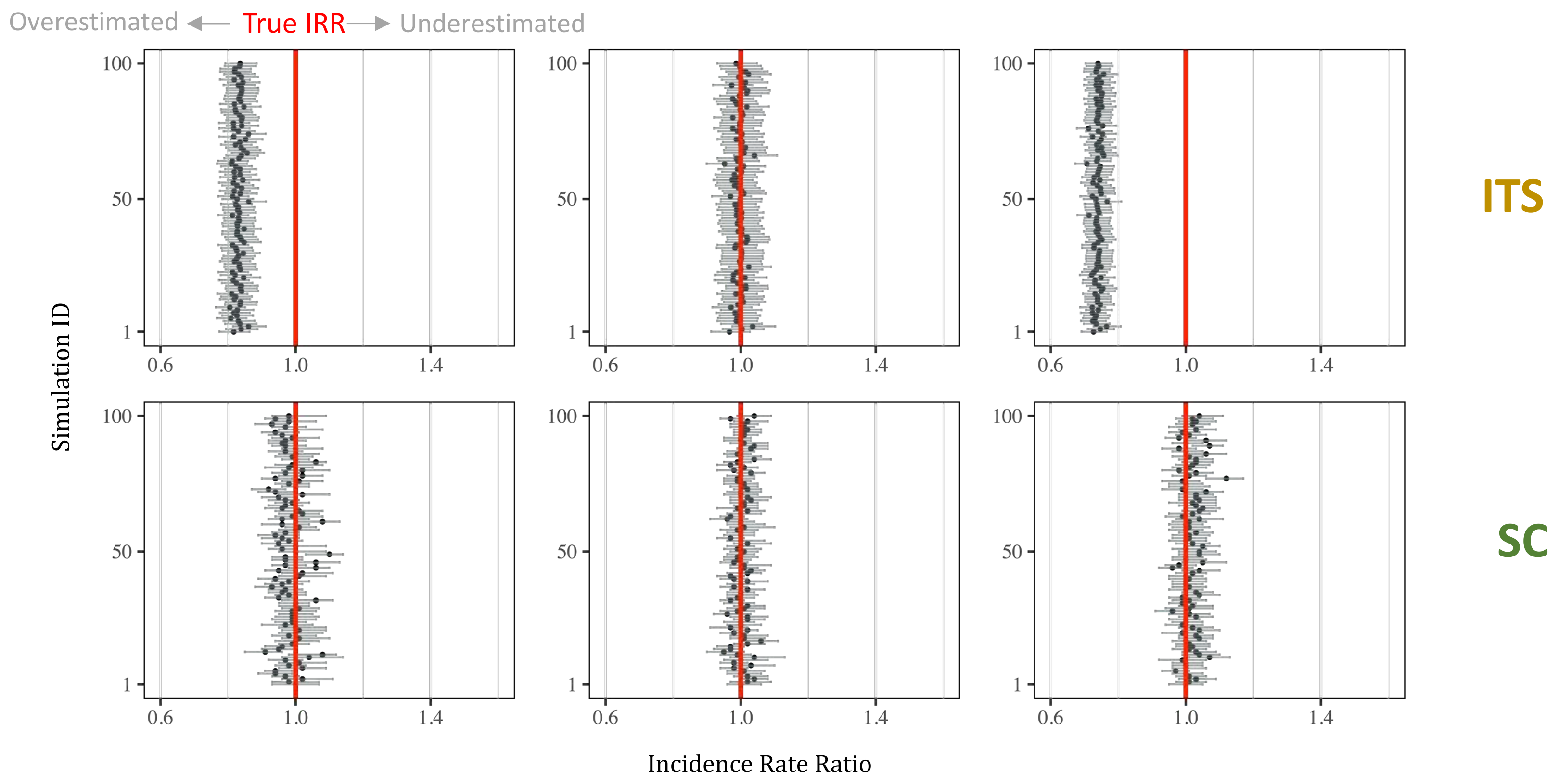


ITS estimates were sometimes biased

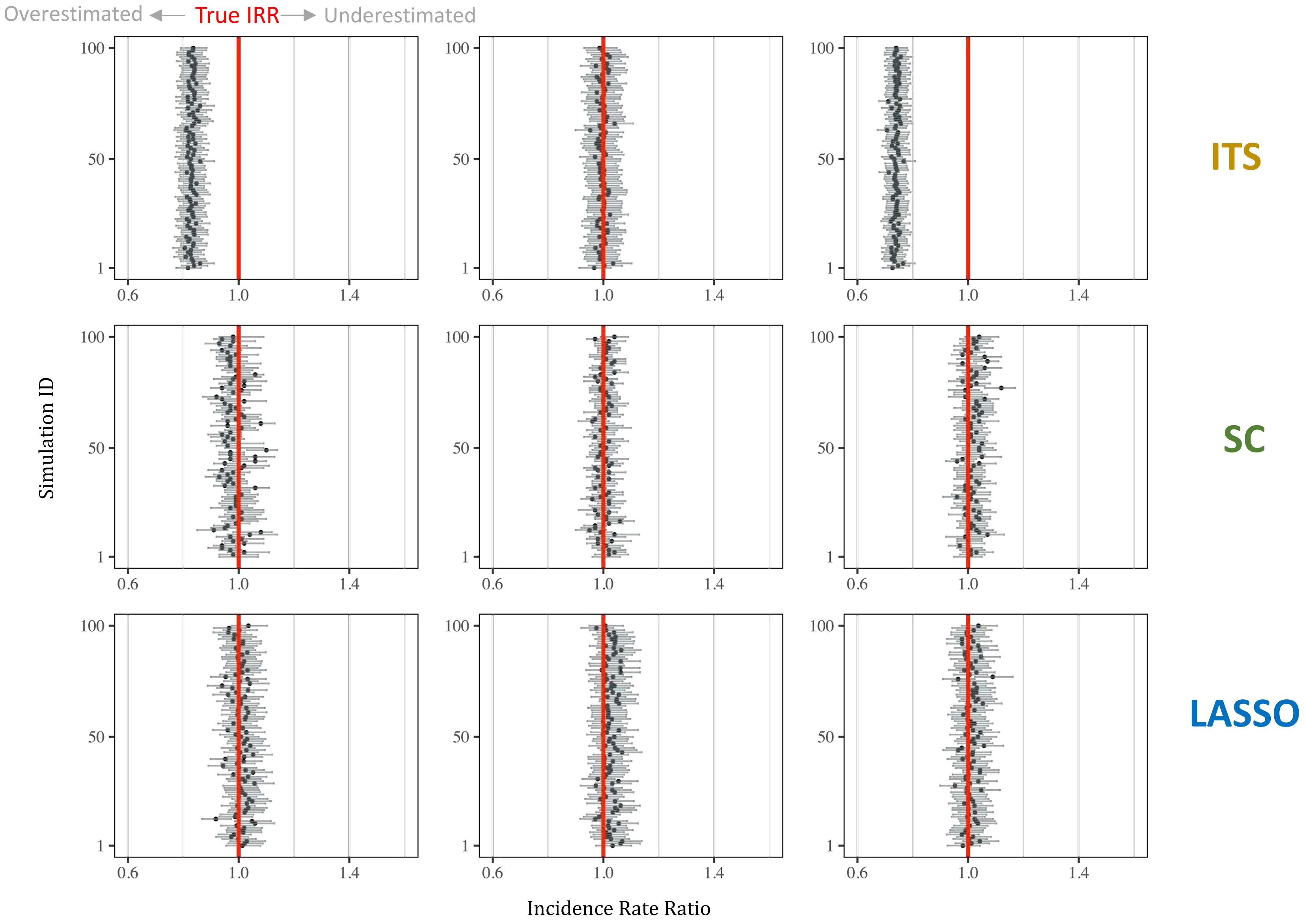


ITS

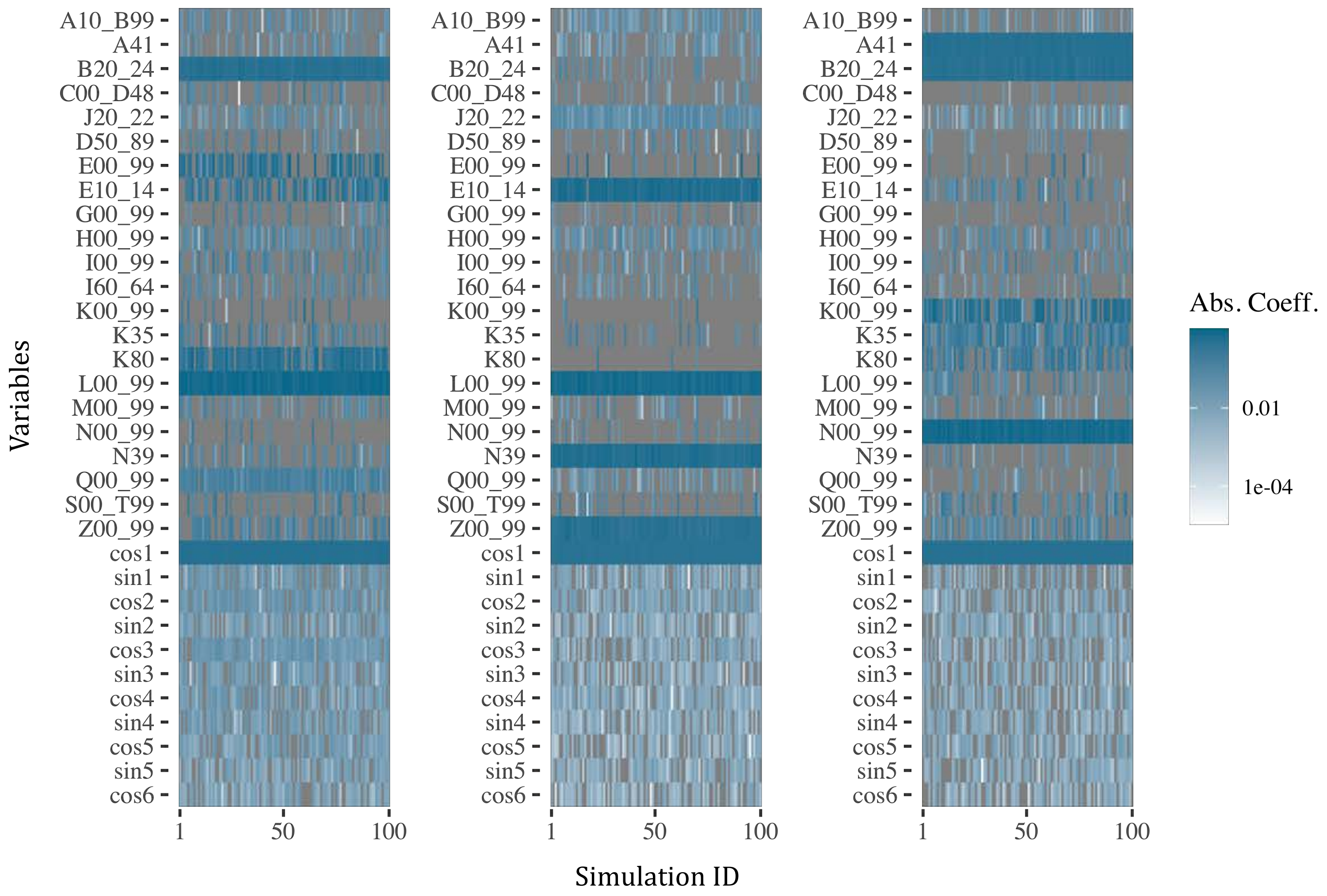
SC estimates were accurate across simulation scenarios



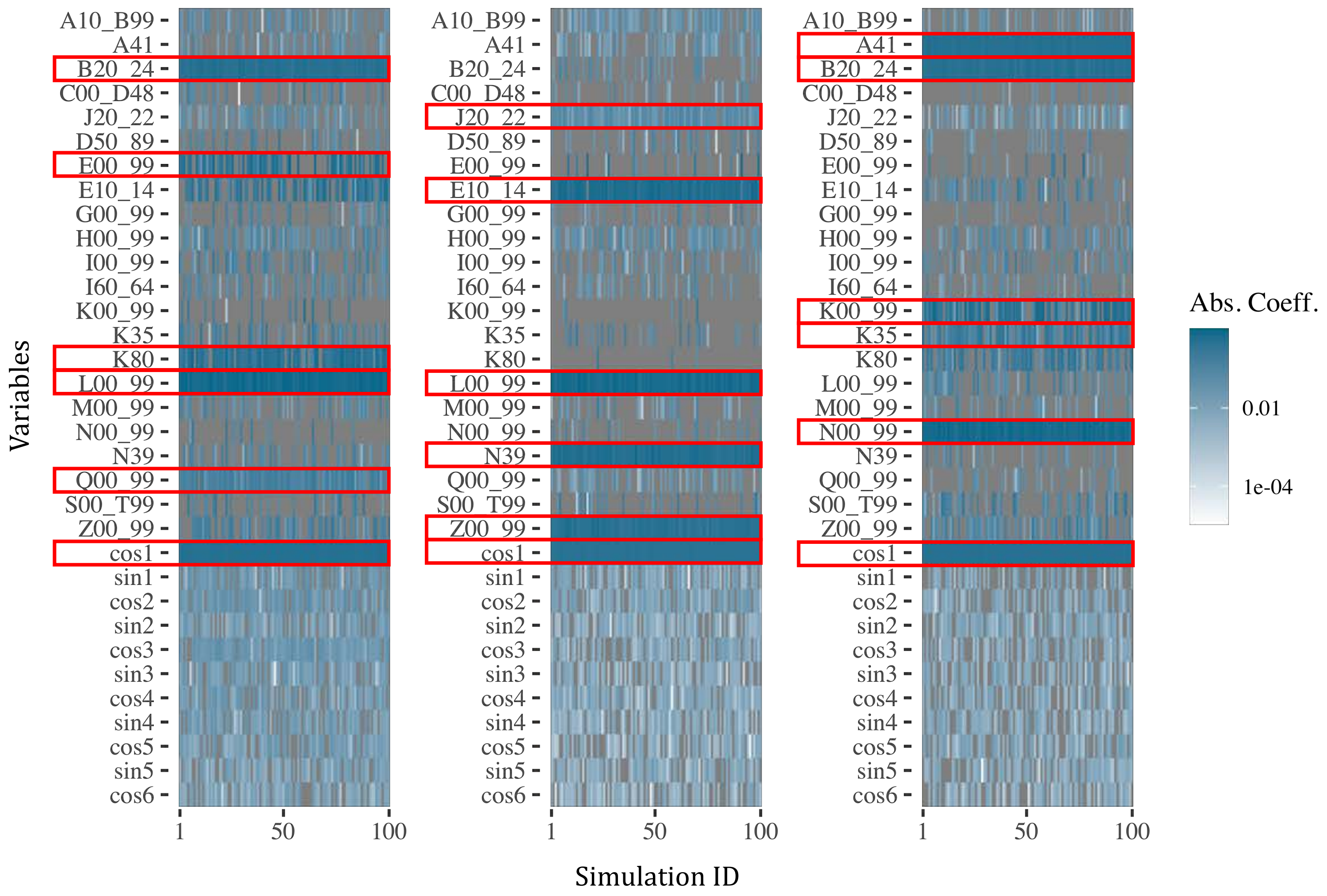
LASSO estimates were accurate across simulation scenarios



LASSO selected the controls used to simulate data



LASSO selected the controls used to simulate data



Take-home messages

- Nice features of LASSO method
  - Accurate estimation
  - Interpretable models
  - Easy to implement (pkg "glmnet"[1])
  - Can reducing confounding

1. Friedman J, Hastie T, Tibshirani R, et al. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. https://cran.r-project.org/web/packages/glmnet/index.html.

Take-home messages

- Nice features of LASSO method
  - Accurate estimation
  - Interpretable models
  - Easy to implement (pkg "glmnet"[1])
  - Can reducing confounding

- Some remaining challenges
  - Methods to obtain confidence intervals not readily available
  - Suboptimal performance in sparse data

1. Friedman J, Hastie T, Tibshirani R, et al. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. https://cran.r-project.org/web/packages/glmnet/index.html.

Take-home messages

- Nice features of LASSO method
  - Accurate estimation
  - Interpretable models
  - Easy to implement (pkg “glmnet”[1])
  - Can reducing confounding

- Some remaining challenges
  - Methods to obtain confidence intervals not readily available
  - Suboptimal performance in sparse data

Acknowledgement

Co-authors	*Helpful discussions*
Matthieu Domenech de Cellès	Daniel Weinberger
Sarah Kramer	Iris Artin
Marco Piccininni	Annette Aigner
Jessica L. Rohmann	Madlen Schranz
Tobias Kurth	Elizabeth Goult
Sylvie Escolano	Laura Barrero Guevara
Ulrike Grittner	

JOURNAL ARTICLE ACCEPTED MANUSCRIPT

Using LASSO regression to estimate the population-level impact of pneumococcal conjugate vaccines

Anabelle Wong, Sarah C Kramer, Marco Piccininni, Jessica L Rohmann, Tobias Kurth, Sylvie Escolano, Ulrike Grittner, Matthieu Domenech de Cellès

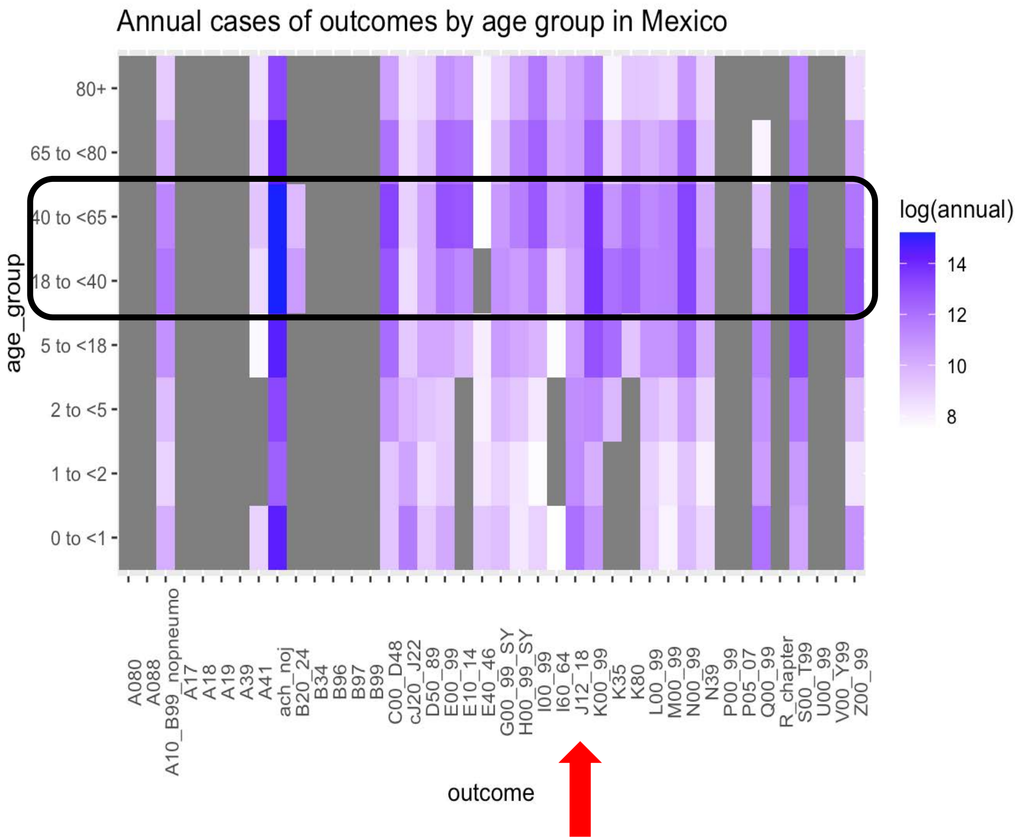
American Journal of Epidemiology, kwad061, https://doi.org/10.1093/aje/kwad061

Published: 17 March 2023 **Article history** ▾

1. Friedman J, Hastie T, Tibshirani R, et al. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. https://cran.r-project.org/web/packages/glmnet/index.html.

Q & A

We simulated outcome based on real data

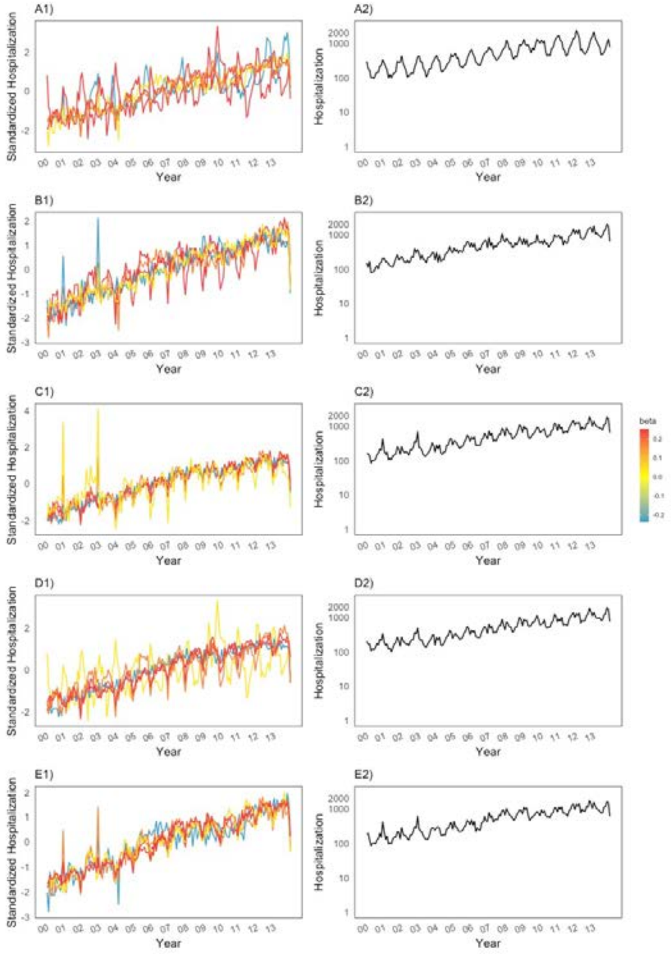


$$Y_t \sim \text{Poisson}(\mu_t)$$

$$\ln(\mu_t) = \alpha + \ln(NRH_t) + \sum_{i=1}^{n} \beta_i X_{it} + S_t + \gamma \mathbb{1}(t \geq t_{vac})$$

$$where\ \alpha = \ln\left(\frac{\bar{Y}}{\overline{NRH}}\right)$$

$$and\ S_t = \sum_{s=1}^{6} \delta_s \cos\left(\frac{2\pi st}{12}\right) + \sum_{s=1}^{5} \zeta_s \sin\left(\frac{2\pi st}{12}\right)$$



- Draw 5 controls & assign beta (x5)
- Draw 10 controls & assign beta (x5)
- 10% binomial subsample from 1st set (x1)

*Eliminate if annual max:min ratio > 10 (unrealistic)

Sensitivity test

- Instead of a null-impact vaccine, we tested a vaccine with VE=10%