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
The pneumococcal conjugate vaccines (PCVs)

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

1. Ganaie et al. (2020) *mBio*
The challenges in estimating PCV impact

• Population impact: direct effect + indirect effect\(^1\)

---

The challenges in estimating PCV impact

• Population impact: direct effect + indirect effect
• Indirect effect cannot be easily estimated in RCTs

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

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

How to estimate vaccine impact?

Counterfactual cases
Observed cases

PCV introduced

Defined evaluation period
How to estimate vaccine impact?

**Counterfactual cases**

**Observed cases**

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

$$IRR = 0.8 \Rightarrow 20\% \text{ reduction}$$
# How to predict counterfactual outcome?

<table>
<thead>
<tr>
<th>Rely on outcome of interest</th>
<th>Synthetic control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interrupted Time Series (ITS)</strong></td>
<td><strong>ITS + offset</strong></td>
</tr>
</tbody>
</table>

1. Thorrington et al. (2018) *BMC Medicine*
2. Bruhn et al. (2017) *PNAS*
# How to predict counterfactual outcome?

<table>
<thead>
<tr>
<th>Rely on outcome of interest</th>
<th>Synthetic control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interrupted Time Series (ITS)</strong></td>
<td><strong>ITS + offset</strong></td>
</tr>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
</tbody>
</table>

| ![Graph](image5) | ![Graph](image6) | ![Graph](image7) | **LASSO regression¹** |

---

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

---

¹. Bernal et al. (2017) *Int J Epidemiol*

---

- How to predict counterfactual outcome:
  - Rely on outcome of interest
  - Synthetic control
  - **Interrupted Time Series (ITS)**
    - ITS + offset
    - Hand-picked controls
  - **Data driven**
    - Bayesian variable selection
    - LASSO regression

---

Image: Bernal et al. (2017) *Int J Epidemiol*

Image: Bernal et al. (2018) *Int J Epidemiol*
How to predict counterfactual outcome?

<table>
<thead>
<tr>
<th>Rely on outcome of interest</th>
<th>Synthetic control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interrupted Time Series (ITS)</strong></td>
<td><strong>Hand-picked controls</strong></td>
</tr>
<tr>
<td>IT$\text{S} +$ offset</td>
<td><strong>Bayesian variable selection</strong></td>
</tr>
<tr>
<td><strong>Data driven</strong></td>
<td><strong>LASSO regression$^1$</strong></td>
</tr>
</tbody>
</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
SC estimates were accurate across simulation scenarios.
LASSO estimates were accurate across simulation scenarios

<table>
<thead>
<tr>
<th>Simulation ID</th>
<th>Incidence Rate Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ITS</td>
</tr>
<tr>
<td></td>
<td>SC</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
</tr>
</tbody>
</table>
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

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

Take-home messages

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

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

Acknowledgement

Co-authors
Matthieu Domenech de Cellès
Sarah Kramer
Marco Piccininni
Jessica L. Rohmann
Tobias Kurth
Sylvie Escolano
Ulrike Grittner

Helpful discussions
Daniel Weinberger
Iris Artin
Annette Aigner
Madlen Schranz
Elizabeth Goult
Laura Barrero Guevara

Q & A
We simulated outcome based on real data

\[ Y_t \sim \text{Poisson}(\mu_t) \]

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

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

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

- 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%