

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

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.

* Correspondence: wong@mpiib-berlin.mpg.de

Abstract: The pneumococcal conjugate vaccines (PCVs) protect against diseases caused by *Streptococcus pneumoniae*, such as meningitis, bacteremia, and pneumonia. It is challenging to estimate their population-level impact due to the lack of a perfect control population and the subtleness of signals when the endpoint – like all-cause pneumonia – is non-specific. Here we present a new approach to estimate the impact of PCVs – using LASSO regression to predict the counterfactual outcome for vaccine impact inference. First, we used a simulation study to test the performance of LASSO regression and established methods including the synthetic control (SC) approach. We found that LASSO achieved accurate and precise estimation, even in complex simulation scenarios where causal variables were removed. Then we applied LASSO to published real-world data from Chile, Ecuador, Mexico, and the US, and found that it yielded similar estimates of vaccine impact to SC. Overall, the performance of LASSO was comparable to SC. Our method is easy to implement and to interpret and can therefore be applied to study the impact of PCVs and of other vaccines.