UC Riverside
UC Riverside Previously Published Works

Title

Estimating the population-level impact of vaccines using synthetic controls

Permalink

https://escholarship.org/uc/item/9z17g7gf

Journal

Proceedings of the National Academy of Sciences of the United States of America, 114(7)

ISSN

0027-8424

Authors

Bruhn, Christian AW
Hetterich, Stephen
Schuck-Paim, Cynthia
et al.

Publication Date

2017-02-14

DOI

10.1073/pnas.1612833114

Peer reviewed

Estimating the population-level impact of vaccines using synthetic controls

Christian A. W. Bruhn[a], Stephen Hetterich[b], Cynthia Schuck-Paim[b], Esra Kürüm[a,c], Robert J. Taylor[b], Roger Lustig[b], Eugene D. Shapiro[a,d], Joshua L. Warren[a,e], Lone Simonsen[b,f,g], and Daniel M. Weinberger[a,1]

[a]Department of Epidemiology of Microbial Diseases, Yale University School of Public Health, New Haven, CT 06520; [b]Sage Analytica, Portland, ME 04101; [c]Department of Statistics, University of California, Riverside, CA 92521; [d]Department of Pediatrics, Yale School of Medicine, New Haven, CT 06520; [e]Department of Biostatistics, Yale School of Public Health, New Haven, CT 06520; [f]Milken Institute School of Public Health, George Washington University, Washington, DC 20052; and [g]Department of Public Health, University of Copenhagen, 1017 Copenhagen, Denmark

When a new vaccine is introduced, it is critical to monitor trends in disease rates to ensure that the vaccine is effective and to quantify its impact. However, estimates from observational studies can be confounded by unrelated changes in healthcare utilization, changes in the underlying health of the population, or changes in reporting. Other diseases are often used to detect and adjust for these changes, but choosing an appropriate control disease a priori is a major challenge. The "synthetic controls" (causal impact) method, which was originally developed for website analytics and social sciences, provides an appealing solution. With this approach, potential comparison time series are combined into a composite and are used to generate a counterfactual estimate, which can be compared with the time series of interest after the intervention. We sought to estimate changes in hospitalizations for all-cause pneumonia associated with the introduction of pneumococcal conjugate vaccines (PCVs) in five countries in the Americas. Using synthetic controls, we found a substantial decline in hospitalizations for all-cause pneumonia in infants in all five countries (average of 20%), whereas estimates for young and middle-aged adults varied by country and were potentially influenced by the 2009 influenza pandemic. In contrast to previous reports, we did not detect a decline in all-cause pneumonia in older adults in any country. Synthetic controls promise to increase the accuracy of studies of vaccine impact and to increase comparability of results between populations compared with alternative approaches.

pneumococcal conjugate vaccines | synthetic controls | *Streptococcus pneumoniae* | observational study | program evaluation

When a new public health intervention such as a vaccine is introduced, it is critical to monitor disease rates to ensure that the intervention is effective and to quantify its benefit. This information is important for policymakers working to prioritize use of health resources, but obtaining accurate estimates of vaccine impact is a major challenge.

Vaccine impact is typically evaluated by comparing rates or trends of the targeted disease in the years after introduction with rates or trends in the years before introduction using a time-series analysis (1–8). This type of analysis is used to estimate the total benefit of a new vaccine, in both vaccinated and unvaccinated individuals. However, these observational studies cannot easily disentangle changes in disease rates caused by the vaccine from changes caused by unrelated factors. Coincident changes in reporting patterns, access to care, clinical practice, and diagnostic coding practices, as well as social investments that influence the underlying health of the population, can either mask or exaggerate the true effect of a vaccine.

Evaluations of the impact of pneumococcal conjugate vaccines (PCVs) provide a case in point. PCVs target a subset (7, 10, or 13) of the more than 90 known serotypes of pneumococcus. Use of PCVs in infants nearly eliminated the targeted serotypes from vaccinated children and, through indirect effects, unvaccinated adults, with rates of invasive pneumococcal disease (IPD) in children declining by an average of 56% (range: 24–83%) within 3 y in high-income countries (9). However, the incidence of IPD is relatively small compared with the incidence of pneumonia caused by

pneumococcus (10), particularly in older adults (11). Therefore, to inform public health policy, it is critical to quantify the extent to which the vaccine reduces pneumonia hospitalizations in all age groups.

Detecting the population-level impact of vaccines on IPD is relatively straightforward because a large fraction of the cases is preventable by the vaccine, so the decline is dramatic. However, quantifying the impact of the vaccine on rates of pneumonia is more challenging because many pathogens can cause pneumonia, including other bacterial species, viruses, and fungi. Therefore, the relative decline in rates of pneumonia is expected to be far smaller than the decline in rates of IPD, as was observed in randomized controlled trials. For example, a trial in Latin America demonstrated that PCV10 had 67% efficacy against IPD, 22% efficacy against likely bacterial pneumonia, and 10% against radiologically confirmed community-acquired pneumonia (12). When evaluating trends in disease rates over time, these modest vaccine-related changes in rates of pneumonia are difficult to distinguish from background noise and secular trends. Thus, it is not surprising that published estimates of the change in rates of pneumonia after introduction of PCVs vary considerably (13, 14). Better approaches are needed to separate vaccine effects from unrelated changes.

Significance

Pneumococcus, a bacterial pathogen, is among the most important causes of pneumonia globally. Quantifying the impact of pneumococcal conjugate vaccines (PCVs) on pneumonia is challenging due to time trends unrelated to the vaccine. We use a method developed for website analytics and economics called "synthetic controls" to disentangle changes in pneumonia rates caused by the vaccine from changes caused by unrelated factors. We found that PCVs significantly reduce all-cause pneumonia hospitalizations in young children, and reduce hospitalizations for invasive pneumococcal disease and pneumococcal pneumonia in children and adults. In contrast to previous studies, we did not detect a decline in all-cause pneumonia hospitalizations in older adults in any of the five countries following the introduction of the vaccine in children.

To strengthen the argument that an observed decline in disease rates is, in fact, caused by a vaccine, many studies compare changes in the disease of interest to changes in other diseases during the same time period (7, 8). The goal of these comparative analyses is to detect unmeasured confounding and bias in the time series of interest. The general reasoning is that if the disease of interest declines relative to the comparison disease after vaccine use begins, then the decline can be attributed to the vaccine. However, there is no standard methodology for selecting these comparison diseases or using them to adjust estimates of vaccine-associated changes. Ideally, diseases used for comparison would share the same set of potential biases and causal factors as the outcome of interest but would not be influenced by the vaccine (15, 16). Identifying a comparison disease that reasonably fulfills these criteria a priori is a major challenge and requires many qualitative decisions by the analyst.

A method developed for website analytics and economics research, "synthetic controls," provides a potential solution to this problem. With this method, a number of time series that are unaffected by the intervention are optimally weighted according to their fit to the outcome of interest in the period before the intervention, then combined into a composite time series. The synthetic composite gives more weight to the predictor variables that jointly explain the outcome variable best (17, 18). The values for this composite synthetic control are then estimated in the post-intervention period, which produces a counterfactual estimate of what would have happened in the postintervention period had the intervention not occurred, effectively adjusting for unmeasured bias and confounding. This general approach has been applied to various problems, including evaluation of the effects of advertising on website traffic (using websites unaffected by the ad to construct the synthetic control) (18) and evaluation of tobacco control programs on smoking rates (using states that did not have a program to construct the synthetic control) (17). The increasing availability of large electronic healthcare databases now allows this approach to be applied to evaluations of public health interventions.

In this study, we demonstrate how the synthetic control method [causal impact approach of Brodersen et al. (18)] can be adapted to detect and adjust for unmeasured bias and confounding in the evaluation of vaccine programs. Specifically, we applied the synthetic controls method to nationwide administrative databases in Brazil, Chile, Ecuador, Mexico, and the United States to evaluate changes in the burden of hospitalizations for all-cause pneumonia associated with the introduction of PCVs. We show that this approach can effectively adjust for unexplained trends in the data from these five countries and offers important improvements over other commonly used approaches. This approach has the potential to increase the accuracy and improve the comparability of studies of vaccine impact between locations with different local biases and trends.

Results

Unexplained Factors Bias Estimates of Vaccine Impact. We used nationally aggregated, age-specific hospitalization time-series data from Brazil, Chile, Ecuador, Mexico, and the United States (Fig. 1 and Tables S1 and S2). To evaluate changes in pneumonia hospitalizations that occurred after vaccine introduction, we first performed a simple comparison of the number of cases of all-cause pneumonia before and after vaccine introduction, using nonrespiratory hospitalizations as an offset. Among <12-mo-old children, the number of hospitalizations for pneumonia was substantially lower than the counterfactual in the postvaccine period in Brazil [−22%; 95% credible interval (CI): −27%, −17%], Mexico (−17%; 95% CI: −28%, −4%), and the United States (−19%; 95% CI: −23%, −14%). In contrast, the number of hospitalizations for pneumonia did not significantly change compared with the counterfactual in Chile (−15%; 95% CI: −29%, +2%), and increased in Ecuador (+17%; 95% CI: +1%, +35%) (Figs. 1 and 2*A* and Table S3). In adults 80+ years of age, the results also varied between countries when using this model. Notably, in Brazil and Ecuador, hospitalizations for pneumonia increased by 25% (95% CI: 18%, 32%) and 19% (95% CI: 6%, 33%) compared with the counterfactual (Fig. 2*A* and Table S3).

The increase in rates of hospitalization for pneumonia among older adults in Brazil and Ecuador suggested that factors other than the vaccine had influenced the observed trends. To put these patterns into context, we repeated this simple analysis for a broad range of comparison disease categories using data from Brazil (Table S2). These analyses demonstrated that hospitalizations for many other disease categories also increased substantially in the same period among 80+-year-olds (Fig. S1*B*). Among <12-mo-old children, the changes in these other disease categories after PCV10 was introduced were more varied, with both significant increases and decreases (Fig. S1*A*).

Use of Synthetic Controls to Adjust for Unexplained Trends. To adjust for unexplained trends in the data, we used the synthetic control approach. These analyses yielded several notable results. For all five of the countries, the introduction of PCVs was associated with substantial and significant declines in hospitalizations for all-cause pneumonia among children <12 mo of age (Figs. 2 and 3). The decline showed a similar trajectory in Brazil, Ecuador, Mexico, and the United States, and was more pronounced in Chile (though, as discussed in *Sensitivity Analyses and Alternative Models*, the estimates for Chile were less robust to the model structure; Fig. 3). The average decline across all five countries was −20% (95% CI: −27%, −12%). However, among 65- to 79-year-olds and 80+-year-olds, we did not detect a decline in hospitalizations for pneumonia after vaccine introduction in children in any of the five countries; for example, the estimate for 80+-year-olds was +3%; 95% CI: −4%, +10% (Figs. 2 and 3 and Table S3). For the other
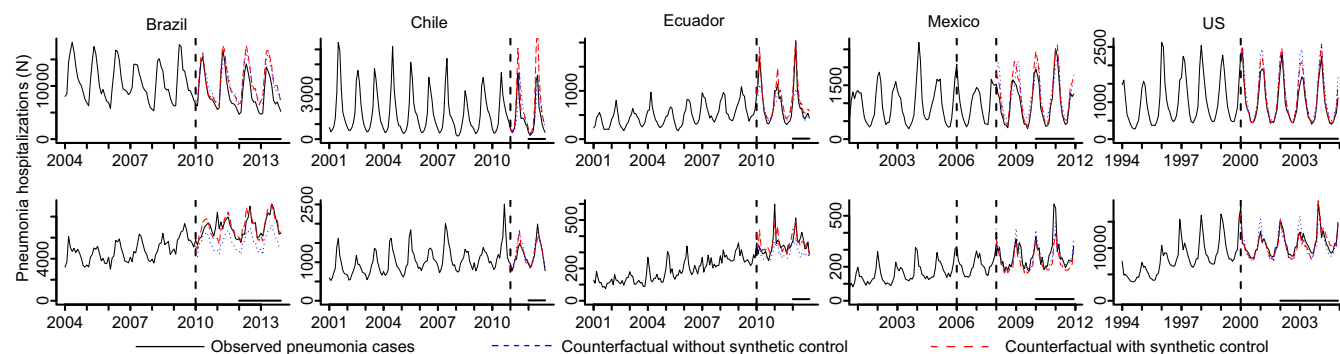
Fig. 1. Time series of pneumonia hospitalizations (cases) from five countries (black line) among (*A*) children <12 mo and (*B*) adults 80+ y, and counterfactual estimates during the postvaccine period from a model that only adjusts for the number of nonrespiratory hospitalizations and seasonality (blue dotted line) or the synthetic control model (red dashed line). The vertical dashed line indicates the time of PCV introduction (the date of introduction varied by region in Mexico). The black horizontal line indicates the period during which the observed/counterfactual rate ratio was calculated.
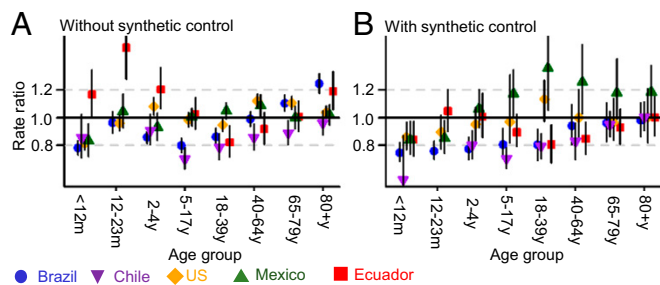
Fig. 2. Changes in hospitalizations for all-cause pneumonia (rate ratio) in five countries using (*A*) a model without synthetic controls and adjusted for nonrespiratory hospitalizations and seasonality only or (*B*) a model with synthetic controls. Brazil (blue circle), Chile (purple inverted triangle), United States (yellow diamond), Mexico (green triangle), and Ecuador (red square). Rate ratios were calculated as the sum of observed hospitalizations for all-cause pneumonia divided by the sum of predicted (counterfactual) hospitalizations during the evaluation period.

age groups, the estimated change differed between countries (Table S3). Notably, the significant declines in older children and young/middle-aged adults in some of the countries might have been biased by the inclusion of the 2009 influenza pandemic during the fitting period (sensitivity analyses below; Figs. S2 and S3).

In some strata, the estimates were similar between the model that simply adjusted for nonrespiratory hospitalizations and the model that adjusted for the synthetic control (e.g., declines of −25% vs. −22% in hospitalizations for all-cause pneumonia among children <12 mo of age in Brazil with and without a synthetic control, respectively; Fig. 1 and Table S3). This consistency would be expected if there are no unexplained biases in the data. However, in other strata, the synthetic control adjustment changed the estimates substantially. Among 80+-year-olds in Brazil, the model without the synthetic control estimated a 25% (95% CI: 18%, 32%) increase in pneumonia rates associated with introduction of PCV, whereas the model with a synthetic control estimated no change in pneumonia rates (−2%, 95% CI: −14%, +11%). The weight given to different variables in the synthetic control differed by age group and country (*SI Results* and Dataset S1).

Validation of Synthetic Controls Using Other Categories of Pneumococcal Disease. Analyses of other categories of pneumococcal disease can be used to evaluate the epidemiological credibility of the estimates of the declines in all-cause pneumonia. The relative decline in disease rates should be correlated with the specificity of the case definition—more-specific definitions (e.g., IPD, pneumococcal pneumonia) should have greater relative declines than less-specific definitions (e.g., all-cause pneumonia) because a larger fraction of cases are caused by vaccine-type pneumococci. When using the synthetic control model, the estimates followed the expected pattern: for children <12 mo of age, the greatest estimated decline was for IPD (−59%; 95% CI: −66%, −52%), followed by pneumococcal/lobar pneumonia (−39%, 95% CI: −66%, −24%), followed by a more-specific definition of all-cause pneumonia (19) (−26%; 95% CI: −34%, −16%), and finally the less-specific definition of all-cause pneumonia that was used for the main analyses (−14%; 95% CI: −23%, −3%; Fig. 4*A*). The same pattern of greater reduction with higher specificity was seen for the point estimates for all of the other age groups (Fig. 4*A*). In contrast, estimates from models that only adjusted for nonrespiratory hospitalizations or linear trend did not follow this expected pattern (Fig. 4 *B* and *C*).

Sensitivity Analyses and Alternative Models. To validate the synthetic control approach, we divided the prevaccine data from each country into different 48-mo training periods and 12-mo evaluation periods. The counterfactual estimates matched the observed

data (rate ratio of 1) in most instances for the <5-y-old age group and age groups over 65 years when using the synthetic controls (Fig. S2). In Brazil and Chile, the models did not perform well (credible intervals for the rate ratios did not include 1) for the age groups that were most affected by the 2009 influenza pandemic (older children and young and middle-aged adults) when the postvaccine evaluation windows included the pandemic period (*SI Results* and Fig. S2). The estimates for Chile and Ecuador were also more subject to short-term variations than for the other countries (Fig. S2). As a further sensitivity analysis, we used different prevaccine training periods and estimated the rate ratios using the same evaluation period; the credible intervals for the rate ratios in older children and adults mostly overlapped 1 when excluding 2009 from the training period (Fig. S3).

The sensitivity of the results to the exclusion of specific variables as potential components of the control differed by country and age group, with highly consistent results for Brazil and the United States, and less-consistent results for the other countries (Fig. S4). The estimates for children <12 mo of age in Ecuador, Mexico, and Chile were sensitive to the specific variables included. In particular, J20–J22 (bronchitis/bronchiolitis) was an influential variable. The exclusion of J20–J22 generally increased heterogeneity of results between studies among children <12 mo and 12–23 mo of age compared with the models that included bronchitis/bronchiolitis (Table S3). In some instances, the models excluding J20–J22 did not properly adjust for pneumonia epidemics (leading to smaller estimates of vaccine-related declines). A notable exception is the estimates for <1-y-old children in Chile, where the effect of the vaccine was possibly overestimated when J20–J22 was included (Figs. 1–3).

We also evaluated two alternative models (*SI Results* and Table S3) that had a simple adjustment for linear trend or used nonrespiratory hospitalizations as the sole covariate in the model (rather than as an offset). In some instances, these models gave results similar to the synthetic control model, but they were less likely to detect an effect of the vaccine in children, and, as noted above, the rate ratio estimates from these alternative models did not follow the expected patterns among different categories of pneumococcal disease, whereas the estimates from the synthetic control models did.

Discussion

We have used synthetic controls to assess the population-level impact on pneumonia of vaccinating children with PCVs. In so doing, we have shown that using synthetic controls provides a promising way to adjust for unmeasured bias and confounding in observational studies that evaluate the impact of public health interventions. Our results showing a reduction in all-cause pneumonia following introduction of PCVs in infants, and to some extent in young children, agree with previous studies. We also confirmed that invasive pneumococcal disease and pneumococcal pneumonia declined in children and adults. However, contrary to previous reports (7, 20), we did not detect a reduction in all-cause pneumonia among adults 65 y and older in
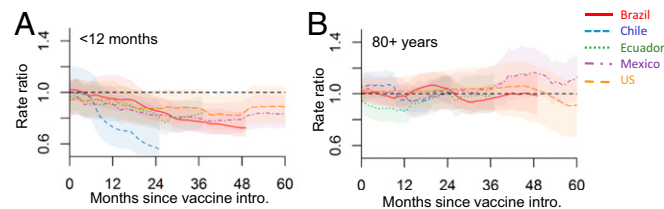
Fig. 3. Changes in hospitalizations for all-cause pneumonia by months since vaccine introduction in Brazil (red), Chile (blue), Ecuador (green), Mexico (purple), and the United States (orange) in (*A*) children <12 mo and (*B*) adults 80+ y; estimates are from the model adjusted using a synthetic control. In each month, the rate ratio is the sum of observed all-cause pneumonia hospitalizations during the previous 12 mo divided by the sum of predicted (counterfactual) hospitalizations in the same period.
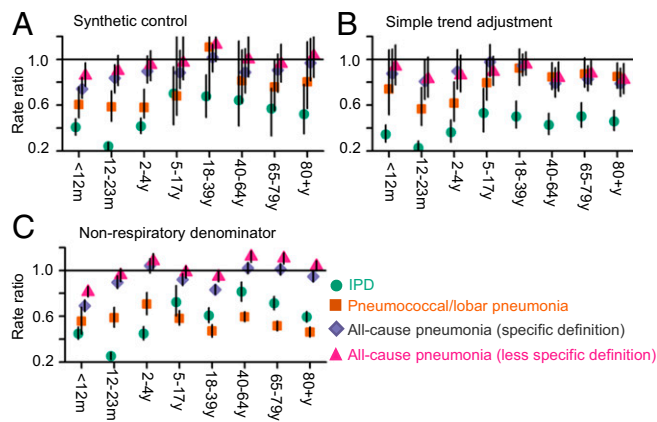
Fig. 4. Changes in hospitalizations in the United States (rate ratios) due to IPD (turquoise), lobar/pneumococcal pneumonia (orange), all-cause pneumonia [Grijalva and coworkers' (19) definition, purple], all-cause pneumonia (as defined in this study, pink) adjusted using (*A*) synthetic controls; (*B*) linear trend, nonrespiratory hospitalizations (as the denominator), and seasonality; and (*C*) only nonrespiratory hospitalizations (as the denominator) and seasonality.

any of the five countries. This is notable because the burden of all-cause pneumonia in this age group is large, and the fraction that is indirectly preventable by vaccinating children might be smaller than previously estimated (7, 20). Declines were also observed in older children and younger adult age groups in some countries, but validation and sensitivity analyses suggest that some of these declines were biased by inclusion of the 2009 influenza pandemic during the prevaccine period.

Our analyses show that although observational studies of vaccine impact come with multiple caveats, use of synthetic controls has the potential to improve the accuracy of the estimates. This approach to adjusting for unmeasured bias and confounding, irrespective of source, increases comparability of such assessments across settings.

Use of Synthetic Controls to Evaluate Impacts of Vaccination. Aside from conducting a large and expensive cluster-randomized trial, time trend analyses are the only method that can quantify both the direct effects of being vaccinated and the indirect effects that result from disrupting transmission. However, because any number of factors can influence disease rates around the time of vaccine introduction, time-series analyses are susceptible both to incorrectly attributing unrelated changes to a vaccine and to missing a vaccine-related change hidden by an unrelated factor. Analyses that incorporate synthetic controls to adjust for unmeasured bias and confounding have several distinct advantages over approaches currently used; first, they are more likely to correctly control for these unrelated changes than other methods, potentially increasing accuracy as well as comparability of results between studies and study settings. Second, the analyst is not required to handpick individual comparison diseases a priori, a process that itself can introduce bias; instead, the method combines time series of hospitalizations for many disease outcomes into a single synthetic composite time series, with each component weighted according to its fit to the disease of interest in the preintervention period. The result is a time series that is more likely to behave like the disease of interest in the postintervention period than any single comparator. Third, the synthetic control eliminates the need to assume that any trend beginning in the prevaccine period continues linearly into the postvaccine period, as is done with the commonly used interrupted time-series analysis. Fourth, the analytic framework can readily be used to incorporate any time series relevant to the outcome of interest into the synthetic control. For example, time series of virological data [e.g., confirmed influenza or respiratory syncytial virus (RSV) infections] could be used to

adjust epidemics of pneumonia, assuming they are not themselves influenced by the vaccine.

Synthetic controls can also be implemented for different types of study designs and different levels of data availability. For example, synthetic controls could be used for prospective studies of pneumonia trends or to adjust for changes in ascertainment over time for studies of invasive bacterial diseases (using time series of other pathogens as components of the synthetic control) (21, 22). In settings where it is not feasible to extract a large number of time series to use as components of the synthetic control (e.g., in a setting with only paper health records), the approach might still be useful, albeit with fewer contributing time series. In such a situation, studies from countries with electronic data might be used to guide selection of time series that are most likely to be informative. In all of these scenarios, the synthetic control approach provides the important advantage of accounting for model uncertainty (and subsequent uncertainty in the estimates of the counterfactual) in a way that current approaches do not.

To use the synthetic control approach optimally, we propose the following guidelines. First, we suggest presenting results from both a model that adjusts only for all-cause hospitalizations or population size and one that adjusts using the synthetic control; presenting results from both models increases the likelihood of detecting potential biases, as was the case for the older adults in Brazil. Second, we suggest presenting results of sensitivity analyses in which the most heavily weighted components of the control are removed, and the models are rerun (Fig. S4). Third, we suggest validating the model by using one subset of prevaccine data to predict values in a second subset of prevaccine data (Fig. S2), with the expectation that there should be no effect if the relationships among the variables are stable. This approach can help to identify potential issues with the analyses, as with the effect of the 2009 influenza pandemic in older children and young/middle-aged adults in Brazil (Fig. S3). Fourth, whenever possible, analyses of positive controls—time series for which one is highly confident that an intervention will have an impact—should be included; for example, invasive pneumococcal disease in the case of PCVs.

Several caveats must be noted when interpreting the results. The synthetic controls method makes two major assumptions: that use of PCVs does not affect the incidence of the diseases that are components of the synthetic control, and that the only change in the relationship between pneumonia and the components of the synthetic control over time is caused by the vaccine. If the contributors to the synthetic control have themselves been subject to a separate intervention (e.g., rotavirus vaccine), this assumption would be violated. If a disease time series is expected to be affected by the vaccine, it must be excluded from the synthetic control; likewise, if reporting or coding for pneumonia changes over time (e.g., due to awareness of pneumonia as a public health issue following PCV introduction), the assumption that the relationship among the time series is consistent over time would be violated. Likewise, some factors affecting the outcome of interest might not be captured by the time series available for inclusion in the synthetic control—for instance, the 2009 influenza pandemic was not fully adjusted by the model, and therefore the relationship between pneumonia and the components of the synthetic control change during this time period. If the assumptions of the synthetic controls model are fulfilled, then the difference between the observed and counterfactual values can be interpreted as the causal effect of the vaccine (16). However, as with any observational study, results must be interpreted cautiously; synthetic controls can help to reduce the effects of unmeasured bias and confounding, but they are unlikely to completely eliminate it.

Impacts of PCVs on Pneumococcal Disease. Assessments of vaccine impact are a particular challenge because there is no "ground-truth" estimate for how much disease rates should be expected to decline following vaccine introduction. We get some approximation from randomized controlled trials and case-control analyses, both of which provide an estimate of the direct protection obtained by receiving the vaccine, although these do not capture the indirect

effect of the vaccine that results from disrupting transmission. The estimates of the PCV-related reduction in hospitalizations of infants for all-cause pneumonia produced by the synthetic controls approach are larger than the both 10% against clinical pneumonia found by a randomized controlled trial in Latin America (12) and the 11% reduction in pneumonia hospitalizations in children less than 2 y of age found by a case-control study in Chile (23). Given that trend analyses capture both direct and indirect effects of the vaccine, it is reasonable that our estimates would be larger than those found by clinical trials or case-control studies, which only capture direct effects. Estimates from time trend analyses have been more varied. For instance, a 13% reduction in clinical pneumonia in children <12 mo was found from a population-based surveillance study in Brazil (24), whereas estimates of ~20% and 40% were found using administrative databases from Brazil and the United States, respectively (7, 25).

Aside from comparing estimates from time trend studies to randomized controlled trials, the best ground-truth estimate of vaccine impact that we can obtain is to quantify changes in more-specific categories of pneumococcal disease, where high-quality surveillance studies provide an estimate of the expected decline. When evaluating categories of pneumococcal disease with different specificities, the relative decline should be greater for disease categories that are more specific for pneumococcus (12, 26). Both the synthetic control models and models with simpler adjustments gave estimates of declines in IPD that were comparable to estimates from active surveillance data from the United States (27). The advantage of the synthetic control approach can be seen when comparing estimates of vaccine-related changes in four categories of pneumococcal disease. The point estimates from the synthetic controls model followed the expected patterns, with greater declines for more specific definitions across most age groups (Fig. 4). In contrast, using simpler adjustment methods, the expected pattern was only seen among children.

In contrast to some previous studies, we found no vaccine-related reductions in pneumonia among people age 65+. Although one trend study from the United States reported no decline in confirmed or presumptive hospitalized pneumonia cases in this age group (28), others that used large administrative databases found significant PCV-related reductions in pneumonia hospitalization in seniors (7, 8, 19, 20). Our secondary analyses, in which we use a simple linear trend to make the adjustment, might explain this difference. Trend-line adjustments in the other studies might have been skewed upward by severe pneumonia seasons during the prevaccine years, thereby leading to larger apparent vaccine impact when extrapolated into the postvaccine period. Synthetic control adjustments are less sensitive to such short-term epidemics than are linear trends, even if they are not immune.

There are at least two possible explanations for our results in older adults; one is that any indirect effect on all-cause pneumonia in older adults is either nonexistent or too small to estimate reliably from time-series data due to insufficient power. Although we did find substantial declines in hospitalizations for IPD and lobar/pneumococcal pneumonia among people 65+ y in the United States after introduction of PCV7, vaccine-type pneumococci might comprise a smaller-than-expected fraction of all-cause pneumonia cases than previously thought. It is also possible that increases in pneumonia hospitalizations due to other pathogens either offset any indirect benefit or mask declines due to pneumococcus. A second possibility is that our synthetic control approach might not fully adjust for unmeasured bias and confounding, so true declines might be masked. Regardless, the burden of pneumonia in older adults is tremendous, so small relative declines (even if they cannot be reliably measured from time-series data) could translate into large numbers of cases prevented.

We found some unexplained variability in estimates of vaccine-associated changes between countries, particularly among young and middle-aged adults; these were the age groups with the highest hospitalization rates during the 2009 pandemic, and significant declines in pneumonia in these age groups were notably only found in countries where the baseline period included the pandemic. Including the pandemic during the fitting period has the potential to bias the counterfactual upward and thus exaggerate the estimated declines, which the sensitivity analyses in Figs. S2 and S3 suggest might have occurred. Alternative approaches to adjust for the pandemic (including the inclusion of time series of virological data as components of the synthetic control) could help to adjust for these biases. The variability in the estimated impacts of the vaccine between countries for some age groups highlights the perils of trying to interpret data from a single country on its own, without context.

Our results were sensitive to some modeling choices and not to others. Inclusion of bronchiolitis/bronchitis (J20–J22) time series as a component of the synthetic control had a pronounced effect on the estimates; for instance, the estimated magnitude of a vaccine-associated effect on all-cause pneumonia for each age group in Chile was substantially larger when bronchiolitis/bronchitis (J20–J22) was included in the synthetic control (Table S3). The validity of including bronchiolitis/bronchitis in the model is debatable. On one hand, bronchiolitis/bronchitis are acute respiratory illnesses strongly associated with RSV in young children; they likely share many causal mechanisms with all-cause pneumonia and can potentially explain short-term epidemic patterns (i.e., if epidemics of pneumonia are also caused by RSV or some other shared environmental factor). However, PCVs could plausibly influence the rate of bronchiolitis/bronchitis hospitalizations as indicated by results showing RSV-associated disease decreasing in the United States following PCV introduction, in violation of one of the model assumptions (29).

Although the use of synthetic controls is not a guarantee against bias and confounding, we believe that this approach will increase the probability of obtaining accurate and interpretable estimates of vaccine impact that can be shared with public health decision-makers. Despite the complex computing involved, using the synthetic control method is straightforward, thanks to an R package developed by Brodersen et al. (18). We include code for implementing this package with administrative hospitalization data (Datasets S2–S4). The adoption of this methodology has the potential to improve the quality and transparency of vaccine benefit estimates as well as the comparability between studies conducted in different settings.

Methods

Data Sources and Definitions. We used routinely collected administrative data on hospitalizations from Brazil, Chile, Ecuador, Mexico, and 10 states in the United States (Table S1). The unifying characteristic of the five databases is that they all use International Classification of Diseases (ICD) codes to classify the cause of hospitalizations. The primary outcome for all analyses was all-cause pneumonia. For the components of the synthetic control, we used the ICD chapters, which group disease codes thematically (e.g., infectious and parasitic diseases, diseases of the respiratory system, etc.). Some codes that identified specific diseases were also used. Codes that could potentially be influenced by the introduction of PCVs were excluded from the covariates (*SI Methods*). The Human Investigation Committee at Yale University determined that this study is exempt from review.

Preprocessing Data. All of the time series were log-transformed and standardized before being used for analysis. This transformation helped to minimize the effects of epidemics on the long-term trends and associations. After model fitting and before calculating the rate ratios and cases prevented, the observed and predicted values were transformed back to the original scale. For data from Brazil, additional preprocessing was required to adjust for an abrupt shift in coding that occurred at the start of 2008 due to a change in reimbursement policies (*SI Methods*).

Estimating Vaccine-Associated Declines. When estimating the effect of a vaccine using time-series data, it is necessary to compare the observed number of cases with a counterfactual estimate of what the number of cases would have been if the vaccine had not been introduced. We compare the synthetic control approach with other approaches that are commonly used to generate counterfactuals. In a simple pre/postanalysis (model 1; Eq. **1**), the counterfactual for the postvaccine period is based on the seasonally adjusted mean number of cases

for the prevaccine period. We adjusted for seasonal variations using indicator variables for month of the year [$c_k*I\{month_k = m(t)\}$]. Changes in population size or changes in the volume of nonrespiratory hospitalizations (NRH) were adjusted using an offset:

$$\log(\text{pneumonia cases}_t) = \beta_0 + \sum_k \left[c_k * I\{month_k = m(t)\} \right] + \log(\text{NRH}_t). \quad [1]$$

With the synthetic control approach [causal impact method of Brodersen et al. (18)] (model 2; Eq. **1**), we used time series for specific other diseases to adjust for trends unrelated to the vaccine. The control time series were incorporated into a regression model fit to pneumonia data from the pre-vaccine period. Bayesian variable selection was used to weight the different control time series, effectively giving more weight to those predictor variables that jointly explain the outcome variable best. These weights are used to generate a composite control variable (the synthetic control), which adjusts the counterfactual for changes in these other time series. Notably, log (NRH$_t$) can be included as one of the potential control variables.

$$\log(\text{pneumonia cases}_t) = \beta_0 + \sum_k \left[c_k * I\{month_k = m(t)\} \right] +$$
$$\sum \left[\text{binary\_inclusion\_indicator}_j * \text{coeff}_j * \log(\text{control\_time series}_{jt}) \right] \quad [2]$$

In this model, $\Sigma[\text{binary\_inclusion\_indicator}_j*\text{coeff}_j*\log(\text{control\_time series}_{jt})]$ is the set of control time series, weighted by their inclusion indicators and regression coefficients. The binary_inclusion_indicator$_j$ are independent Bernoulli (π)-distributed random variables that determine the presence or absence of a particular control series in the model (0: excluded; 1: included).

For the variable selection step, spike and slab priors were used with equal prior probability for inclusion for all covariates (18), and this probability was set so that the prior inclusion probability for each variable was $\pi = 0.5$. We also evaluated models that excluded J20–J22 as a potential component of the synthetic control (a heavily weighted covariate that might potentially be influenced by the vaccine), as well as models that performed simple adjustments for linear trend or that included log(NRH$_t$) as the only covariate (rather than as an offset) (*SI Results*). Each of the models was fit to pneumonia data from the pre-PCV period and used to generate counterfactual estimates for a post-PCV evaluation period. We used the bsts (30) and CausalImpact (18) packages in R for model fitting. For details on the calculation of the counterfactuals and rate ratios and additional details on the model, see *SI Methods*. The R code used for these analyses and the time series are included in Datasets S2–S10 and can be found at https://github.com/weinbergerlab/synthetic-control/.

1. Hariri S, et al.; HPV-IMPACT Working Group (2015) Population-based trends in high-grade cervical lesions in the early human papillomavirus vaccine era in the United States. *Cancer* 121(16):2775–2781.
2. Dijkstra F, Donker GA, Wilbrink B, Van Gageldonk-Lafeber AB, Van Der Sande MA (2009) Long time trends in influenza-like illness and associated determinants in The Netherlands. *Epidemiol Infect* 137(4):473–479.
3. Chang DH, et al. (2016) Trends in U.S. hospitalizations and inpatient deaths from pneumonia and influenza, 1996-2011. *Vaccine* 34(4):486–494.
4. Davis MM, King JC, Jr, Moag L, Cummings G, Magder LS (2008) Countywide school-based influenza immunization: Direct and indirect impact on student absenteeism. *Pediatrics* 122(1):e260–e265.
5. Richardson V, et al. (2010) Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. *N Engl J Med* 362(4):299–305.
6. do Carmo GMI, et al. (2011) Decline in diarrhea mortality and admissions after routine childhood rotavirus immunization in Brazil: A time-series analysis. *PLoS Med* 8(4):e1001024.
7. Grijalva CG, et al. (2007) Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: A time-series analysis. *Lancet* 369(9568):1179–1186.
8. Simonsen L, et al. (2014) Effect of 13-valent pneumococcal conjugate vaccine on admissions to hospital 2 years after its introduction in the USA: A time series analysis. *Lancet Respir Med* 2(5):387–394.
9. Feikin DR, et al.; Serotype Replacement Study Group (2013) Serotype-specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: A pooled analysis of multiple surveillance sites. *PLoS Med* 10(9):e1001517.
10. O'Brien KL, et al.; Hib and Pneumococcal Global Burden of Disease Study Team (2009) Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: Global estimates. *Lancet* 374(9693):893–902.
11. Said MA, et al.; AGEDD Adult Pneumococcal Burden Study Team (2013) Estimating the burden of pneumococcal pneumonia among adults: A systematic review and meta-analysis of diagnostic techniques. *PLoS One* 8(4):e60273.
12. Tregnaghi MW, et al.; COMPAS Group (2014) Efficacy of pneumococcal nontypable *Haemophilus influenzae* protein D conjugate vaccine (PHiD-CV) in young Latin American children: A double-blind randomized controlled trial. *PLoS Med* 11(6):e1001657.
13. Loo JD, et al. (2014) Systematic review of the indirect effect of pneumococcal conjugate vaccine dosing schedules on pneumococcal disease and colonization. *Pediatr Infect Dis J* 33(Suppl 2):S161–S171.
14. Loo JD, et al. (2014) Systematic review of the effect of pneumococcal conjugate vaccine dosing schedules on prevention of pneumonia. *Pediatr Infect Dis J* 33(Suppl 2):S140–S151.
15. Lipsitch M, Tchetgen Tchetgen E, Cohen T (2010) Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* 21(3):383–388.
16. Lipsitch M, Jha A, Simonsen L (2016) Observational studies and the difficult quest for causality: Lessons from vaccine effectiveness and impact studies. *Int J Epidemiol*, 10.1093/ije/dyw124.
17. Abadie A, Diamond A, Hainmueller J (2012) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J Am Stat Assoc* 105(490):493–505.
18. Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL (2015) Inferring causal impact using Bayesian structural time-series models. *Ann Appl Stat* 9(1):247–274.
19. Griffin MR, Zhu Y, Moore MR, Whitney CG, Grijalva CG (2013) U.S. hospitalizations for pneumonia after a decade of pneumococcal vaccination. *N Engl J Med* 369(2):155–163.
20. Simonsen L, et al. (2011) Impact of pneumococcal conjugate vaccination of infants on pneumonia and influenza hospitalization and mortality in all age groups in the United States. *MBio* 2(1):e00309–e00310.
21. Miller E, Andrews NJ, Waight PA, Slack MPE, George RC (2011) Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: An observational cohort study. *Lancet Infect Dis* 11(10):760–768.
22. Flasche S, Slack M, Miller E (2011) Long term trends introduce a potential bias when evaluating the impact of the pneumococcal conjugate vaccination programme in England and Wales. *Euro Surveill* 16(20):19868.
23. Diaz J, et al. (2016) Effectiveness of the 10-valent pneumococcal conjugate vaccine (PCV-10) in children in Chile: A nested case-control study using nationwide pneumonia morbidity and mortality surveillance data. *PLoS One* 11(4):e0153141.
24. Sgambatti S, et al. (2016) Early impact of 10-valent pneumococcal conjugate vaccine in childhood pneumonia hospitalizations using primary data from an active population-based surveillance. *Vaccine* 34(5):663–670.
25. Afonso ET, et al. (2013) Effect of 10-valent pneumococcal vaccine on pneumonia among children, Brazil. *Emerg Infect Dis* 19(4):589–597.
26. Madhi SA, Klugman KP (2007) World Health Organisation definition of "radiologically-confirmed pneumonia" may under-estimate the true public health value of conjugate pneumococcal vaccines. *Vaccine* 25(13):2413–2419.
27. Pilishvili T, et al.; Active Bacterial Core Surveillance/Emerging Infections Program Network (2010) Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J Infect Dis* 201(1):32–41.
28. Nelson JC, et al. (2008) Impact of the introduction of pneumococcal conjugate vaccine on rates of community acquired pneumonia in children and adults. *Vaccine* 26(38):4947–4954.
29. Weinberger DM, Klugman KP, Steiner CA, Simonsen L, Viboud C (2015) Association between respiratory syncytial virus activity and pneumococcal disease in infants: A time series analysis of US hospitalization data. *PLoS Med* 12(1):e1001776.
30. Scott SL (2015) bsts: Bayesian Structural Time Series. R package version 0.6.2. Available at https://cran.r-project.org/web/packages/bsts/. Accessed January 20, 2017.
31. Harvey AC (1990) *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge Univ Press, Cambridge, UK).
32. Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36(3):1–48.

APPLIED BIOLOGICAL SCIENCES

ECONOMIC SCIENCES