Characterizing Vaccine-associated Risks Using Cubic Smoothing Splines

M. Alan Brookhart*, Alexander M. Walker, Yun Lu, Laura Polakowski, Jie Li, Corrie Paeglow, Tosmai Puenpatom, Hector Izurieta, and Gregory W. Daniel

* Correspondence to Dr. M. Alan Brookhart, Department of Epidemiology, UNC Gillings School of Global Public Health, UNC Chapel Hill, McGavran-Greenberg, CB #7435, Chapel Hill, NC 27599-7435 (e-mail: abrookhart@unc.edu).

Initially submitted November 1, 2011; accepted for publication February 28, 2012.

Estimating risks associated with the use of childhood vaccines is challenging. The authors propose a new approach for studying short-term vaccine-related risks. The method uses a cubic smoothing spline to flexibly estimate the daily risk of an event after vaccination. The predicted incidence rates from the spline regression are then compared with the expected rates under a log-linear trend that excludes the days surrounding vaccination. The 2 models are then used to estimate the excess cumulative incidence attributable to the vaccination during the 42-day period after vaccination. Confidence intervals are obtained using a model-based bootstrap procedure. The method is applied to a study of known effects (positive controls) and expected noneffects (negative controls) of the measles, mumps, and rubella and measles, mumps, rubella, and varicella vaccines among children who are 1 year of age. The splines revealed well-resolved spikes in fever, rash, and adenopathy diagnoses, with the maximum incidence occurring between 9 and 11 days after vaccination. For the negative control outcomes, the spline model yielded a predicted incidence more consistent with the modeled day-specific risks, although there was evidence of increased risk of diagnoses of congenital malformations after vaccination, possibly because of a "provider visit effect." The proposed approach may be useful for vaccine safety surveillance.

nonparametric regression; pharmacoepidemiology; self-controlled design; vaccine safety

Abbreviations: ECI, excess cumulative incidence; ITP, idiopathic thrombocytopenic purpura; MMR, measles, mumps, and rubella; MMRV, measles, mumps, rubella, and varicella.

For universally recommended vaccines such as the measles, mumps, and rubella (MMR) vaccine, the absence of a ready comparator cohort complicates the estimation of risks for adverse effects. Almost all children receive the vaccine, and those who do not either possess a contraindication for vaccination or experience substandard medical care. Therefore, any approach that compares incidence rates between cohorts of vaccinated and unvaccinated children will be imprecise because of the small comparison group and may be subject to intractable confounding bias. Furthermore, if administrative claims data are used for exposure identification, absence of a claim for vaccination may not be an adequate method for identifying lack of exposure.

Self-controlled designs represent an alternative to traditional cohort approaches (1). These methods use individuals as their own controls by identifying risk and control periods within each individual. For example, the case-crossover design is a case-only design that compares the frequency of exposure in a period immediately before the event (the hazard period) with some chosen comparator period that may be either before or after the event (2–4). The self-controlled case-series design compares the risks of events in a hazard period usually immediately after the vaccination to a chosen control period that may be either before the vaccination or sometime after the hazard period (5, 6). The optimal size and placement of the risk and control windows depends on an induction period and other patterns of health care use that may be incompletely understood.

In the present article, we propose an approach for studying risks associated with childhood vaccinations that is sensitive to time variation in risk. The method yields an estimate of an easy-to-interpret measure of excess vaccine-induced illness and also provides graphical output that
allows the assessment of various assumptions made by the design. The approach first uses a cubic smoothing spline to flexibly model the daily risk of events after vaccination. To estimate the counterfactual incidence rates that would have occurred had the children not been vaccinated, we fit a log-linear trend to the incidence data excluding the days surrounding the vaccination. We then used the difference between 2 models to estimate the excess cumulative incidence attributable to the vaccination during the 42-day period after vaccination. Inference was obtained using a model-based bootstrap procedure. We applied the method in a study of known effects (positive controls) and expected noneffects (negative controls) of the MMR and measles, in a study of known effects (positive controls) and expected

period after vaccination. Inference was obtained using a
dence attributable to the vaccination during the 42-day

linear trend to the incidence data excluding the days sur-

rounding the vaccination. We then used the difference between 2 models to estimate the excess cumulative incidence attributable to the vaccination during the 42-day period after vaccination. Inference was obtained using a model-based bootstrap procedure. We applied the method in a study of known effects (positive controls) and expected noneffects (negative controls) of the MMR and measles, mumps, rubella, and varicella (MMRV) vaccines.

MATERIALS AND METHODS

Data
We used data from the HealthCore Integrated Research Database (HealthCore, Inc., Wilmington, Delaware). This database contains health care claims information for hospitalizations and provider visits with their associated procedures and diagnoses, plus pharmacy claims from 14 Blue Cross/Blue Shield-licensed health plans. Within this population, we identified records of children with continuous health insurance coverage between 12 and 24 months of age during the study period of October 11, 2004, through January 31, 2010, who had claims for a physician visit in which they received the MMR or MMRV vaccine. We identified claims from physician visits, emergency department visits, and hospitalizations from 56 days before vaccination to 84 days after vaccination, and we noted those for which the accompanying diagnoses included events known to be associated with MMR and MMRV vaccination. These positive controls were fever, adenopathy, rash, and idiopathic thrombocytopenic purpura (ITP). We also identified diagnoses believed to be unrelated to vaccination to serve as negative controls. These were injuries, urinary tract infections, and congenital malformations. The specific definition of each outcome and the codes for identifying ITP onset date are provided in the Appendix Table 1. Additionally, we identified dates of any other vaccinations occurring during the study period. These dates were used as censoring times in the analysis.

Statistical analysis
We indexed time relative to the vaccination, with time 0 being the date the vaccination was received. The data used in the analysis began for all patients at day −56 (before vaccination) and ended on day 84 (after vaccination). Records were censored after the child experienced the event under study or received another type of vaccination. We modeled the risk of each event during each day after vaccination using a Poisson regression model in which the association between time since vaccination and outcomes risk was flexibly modeled with a cubic smoothing spline

estimated using the gam function in R for MacOS X, version 2.10.1 (7). Cubic smoothing splines are defined as the function f(·) that minimizes the following penalized least-squares objective function:

where a ≤ x1 ≤ ... ≤ xn ≤ b and f(·) must have continuous first and second derivatives (8). The solution to this problem is a spline that has knots at the unique values of xi. The penalty term penalizes the fit for nonlinearity and reduces the effective dimensionality of the spline and makes it identifiable (estimable). We set the smoothing parameter for the spline term to be 0.5. This provided adequate flexibility in the fitted spline to capture peaks in incidence that occurred over a period of 3–4 days. Because each observation is a single day (relative to vaccination) and the outcome is the count of events occurring on that day among the number of children at risk, the Poisson offset was set to be the log number of children at risk each day. To estimate the event rate in patients had they not been vaccinated (the counterfactual event rate), we fit a Poisson regression model in which the effect of time was modeled through a log-linear term over the interval −56 to 84 days, omitting days −7 through 42. The log-linear term was used to represent age-related changes in event risk. The omission of the time from day −7 through day 42 was intended to prevent the fit of the linear model from being influenced by either the “healthy vaccinee effect” (decreased risk of the event immediately before vaccination) or the effect of the vaccine itself. To capture weekly periodicity in the use of outpatient health care, we also included a categorical term for time, modulo 7, that represented the remainder of time divided by 7. Finally, we note that because each child could only experience any particular event once in our analysis, there was no need to adjust for repeated observations in either the spline model or the log-linear trend model.

We estimated the excess cumulative incidence by summing across the 42-day postvaccination period the difference in the fitted daily risk of events between the spline and the log-linear trend model. To describe this more precisely, we let s_i be the estimated probability of experiencing an event on day i under the spline model and p_i be the estimated probability of experiencing an event on day i under the log-linear trend model. We estimated the excess cumulative incidence (ECI) per 1,000 children vaccinated as

where \( \sum_{i=1}^{42} n_{x,i} s_i \) is the expected number events under the spline model and \( \sum_{i=1}^{42} \hat{n}_{i} \hat{p}_i \) is the expected number events under the log-linear trend model per 1,000 vaccinated children. At day 1, we set \( \hat{n}_{x,1} = 1,000 \), and subsequently \( n_{x,i} \)
was defined recursively as $\hat{n}_{t,i} = \hat{n}_{t,i-1}(1 - \hat{s}_{t,i-1})$. Similarly, we set $\hat{n}_{p,1} = 1,000$, and subsequently $n_{p,i}$ was defined recursively as $\hat{n}_{p,i} = \hat{n}_{p,i-1}(1 - \hat{p}_{i-1})$. This is a convenient approximation of the difference between the product-limit estimates of the fitted cumulative risks in vaccinated and unvaccinated children over the 42 days after vaccination.

To obtain confidence intervals (and tests) for the estimated ECI, we used a model-based bootstrap procedure implemented as follows: Let $Y_{i}$ be the number of events, $n_{i}$ be the number of patients at risk, and $c_{i}$ be the number of patients censored on day $i$, $i = -56, \ldots, 84$. We repeatedly regenerared the data for each bootstrap sample, taking the observed relative frequencies of censoring and events on each day as true probabilities of occurrence. For the $b$th bootstrap sample, we simulated $Y_{1,b}$ (the number of events on the first day) as a binomial random variable with size $n$ (the total number of children in the study) and probability $c_{1}/n$. We then simulated $c_{1,b}$ (the number of children censored on the first day) as a binomial random variable with size $n$ and probability $c_{1}/n$. We then computed the number of patients at risk on the second day in the bootstrap sample, $n_{2,b} = n - c_{1,b} - Y_{1,b}$, and repeated the above process for day 2. We continued this process until we had simulated event counts for each day. Within each bootstrap sample, we reestimated the log-linear trend (again omitting days 7 through 42) and the spline model (using the postvaccination incidence data) and then recomputed the ECI. We conducted 250 iterations for the bootstrap procedure and computed the 95% confidence intervals for the ECI by taking the 2.5th percentile and 97.5th percentile of the bootstrap distribution. For descriptive purposes, we noted the day on which the fitted cubic spline function reached its maximum. To obtain an approximate interval during which the excess was greatest, we identified the first and last days on which the function exceeded 50% of the peak risk. We used the bootstrap samples to estimate the 95% confidence intervals for peak and the start and end dates for the excess risk interval. By predefining a threshold, this method enables an automated system to determine the excess risk intervals. The Quorum Institutional Review Board approved this research.

**RESULTS**

We identified 375,094 children who received an MMR or MMRV vaccine during the study period. We excluded 151,989 of those who had a gap of more than 2 days in their health insurance coverage between 12 and 24 months of age. This left us with a cohort of 223,105 children. Of these, 53,667 received other vaccinations during follow up and were therefore censored. In Figures 1–7, we plot the daily incidence rates (asterisks), fitted log-linear trends (dashed lines), and fitted splines (solid lines) for fever, rash, adenopathy, ITP, injuries, urinary tract infections, and congenital malformations. For the positive controls, the graphs revealed sharply increased risks of fever, rash, and adenopathy after vaccination. Fever and rash spiked sharply, peaking around 10 days after vaccination. The incidence of adenopathy peaked at the same time but remained elevated through the 20th day after vaccination and returned to baseline by the 40th day after vaccination. For the positive control outcomes, the baseline incidence revealed strong weekly periodicity, a slightly decreasing incidence of rash, and a slightly increasing incidence of fever. The incidence rate of ITP was very low; however, the graphs revealed an increased incidence 11–26 days after vaccination. For the negative control outcomes, the graphs revealed increasing rates of injury and decreasing rates of urinary tract infections and strong weekly periodicity. There was no evidence of elevated rates of injury after vaccination beyond the trend; however, we observed a slight peak in the incidence of urinary tract infection 9 days after vaccination. We did not observe evidence of peak or sharply increased risk in congenital malformation diagnosis during the days after vaccination.

In Table 1, we provide the point estimates and 95% confidence intervals for the ECI statistic for all outcomes, as well as the excess risk interval for persons outcomes in which the lower bound of the ECI was greater than zero. The ECI for all positive control outcomes, including ITP, were significantly elevated in the 42 days after vaccination. For the negative control outcomes, there was no significant elevation in the ECI of either urinary tract infection or injury; however, diagnosis of congenital malformation was increased in the postvaccination period relative to the baseline rate. The excess risk intervals for fever and rash were similar at 6–11 and 9–13 days, respectively. The excess risk interval for adenopathy was broader, running from 8–20 days. The excess risk interval for visits for congenital malformations covered essentially all of the postvaccination risk period of 42 days.

**DISCUSSION**

We have proposed an approach for studying short-term risks associated with exposure to childhood vaccines. The approach is motivated similarly to the method proposed by Serfling to identify excess influenza-related mortality (9). Our approach would be appropriate when a comparison between vaccinated and unvaccinated children may be strongly confounded by unmeasured characteristics. We applied the method in a very large cohort of children in the United States who received either the MMR or the MMRV vaccine during their second year of life. The method revealed sharply increased risks of fever, rash, and adenopathy during the weeks after vaccination, outcomes that are known to be caused by the MMR/MMRV vaccine. The method also revealed increased risk of ITP, a rare adverse event that has been linked with the vaccine. The method correctly ruled out the negative control outcomes of injury and urinary tract infection but did show a temporal association between vaccination and an increased risk of a diagnosis of congenital malformation.

Our plots of disease incidence revealed complex patterns of health care use that need to be addressed in a self-controlled study of vaccine effects. For example, many of the outcomes occurred at a decreased rate in the 7–10 days before vaccination. This could be explained by either a healthy vaccinee effect (i.e., children with active illness having vaccines withheld) or a delay in seeking medical care.
Figure 1. Observed and predicted fever incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.

Figure 2. Observed and predicted rash incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.
Figure 3. Observed and predicted adenopathy incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.

Figure 4. Observed and predicted idiopathic thrombocytopenia purpura incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.
Figure 5. Observed and predicted injury incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.

Figure 6. Observed and predicted urinary tract infection incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.
care because of the pending physician appointment. Many of the outcomes exhibited a spike on the day the vaccination was administered. This does not necessarily reflect a vaccine effect; rather, it could indicate the reporting of other conditions observed during the index office visit. We also observed evidence of time/age trends in disease incidence. Specifically, we observed decreasing incidence of rash and congenital malformations but increasing rates of fever and injury. Finally, all of the outcomes except ITP and injury revealed fairly strong weekly periodicity.

The graphs also revealed time trends in disease incidence. We observed increasing risk of injuries with time since vaccination. This is likely a result of toddlers becoming increasingly more mobile and prone to falls during their first year of life. We also observed a decreasing risk of urinary tract infections. This is consistent with a previous report that the prevalence of urinary tract infections among children under 2 years of age who present with fever decreases with advancing age (10).

Our plots of disease incidence also revealed substantial heterogeneity in the timing of vaccine effects. For example,

![Figure 7. Observed and predicted congenital malformation incidence, 2004–2010. The daily incidence rates are shown with asterisks, the fitted log-linear trend (the counterfactual incidence) is shown with a dashed line, and the fitted spline (the vaccine effect) is shown with a solid line.](image-url)

<table>
<thead>
<tr>
<th>Table 1. Estimated Excess Cumulative Incidence of Each Outcome in the 42 Days After Vaccination per 1,000 Children Vaccinated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Fever</td>
</tr>
<tr>
<td>Rash</td>
</tr>
<tr>
<td>Adenopathy</td>
</tr>
<tr>
<td>Idiopathic thrombocytopenia purpura</td>
</tr>
<tr>
<td>Injury</td>
</tr>
<tr>
<td>Urinary tract infections</td>
</tr>
<tr>
<td>Congenital malformation</td>
</tr>
</tbody>
</table>

fever and rash exhibited well-resolved spikes in incidence occurring approximately 10 days after vaccination. However, the incidence of adenopathy increased 8 days after vaccination but was still at 50% of its peak 20 days after vaccination. Similarly, most ITP occurred between 11 and 26 days after vaccination, but new cases continued to be reported after 40 days. The complex patterns of health-care use, age-related trends in event risk, weekly periodicity, and the heterogeneity in vaccine effects highlight the complexities in prespecifying hazard and control windows for a self-controlled design and also suggest that a single specification may not work well for all outcomes. To estimate the excess cumulative incidence, our approach also requires the analyst to prespecify hazard and healthy vaccinee periods. However, the graphic output that is produced by our analytic approach allows one to identify bias that could result from poor placement or sizing of these intervals. It is also worth noting that the graphs could also be used to identify the appropriate size and placement for the hazard and control intervals for self-controlled designs.

We speculate that the failure of our approach to rule out congenital malformations, a negative control outcome, may be a result of a “provider visit effect.” It is possible that during the index physician visit, potential health problems are noticed and the child is referred to a specialist physician who then makes the diagnosis. This phenomenon, if present, would tend to lead to increased rates of diagnoses made by subspecialty physicians in the weeks after vaccination. As implemented, our method would tend to falsely identify these diagnoses as an effect of the vaccination. One could test for the presence of a provider visit effect by examining whether the incidence of specific outcomes was elevated after a physician visit in which the vaccine was not administered.

Our method possesses other limitations that should be noted. First, the approach may attribute adverse effects to the study vaccine that are caused by another vaccine with which it is frequently coadministered. For example, the MMR and MMRV vaccines may currently be given at the same time as hepatitis A, Haemophilus influenzae, or Streptococcus pneumoniae vaccinations; thus, adverse effects of these vaccines could be falsely attributed to the MMR and MMRV vaccines or vice-versa. This kind of confounding could be explored by restricting the analysis to children who received only the MMR or MMRV vaccine on the index date, although coadministration of the individual MMR and varicella vaccines on the same day could make this approach more difficult. The method also assumes that the log-linear term fit to the pre- and postvaccination data is a good model for the counterfactual incidence. For various reasons, this assumption may not hold. For example, the age-related changes that this term is designed to represent may not be modeled well with a log-linear term. Another limitation of the proposed method is its reliance on a smoothing parameter that must be either supplied by the analyst or estimated in a data-adaptive way using cross-validation. At one extreme setting of this parameter, the spline will interpolate the data points; at the other extreme, it will fit a straight line to the data. The data-adaptive setting of this parameter tended to over-smooth the spikes, as much of the non-spike data could have been modeled well with a log-linear term. Therefore, we manually set the parameter by inspection so that the fit would be just flexible enough to detect the sharp spikes in fever and rash. It is worth noting that for large data sets and reasonably frequent outcomes, one could estimate the ECI without modeling the postvaccination data. The empirical postvaccination incidence could simply be taken as the expected incidence.

Our analysis illustrates the challenges of untangling vaccine effects from complex patterns of health-care use. The method that we have proposed is intuitive and easily implemented. In addition to self-controlled designs, it appears to be a reasonable design choice for childhood vaccine safety studies when comparisons between vaccinated and unvaccinated children may be infeasible because of unmeasured differences between the groups or an inadequate number of unvaccinated children.

There is increasing interest in automated methods for medical product safety surveillance. The appropriate role of different methodologies and databases in safety surveillance, from hypothesis generation to hypothesis testing, is an area of active debate and research (11–14). Because our method was sensitive to the known impacts of the MMR and MMRV vaccines and able to rule out 2 of 3 negative control outcomes, it could also be a reasonable method to use alongside self-controlled designs for childhood vaccine safety surveillance. However, any positive associations found by such a system would need to be carefully investigated to rule out possible provider visit effects or related phenomena.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina Chapel Hill, Chapel Hill, North Carolina (M. Alan Brookhart); World Health Information Science Consultants, Newton, Massachusetts (Alexander M. Walker); CBER, US Food and Drug Administration, Rockville, Maryland (Yun Lu, Laura Polakowski, Hector Izurieta); and HealthCore, Wilmington, Delaware (Jie Li, Corrie Paeglow, Tosmai Puenpatom, Gregory W. Daniel).

Conflict of interest: none declared.

REFERENCES


---

**Appendix Table 1.** Codes for Positive and Negative Controls

<table>
<thead>
<tr>
<th>Category</th>
<th><em>International Classification of Diseases, Ninth Revision, Codes</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative controls</td>
<td></td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>599.xx</td>
</tr>
<tr>
<td>Congenital diseases</td>
<td>740.0x, 740.1x, 740.2x, 741.0x, 741.9x, 742.xx, 743.xx, 744.xx, 748.xx, 749.xx, 745.xx, 746.xx, 747.xx, 750.xx, 751.xx, 752.xx, 753.xx, 754.xx, 755.xx, 756.xx, 757.xx, 758.0x, 758.2x, 758.3x, 758.4x, 758.5x, 758.6x, 758.7x, 758.8x, 759.xx, 7591, V13.69</td>
</tr>
<tr>
<td>Positive controls</td>
<td></td>
</tr>
<tr>
<td>Rash</td>
<td>057.9, 782.1, 782.9</td>
</tr>
<tr>
<td>Fever</td>
<td>780.60, 780.63, 780.64</td>
</tr>
<tr>
<td>Adenopathy</td>
<td>683, 785.6, 289.2, 289.3</td>
</tr>
<tr>
<td>ITP diagnosis</td>
<td></td>
</tr>
<tr>
<td>Broad ITP definition</td>
<td>287.3x, 287.4x, 287.5x</td>
</tr>
<tr>
<td>Narrow ITP definition</td>
<td>287.31</td>
</tr>
</tbody>
</table>

Abbreviation: ITP, idiopathic thrombocytopenic purpura.