2006

Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations

Jason M. Glanz

David L. McClure

Stanley Xu

Simon J. Hambidge

Martin Lee

See next page for additional authors

Recommended Citation

Glanz, Jason M.; McClure, David L.; Xu, Stanley; Hambidge, Simon J.; Lee, Martin; Kolczak, Margarette S.; Kleinman, Ken; Mullooly, John P.; and France, Eric K., "Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations" (2006). *Journal of Clinical Epidemiology*. 25.
10.1016/j.jclinepi.2005.11.012

Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations

Jason M. Glanz, David L. McClure, Stanley Xu, Simon J. Hambidge, Martin Lee, Margarette S. Kolczak, Ken Kleinman, John P. Mullooly, Eric K. France

Abstract

Objective: We conducted a simulation study to empirically compare four study designs [cohort, case–control, risk-interval, self-controlled case series (SCCS)] used to assess vaccine safety.

Study Design and Methods: Using Vaccine Safety Datalink data (a Centers for Disease Control and Prevention-funded project), we simulated 250 case sets of an acute illness within a cohort of vaccinated and unvaccinated children. We constructed the other three study designs from the cohort at three different incident rate ratios (IRRs, 2.00, 3.00, and 4.00), 15 levels of decreasing disease incidence, and two confounding levels (20%, 40%) for both fixed and seasonal confounding. Each of the design-specific study samples was analyzed with a regression model. The design-specific $\hat{\beta}$ estimates were compared.

Results: The $\hat{\beta}$ estimates of the case–control, risk-interval, and SCCS designs were within 5% of the true risk parameters or cohort estimates. However, the case–control's estimates were less precise, less powerful, and biased by fixed confounding. The estimates of SCCS and risk-interval designs were biased by unadjusted seasonal confounding.

Conclusions: All the methods were valid designs, with contrasting strengths and weaknesses. In particular, the SCCS method proved to be an efficient and valid alternative to the cohort method.

Keywords: Simulation study; Cohort; Case–control; Risk-interval; Self-controlled case series (SCCS); Bias (epidemiology); Confounding factors (epidemiology)

1. Introduction

The most widely accepted methods for evaluating vaccine safety have been study designs that compare distinct exposed and unexposed, or diseased and nondiseased populations. These study methods include prospective designs such as the cohort, and retrospective designs such as the case–control. This investigation evaluates these traditional study designs as well as two newer designs in a simulated analysis of a known, rare, and acute vaccine reaction: idiopathic thrombocytopenic purpura (ITP) after measles-mumps-rubella (MMR) vaccination [1,2].

In a cohort study, a group of healthy vaccinated and unvaccinated individuals are followed forward in time, and the incidence of illness in the two groups is compared. This design provides a direct estimate of effect (the incidence rate ratio, IRR), is well suited for rare exposures, and can be used to analyze multiple outcomes [3,4]. It can, however, be difficult and costly to implement when the disease is rare, and because vaccine safety studies typically involve populations with high vaccine coverage rates, there may be few unvaccinated controls available. The design is also susceptible to biases that can be introduced by comparing vaccinated and unvaccinated populations, as these groups may differ by ethnicity, socioeconomic status, and underlying health states [5].

In nested case–control studies, individuals who experienced a particular event over a defined time period are identified. This group of cases is then compared to a control group of event-free individuals from the same time period, who are often matched to the cases on variables such as gender, managed care organization (MCO), and age [1,6–8]. This design is economical and well suited for rare illnesses. In addition, because the cases are typically matched to the controls by age and calendar time (e.g., the age at the date of diagnosis), particular time-varying confounders, such as age and seasonality, are adjusted for by proxy. As with the cohort method, however, confounding variables related to both the outcome and vaccination status—as well as other time-varying factors such as underlying health states—will bias the case-control design.

Since 1995, alternative methods known as the risk-interval (or vaccinated cohort) and self-controlled case series (SCCS) study designs have been used for vaccine safety studies [2,7,9–15]. These designs differ from more traditional methods in that time intervals both before and after vaccination *within the same individual* are used to classify a person as exposed or unexposed. In the risk-interval design, incidence rates for risk and nonrisk time periods are compared, but only vaccinated individuals are included in the study. A time period immediately following vaccination is designated as the risk-interval, and events that occur during this period are classified as exposed cases. Time periods outside of the risk-interval—before and after the vaccination—are considered the nonrisk (or control) periods, where occurrences of illness are classified as unexposed cases. Because only vaccinated individuals are included in the study, biases introduced by comparing vaccinated and unvaccinated populations are minimized. In addition, because control time periods both before vaccination and after the risk period are included in the analysis, the design is ideal for assessing the risk of acute, self-limiting events following vaccination.

The SCCS method is a similar design in which incidence rates for risk and nonrisk time periods are compared, but only cases are necessary for the analysis [14–17]. The study population comprises only cases that occur over a predefined observation period, and each case acts as its own control, thereby controlling for both measured and unmeasured confounding variables that do not vary over time. With the SCCS method, multiple occurrences of independent events within an individual can be analyzed. Theoretical calculations have also demonstrated that the method's statistical power closely approximates that of a cohort study when the vaccination coverage rate is high and the periods of risk following vaccination are short [14,15]. To our knowledge, however, these assertions have not been validated empirically.

Possible limitations of the risk-interval and SCCS methods stem from their ability to implicitly control for time-varying confounders, such as seasonality or age. In contrast to the case–control analysis, these covariates cannot be adjusted for by proxy in the risk-interval and SCCS analyses. Instead, time-varying confounders must be explicitly defined as either continuous functions or categorical variables and added to multiple Poisson regression models [12,14]. Mis-specifying such variables can lead to biased results—particularly when the event is rare [18].

To address some of the gaps in the current literature, we conducted a simulation study that evaluated the bias and precision of the four study designs' IRR estimates, the stability of the design-specific IRR estimates at different levels of disease incidence, and each design's ability to handle unmeasured confounding.

2. Materials and methods

2.1. Data

This study was conducted under the Vaccine Safety Datalink (VSD), a Centers for Disease Control and Prevention-funded project that links large administrative databases from eight MCOs located across the United States. The focus of the VSD is to conduct epidemiologic studies of vaccine safety [19]. Currently, the VSD databases contain health care data from 1991 to 2003, representing a cohort of over 5,000,000 children younger than 18 years of age. For this study, we used VSD data through year 2000 from five of the MCO sites.

2.2. Cohort construction and simulation

We first constructed a retrospective cohort study population using the following VSD data fields: MCO, birth date, gender, membership dates, and MMR vaccination dates. To ensure a balanced distribution of important variables among the study groups, each MMR vaccinated child was matched to one unvaccinated child by gender, MCO, and age (within 7 days) at the date of the vaccination (n = 2,774,122). Up to 365 days before and after the matched dates were used as follow-up times (i.e., the observation periods). Unvaccinated children did not receive a vaccination during their entire follow-up time of up to 730 days surrounding the matched date. In the exposed children, the 42-day period following vaccination was defined as exposed person-time. The 6-week postvaccination period is an exposure time interval in which ITP has been attributed to the MMR vaccine [1,2]. All of the time outside of the 42-day risk period was designated as unexposed person-time. On average, each cohort member contributed 591 days of person-time follow-up.

After the study population was constructed, cases of ITP were simulated on a specific date (diagnosis date) within the defined follow-up times at a fixed IRR. Exposed cases were simulated in the 42-day risk periods, while unexposed cases were simulated in the time periods outside of the risk periods. In the unexposed (or unvaccinated) subjects, cases were simulated within the entire 365-day periods before or after the matching date. The following probabilistic model was used to simulate the cases:

$$\pi = pt\frac{1}{1+e^{-(\beta_0+\beta_1 x_1)}} \tag{1}$$

where π is the probability of being a case, β_0 is the intercept of the model, β_1 is the main parameter of interest, x_1 is the exposure indicator (1 = exposed and 0 = unexposed), and pt represents person-time contributed. For unexposed individuals (x_1 = 0), π is a function of pt and β_0. To approximate β_0, we used the estimated annual incidence rate of ITP (eight cases per 100,000 children) [20]. In the cohort population, the estimated incidence rate (cases per person-day) among the unexposed was 2.19×10^{-7}; the natural log of this number is β_0. The β_1s were chosen to be 0.693, 1.099, and 1.386 for IRR levels 2.00, 3.00, and 4.00, respectively. We chose these IRR levels because they represent strengths of associations that tend to influence vaccination policy.

The form of equation 1 implies that the probability of being a case is proportional to the amount of person-time contributed. The second term of eq. (1) represents the probability of being a case on any day during the follow-up period. But, because our study population was so large, it was not feasible to simulate cases based on each day's probability. Instead, we used eq. (1) to simulate the cases within each subject's predefined follow-up periods: prevaccination unexposed, postvaccination exposed, and postvaccination unexposed. Given the maximum amount of person-time contributed (i.e., 730 days) and the range of β_0 and β_1, the probability of being a case could never be greater than 1.

The goal of the simulation was to compare the gold standard $\hat{\beta}_1$estimates of the cohort to those of the risk-interval, SCCS, and case–control study designs. Therefore, we first created a cohort gold standard population from the actual VSD data and simulated the cases to create each of the three fixed β_1 levels (IRR = 2.00, 3.00, and 4.00). We then used these cohort populations to build the populations for the three remaining study designs. Each design was nested within the cohort so that direct comparisons could be made. In addition, the $\hat{\beta}_1$ estimates of all four designs were compared to the true, fixed β_1 values.

To determine a suitable number of iterations, we simulated 100, 250, 500, 1,000, 1,500, and 2,000 cohort populations at an IRR of 2.00. As the number of iterations increased, the variances of the $\hat{\beta}_1$ estimates expectedly decreased, but the differences between the mean $\hat{\beta}_1$ estimates demonstrated little variability. The differences between the mean $\hat{\beta}_1$ estimates and the true β_1 ($\beta_1 = 0.693$) remained within 2.5 to 3.0% as the number of iterations increased from 250 to 2,000. Thus, at each of the three IRR levels, we constructed 250 cohort, risk-interval, SCCS and case–control study populations—the latter three of which were nested within the cohort. Each of the 250 simulated study populations was analyzed with an appropriate regression model, and the design-specific $\hat{\beta}_1$ estimates of the nested designs were compared to those of the cohort and to the true β_1 values (Fig. 1).

To evaluate the stability of the different study designs, we continuously reduced the case population by 20% in the cohort design and repeated the simulation. The case populations were decreased by lowering the baseline disease incidence rate, that is, decreasing the value of β_0 in the probabilistic model used to simulate the cases. For each successive case population, the amount of person-time contributed (the denominator of the incidence rate) remained constant. The objective of this study component was to demonstrate—at various IRR levels—how the design-specific regression estimates differ as the incidence of disease decreases.

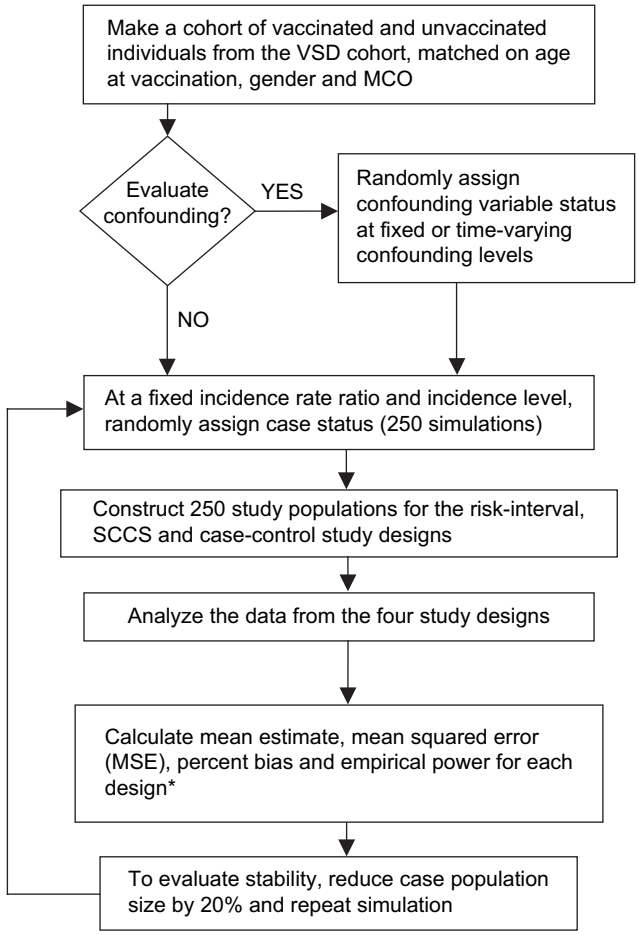

Fig. 1. Summary of steps for the simulation study. *See text for definitions.

2.3. The nested study designs

2.3.1. Case–control

The simulated cases from the cohort were identified and matched to nondiseased controls by age (within ±7 days) at the simulated diagnosis date, MCO, and gender. Cases and all of their matched controls represented risk sets [21]. In our primary analysis, four controls were randomly sampled from each risk set. A 1:4 case-to-control match was sufficient to detect an odds ratio of 1.95 with a case population size of 400 and 80% power (two-sided, $\alpha = 0.05$). We also conducted analyses using case-to-control matches of 1:10, 1:100, and 1:*n* (all available controls). For the 1:*n* design, there were 2 to 1,326 (median, 629) available controls for each case.

A case or matched control was considered exposed if an MMR vaccination date was within 42 days prior to the diagnosis date of the matched case. The sampled risk sets represented matched strata, which were included in conditional logistic regression models to estimate odds ratios (OR) for the risk of the event in the 42-day risk period. The $\hat{\beta}_1$ estimates for the ORs of the case–control design were compared

to the $\hat{\beta}_1$ estimates for the IRRs of the cohort design and to the true β_1 values; the OR approximates the IRR when the incidence of disease is rare and the risk periods are short [14,22].

2.3.2. Risk-interval

The risk-interval analysis was limited to the MMR vaccinated children from the cohort population (n = 1,387,061). The 42-day periods following vaccination represented the exposed risk intervals, and the time outside of this period—up to 365 days before and 42 to 364 days after vaccination—represented the unexposed nonrisk intervals. Because this study population was half the size of the cohort population, only cases that had been simulated during the observation periods (i.e., up to 365 days before and after the vaccination date) of the vaccinated cohort subjects were included. The risk-interval data were analyzed with Poisson regression.

2.3.3. Self-control case series (SCCS)

The simulated cases from the cohort population were included in the SCCS analysis. Once the cases were identified, the follow-up times—up to 365 days before and after the matched cohort vaccination date—were used to ascertain exposure status. The incidence of simulated ITP in the 42-day risk period following vaccination was compared to the incidence in the time periods outside of the risk period. Conditional Poisson regression was used to estimate the IRR of ITP in the 42-day risk window, treating each case as a unique stratum [15].

2.4. Confounding

The final objective was to examine how well each design handled unmeasured confounding, which we simulated in two different forms. The first form was a fixed confounding variable that did not vary over time, and the second was a time-varying confounder that represented a fluctuating risk of illness across the follow-up periods.

2.4.1. Unmeasured/unknown fixed confounding

First, we simulated a hypothetical secondary disease state (the confounder) in the cohort, so that the prevalence of disease was both higher in the vaccinated than in the unvaccinated and associated with the primary outcome of interest (ITP). The overall prevalence of the secondary disease state was 10%, representing the percentage of VSD cohort members with a hypothetical preexisting condition that rendered them at high-risk for developing ITP. This 10% prevalence was disproportionately distributed so that 16% of the vaccinated and 4% of the unvaccinated were classified as high risk. Then, using the form of eq. (1) with vaccination exposure (x_1) and high-risk status (x_2) as dichotomous covariates, we simulated two sets of 250 cohort populations at a vaccine-associated IRR of 2.00 ($\beta_1 = 0.693$) and at two different high risk-associated IRRs of 4.50 ($\beta_2 = 1.50$) and 12.20 ($\beta_2 = 2.50$). At both of the fixed β_2s of 1.50 and 2.50, the mean coefficient estimate for vaccination exposure ($\hat{\beta}_1$) was 0.671 (IRR = 1.95). When the high risk variables were removed from both sets of regression models, the $\hat{\beta}_1$ values increased by 20% ($\hat{\beta}_1 = 0.805$, IRR = 2.24) and 40% ($\hat{\beta}_1 = 0.939$, IRR = 2.56), respectively. Although evaluating confounding in a regression model is somewhat arbitrary, a change of 20% or more is often described as meaningful [23,24].

From each of the 250 simulated cohort populations, the study populations for the other designs were created. The design-specific data were analyzed as if the confounding bias was unmeasured or unknown, and the results were compared to the adjusted regression coefficient estimates of the cohort.

2.4.2. Seasonal time-varying confounding

Next, we simulated 20 and 40% seasonal, time-varying confounding in the cohort. We created a seasonal effect in which the probability of being simulated as a case depended on calendar time as well as vaccination status. A greater percentage of MMR vaccinations were given in the summer months between May and August, mainly due to the vaccination pattern of children ages 4–12 years (school physicals). To create the seasonal effect, we simulated an increased risk for developing ITP between December through March, even though a seasonal pattern of ITP is not known to exist. As with the fixed confounding, we used eq. (1) with vaccination exposure (x_1) as the main effect and added a covariate (x_2) for the time-varying, seasonal factor. Specifically, we simulated 250 cohort populations at two different seasonal, time-varying effect (β_2) levels of 1.30 (IRR = 3.67) and 2.90 (IRR = 18.11). These relatively large effect levels were necessary because of the mild seasonal distribution of MMR administration: 42.5% were given in the four summer months vs. 26.1% during the 4-month, high-risk, winter period. At the two seasonal effect levels, the mean exposure coefficient estimates ($\hat{\beta}_1$) were 0.679 (IRR = 1.97) and 0.702 (IRR = 2.02), respectively. When the seasonal covariates were removed from the regression models, the mean $\hat{\beta}_1$s decreased to 0.538 and 0.417, representing changes of approximately 20 and 40%, respectively. From these simulated cohorts with the infused time-varying, seasonal confounding, we created the other design-specific study populations and observed how the nested designs handled the confounding bias. We examined the bias both as if season was unmeasured and with the seasonal covariates adjusted in the regression analyses.

2.5. Outcome measures

After the design-specific analyses were completed, each of the 250 design-specific regression estimates at each IRR level was compared to the corresponding regression estimate from the simulated cohort design and to the true β_1 values. The differences between regression estimates were expressed as percent biases.

2.5.1. Bias

Bias is presented in two forms:

1. The percent bias from the truth measures the percent difference between the true parameter (β_1) and the regression estimates ($\hat{\beta}_1$) of the four study designs [eq. (2)].

$$\text{Percent Bias from True} = \frac{\text{nested design estimate} - \text{true parameter}}{\text{true parameter}} \times 100 \tag{2}$$

2. The percent bias from the cohort is defined as the percent difference between the cohort regression estimates and the corresponding regression estimates of the risk-interval, SCCS and case–control designs [eq. (3)].

$$\text{Percent Bias from Cohort} = \frac{\text{nested design estimate} - \text{cohort gold standard estimate}}{\text{cohort gold standard estimate}} \times 100 \tag{3}$$

Percent biases are presented at each IRR level using the largest case incidence level (average case population size across the three IRR levels = 392; estimated incidence rate = 8 cases per 100,000 person-years).

2.5.2. Mean squared error

The mean squared error (MSE) measures the average squared difference between the parameter estimate and its true value [eq. (4)]. It contains two components: the variance of the 250 parameter estimates (precision) and the square of the average bias (accuracy).

$$MSE = \frac{\sum_{i=1}^{250}\left(\hat{\beta}_i - \beta\right)^2}{250} = Var\left(\hat{\beta}\right) + \left(\frac{\sum_{i=1}^{250}\hat{\beta}_i - \beta}{250}\right)^2 \tag{4}$$

2.5.3. Empirical power

Empirical power demonstrates, at each incidence level, what percentage of the 250 design-specific IRR estimates was positive and significant ($P < .05$). It is the proportion of the statistical models that had enough cases to generate IRR estimates that rejected the null hypothesis. As the case population decreases, it is less likely the null hypothesis will be rejected.

3. Results

3.1. Bias and precision

The mean estimates, mean percent biases, and MSEs for each design are displayed in Table 1. Across the three IRR

Table 1
Mean estimates of $\hat{\beta}_1$ (mean squared error) and mean percent biases[a,b] (standard errors) with respect to the cohort and true risk parameter, by design, by incidence rate ratio (IRR) level, based on 250 simulations[c]

	True value of $\beta_1 = 0.693$ (IRR = 2.00)			True value of $\beta_1 = 1.099$ (IRR = 3.00)			True value of $\beta_1 = 1.386$ (IRR = 4.00)		
Study designs	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)
Cohort	0.671 (0.044)	—	−3.2 (1.9)	1.067 (0.029)	—	−2.9 (1.0)	1.376 (0.022)	—	−0.8 (0.7)
Risk interval	0.674 (0.048)	−0.6 (1.4)	−2.7 (2.0)	1.066 (0.030)	0.0 (0.3)	−3.0 (1.0)	1.380 (0.023)	0.4 (0.3)	−0.4 (0.7)
SCCS[d]	0.661 (0.048)	−2.7 (1.8)	−4.7 (2.0)	1.055 (0.032)	−1.1 (0.4)	−4.0 (1.0)	1.363 (0.024)	−0.9 (0.3)	−1.7 (0.7)
Case control[e]									
1:4	0.657 (0.075)	−2.7 (1.9)	−5.2 (2.5)	1.074 (0.056)	0.6 (1.0)	−2.3 (1.4)	1.369 (0.043)	−0.4 (0.7)	−1.2 (0.9)
1:10	0.664 (0.056)	−0.8 (2.1)	−4.2 (2.2)	1.068 (0.039)	0.3 (0.7)	−2.8 (1.1)	1.372 (0.034)	−0.3 (0.5)	−1.0 (0.8)
1:100	0.661 (0.049)	−2.6 (1.7)	−4.6 (2.0)	1.063 (0.033)	−0.3 (0.5)	−3.3 (1.0)	1.368 (0.029)	−0.6 (0.4)	−1.3 (0.8)
1:*n*	0.662 (0.049)	−2.6 (1.6)	−4.5 (2.0)	1.063 (0.033)	−0.2 (0.5)	−3.2 (1.0)	1.369 (0.028)	−0.5 (0.4)	−1.3 (0.8)

[a] Percent bias cohort $= \frac{\text{nested design estimate} - \text{cohort gold standard estimate}}{\text{cohort gold standard estimate}} \times 100$.

[b] Percent bias true $= \frac{\text{nested design estimate} - \text{true parameter}}{\text{true parameter}} \times 100$.

[c] The following probabilistic model was used to simulate the cases: π = person time $\frac{1}{1+e^{-(\beta_0+\beta_1 x_1)}}$, where $\beta_0 = -15.3$ and $\beta_1 = 0.693$, 1.099, or 1.386.

[d] Self-controlled case series.

[e] Matched case-to-control ratios 1:4, 1:10, 1:100, 1:*n* (all available).

levels, the cohort design produced mean biases of −3.2 to −0.8% when compared to the true value of β_1. The mean percent biases with respect to the cohort estimates or to the true parameters ranged from −5.2 to 0.6% across the four study methods and three IRR levels. These measures for the case–control design remained within the same range when the case-to-control ratio increased to 1:10, 1:100, and 1:*n*.

The MSEs of the estimates for the risk-interval and SCCS were within 11.0% of the cohort's MSEs across the three IRR levels, while the MSEs of the case–control (1:4) were from 70.5 to 95.5% larger than those of the cohort (Table 1). The case–control's MSEs for case-to-control ratios of 1:10, 1:100, and 1:*n* were between 11.4 to 54.5% larger than those of the cohort. When compared to the cohort, the percentage differences for ratios 1:100 and 1:*n* were approximately equivalent, suggesting no additional efficiency was gained by using more than 100 controls.

Across the three IRR levels, the mean coefficient estimates of the four designs remained within 5.0% of each other as the incidence rate decreased (Fig. 2). However, below an incidence rate of 0.66 cases per 100,000 person-years (33 cases), all of the designs produced mean coefficient estimates that were approximately 5.0 to 50.0% greater than the true effect. At the low incidence levels, fewer of the regression models were able to generate stable, bounded $\hat{\beta}_1$ estimates due to zero cells, that is, either no simulated exposed or unexposed cases. When the models did converge at the low incidence levels, the resulting regression estimates tended to overestimate the true effect, with wide confidence intervals. This occurred because a regression model needed at least one exposed case to generate a bounded estimate. Often, at the low incidence levels, the incidence rate in the exposed with only one case was disproportionately larger than the unexposed incidence rate, leading to a mean $\hat{\beta}$ estimate that overestimated the true effect.

Overall, at each IRR level, the MSEs of the regression estimates increased as the incidence rate decreased, but the differences between the MSEs of the designs varied (Fig. 3). As the case population declined, the MSEs of the case–control were consistently higher than those of the other designs, and the magnitude of the difference increased as the incidence rate decreased. This pattern was observed with all of the case-to-control ratios.

3.2. *Empirical power*

Across all of the IRR levels, the empirical power of the risk-interval and SCCS designs were within 15.4% of the cohort values (Fig. 4). With decreasing case population sizes, the empirical power of the case-control design (1:4) was up to 75.0% lower than that of the other designs. As the case-to-control ratio increased from 1:4 to 1:*n*, the case–control's power approached that of the cohort design,

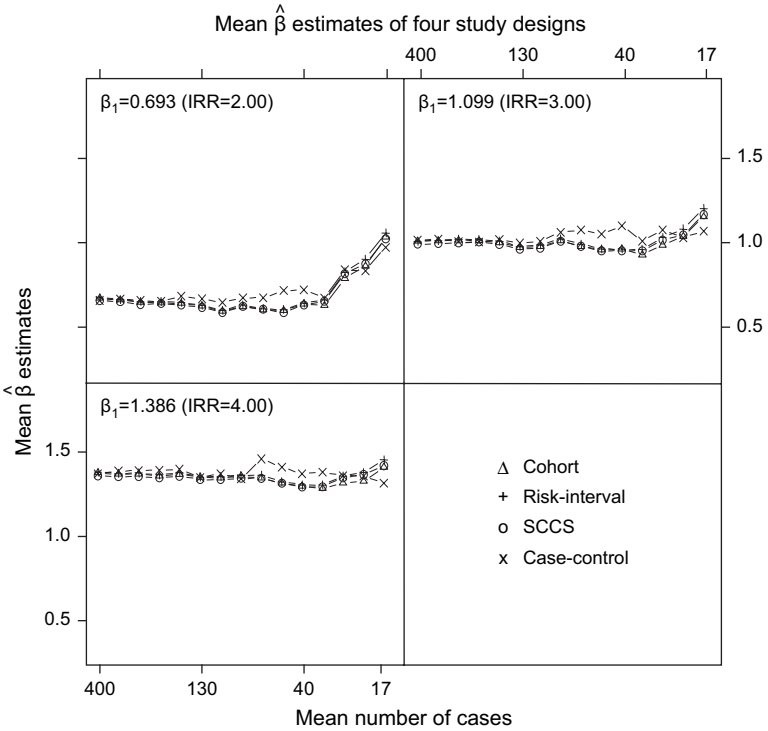

Fig. 2. The mean of each design's 250 regression estimates, by mean case population size, by incidence rate ratio (IRR) levels 2.00, 3.00, and 4.00. The mean case numbers represent the mean case population size of the 250 simulations. The mean case numbers 400, 130, 40, and 17 represent incidence rates of 8.00, 2.70, 0.80, and 0.35 cases per 100,000 person-years.

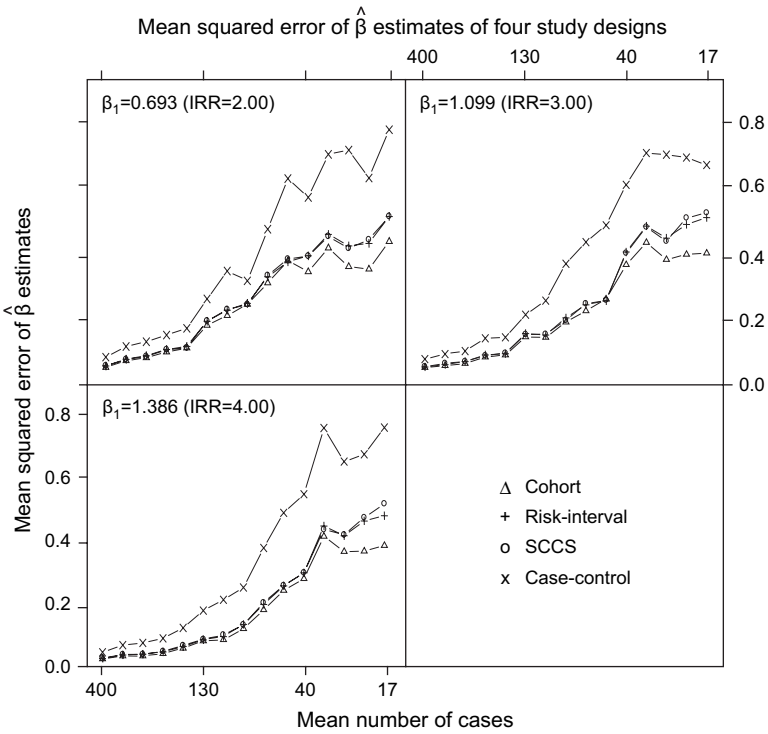

Fig. 3. The mean squared error of each design's 250 regression estimates, by mean case population size, by incidence rate ratio (IRR) levels 2.00, 3.00, and 4.00. The mean case numbers represent the mean case population size of the 250 simulations. The mean case numbers 400, 130, 40, and 17 represent incidence rates of 8.00, 2.70, 0.80, and 0.35 cases per 100,000 person-years.

and it remained within 11.9% of the cohort as the case population decreased (data not shown).

3.3. Confounding

3.3.1. Unmeasured/unknown fixed confounding

In the presence of 20 or 40% confounding, the mean percent biases of the risk-interval and SCCS designs ranged from −4.8% to 0.5% when compared to the adjusted cohort estimates or to the true parameters (Table 2). In contrast, when compared to the adjusted cohort or to the truth, the mean percent biases of the case–control for all case-to-control ratios were approximately 20 and 40% for the two respective levels of confounding.

3.3.2. Seasonal time-varying confounding

At 20 or 40% confounding, the mean percent biases of the case–control estimates for all case-to-control ratios were within 4.9% of the adjusted cohort estimates or the true parameters (Table 3). The estimates of the unadjusted risk-interval and SCCS designs, on the other hand, were biased by approximately −20 and −40% when compared to the adjusted cohort estimates or to the truth. Conversely, the mean percent biases of the adjusted risk-interval and SCCS designs were within 1.8% of the adjusted cohort estimates or the true parameters.

4. Discussion

In this study using vaccine safety databases and simulated cases of a rare, acute illness (ITP) after MMR vaccination, the risk-interval, SCCS, and case–control study designs produced valid IRR estimates that were within 3% of a cohort gold standard. The case–control design was associated with the highest MSEs, the lowest empirical power, and the highest mean percent bias in the presence of unmeasured, fixed confounding. Its estimates, nonetheless, were not biased when the confounding was simulated as a seasonal effect because the cases were matched to the controls by age and diagnosis date. The SCCS and risk-interval designs, in contrast, proved to be as powerful as the cohort (corroborating previously published theoretical results) [14,15], demonstrated the ability to control for unmeasured fixed confounding, and produced mean percent biases that were considerably higher than those of the case–control when the effect of seasonality was not adjusted in the analysis.

While the case–control design proved to be more variable and less powerful than the other designs, the degree of difference decreased as the number of controls for each case increased. At 100 or more controls for each case, the MSE and power of the case–control design approached those of the cohort. When conducting a case–control study (particularly if the analysis is limited to electronic data only), more controls than the customary 4:1 ratio are chosen

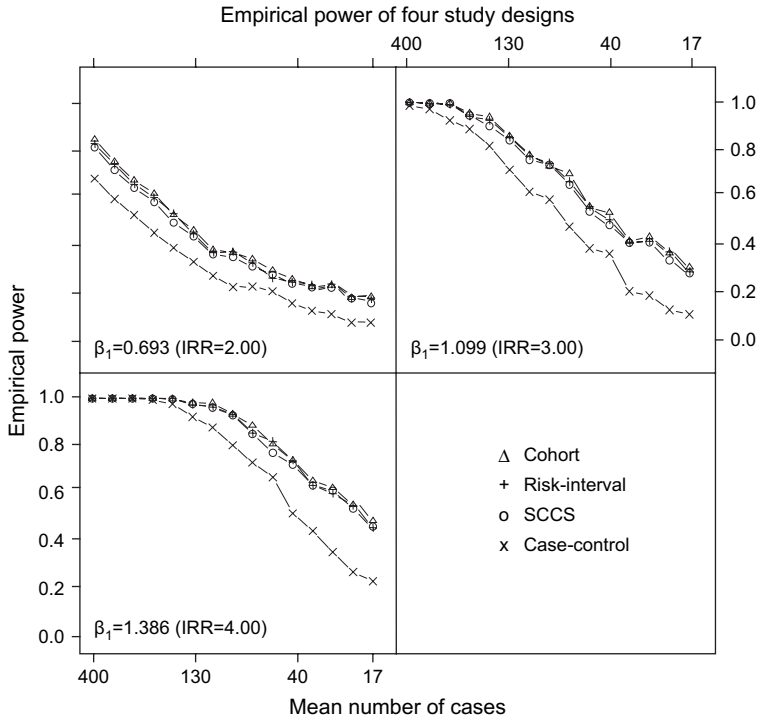

Fig. 4. The proportion of the 250 regression models with positive estimates (>0) and $P < .05$, by study design, by mean case population size, by IRR levels 2.00, 3.00, and 4.00. The mean case numbers 400, 130, 40, and 17 represent incidence rates of 8.00, 2.70, 0.80, and 0.35 cases per 100,000 person-years.

to increase power and to detect effect modification [25,26]. However, when additional data is collected through medical chart review, the case–control design is no longer economical at these higher case-to-control ratios.

We demonstrated empirically that the SCCS and risk-interval designs are biased if a seasonal effect is overlooked. However, we also showed that seasonality can be adjusted in a self-controlled design if a seasonal effect is suspected *a priori* [12,18]. The challenge of conducting such an analysis is that the form of seasonal variable must be explicitly defined prior to conducting the analysis. This can be particularly difficult if the event is rare, as there may not be enough information to estimate the seasonal effect. The nested case–control design, on the other hand, provides a seasonal adjustment by proxy, since the cases are matched to the controls by age and calendar time. We believe that

Table 2
Mean estimates of $\hat{\beta}_1$ (mean squared error) and mean percent biases[a,b] (standard errors) with 20 and 40% unmeasured/unknown fixed confounding, by design, based on 250 simulations and a true β_1 value = 0.693 (IRR = 2.00)

	20% Fixed confounding			40% Fixed confounding		
Study designs	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)
Cohort[c]	0.671 (0.030)	—	−3.1 (1.6)	0.671 (0.015)	—	−3.2 (1.1)
Risk interval[d]	0.674 (0.032)	0.3 (0.4)	−2.8 (1.6)	0.673 (0.015)	0.5 (0.2)	−2.7 (1.1)
SCCS[d,e]	0.660 (0.032)	−1.9 (0.5)	−4.8 (1.6)	0.661 (0.016)	−1.5 (0.3)	−4.7 (1.1)
Case–control[d,f]						
1:4	0.838 (0.076)	26.1 (1.8)	20.9 (2.1)	0.961 (0.101)	45.0 (1.4)	38.7 (1.6)
1:10	0.825 (0.059)	24.5 (1.2)	19.0 (1.9)	0.965 (0.094)	45.3 (0.9)	39.2 (1.3)
1:100	0.822 (0.052)	24.3 (0.9)	18.7 (1.7)	0.965 (0.091)	45.4 (0.8)	39.2 (1.2)
1:*n*	0.823 (0.052)	24.3 (0.8)	18.7 (1.7)	0.965 (0.091)	45.4 (0.7)	39.3 (1.2)

[a] Percent bias cohort $= \dfrac{\text{nested design estimate} - \text{cohort gold standard estimate}}{\text{cohort gold standard adjusted estimate}} \times 100.$

[b] Percent bias true $= \dfrac{\text{nested design estimate} - \text{true parameter}}{\text{true parameter}} \times 100.$

[c] Controlling for x_2, the fixed confounding factor.

[d] Analyzed as if the confounding was unmeasured or unknown.

[e] Self-controlled case series.

[f] Matched case-to-control ratios 1:4, 1:10, 1:100, 1:*n* (all available).

Table 3
Mean estimates of $\hat{\beta}_1$ (mean squared error) and mean percent biases[a,b] (standard errors) with 20 and 40% seasonal time-varying confounding, by design, based on 250 simulations and a true β_1 value = 0.693 (IRR = 2.00)

	20% Seasonal time-varying confounding			40% Seasonal time-varying confounding		
Study Designs	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)	$\hat{\beta}_1$ (MSE)	% Bias cohort (SE)	% Bias true (SE)
Cohort						
Adjusted[c]	0.672 (0.025)	—	−2.1 (1.4)	0.696 (0.007)	—	0.5 (0.8)
Unadjusted[d]	0.540 (0.048)	−20.8 (1.1)	−22.1 (1.4)	0.423 (0.080)	−39.8 (0.3)	−38.9 (0.8)
Risk interval						
Adjusted[c]	0.681 (0.028)	0.2 (0.4)	1.8 (1.5)	0.698 (0.008)	0.2 (0.2)	0.7 (0.8)
Unadjusted[d]	0.534 (0.053)	−23.1 (0.6)	−22.9 (1.5)	0.417 (0.084)	−40.8 (0.4)	−39.8 (0.8)
SCCS[e]						
Adjusted[c]	0.681 (0.029)	0.2 (0.6)	1.8 (1.6)	0.696 (0.010)	−0.1 (0.4)	0.5 (0.9)
Unadjusted[d]	0.537 (0.052)	−22.5 (0.7)	−22.5 (1.5)	0.419 (0.083)	−40.4 (0.4)	−39.5 (0.8)
Case–control[d,f]						
1:4	0.706 (0.043)	4.9 (1.6)	1.9 (1.9)	0.694 (0.013)	−0.3 (0.7)	0.1 (1.1)
1:10	0.684 (0.034)	1.0 (1.1)	−1.3 (1.7)	0.694 (0.011)	−0.3 (0.5)	0.2 (0.9)
1:100	0.678 (0.030)	−0.1 (0.8)	−2.1 (1.6)	0.695 (0.009)	−0.3 (0.4)	0.2 (0.8)
1:*n*	0.679 (0.029)	−0.1 (0.7)	−2.1 (1.6)	0.695 (0.008)	−0.2 (0.3)	0.2 (0.8)

[a] Percent bias cohort $= \dfrac{\text{nested design estimate} - \text{cohort gold standard estimate}}{\text{cohort gold standard estimate}} \times 100.$

[b] Percent bias true $= \dfrac{\text{nested design estimate} - \text{true parameter}}{\text{true parameter}} \times 100.$

[c] Adjusted for the seasonal covariate.

[d] Unadjusted without the seasonal covariate in the regression model.

[e] Self-controlled case series.

[f] Matched case-to-control ratios 1:4, 1:10, 1:100, 1:*n* (all available).

epidemiologists should be aware of this distinction when designing their studies of vaccine safety.

It should also be noted that our results do not suggest that the cohort is the best method for conducting studies of vaccine safety. We used the cohort as the gold standard because it represents the traditional standard in epidemiologic studies. All other study designs, including the risk-interval, SCCS, and case–control designs, are derived from a cohort study design [27]. For a study of vaccine safety, the cohort would be an ideal design if the investigator had access to a large study population of a million or more subjects and if all potential confounding factors could be precisely measured. These caveats are important because, in most instances, potential vaccine-associated adverse events are rare, few unvaccinated controls are available, and vaccinated and unvaccinated populations may differ by confounding variables that are absent from medical databases [5]. When these critical requirements cannot be addressed in practice, alternate designs should be considered. To evaluate these designs, we simulated a perfectly controlled cohort population and constructed the alternative designs from the cohort. This allowed us to make direct, valid comparisons.

A theoretical limitation of our results is that the various study designs used different analytic techniques to generate risk estimates. It is possible that differences between the designs may have been due to the regression methods, rather than inherent differences in the designs themselves. But, in theory, if the study is properly constructed—that is, the outcome and exposure are well measured and biases are minimized—the estimates derived from these designs should be unbiased estimates of the gold standard parameter. In other words, a well-executed case–control study should generate an unbiased estimate of a cohort relative risk estimate. We believe our simulation study created this scenario.

Another limitation is the simplicity of the simulation. Our study compared the four designs in the presence of one vaccination pattern (MMR), one type of time-varying confounding (seasonality), and a perfectly measured exposure and outcome. We did not explore several factors that can influence vaccine safety studies, including variable vaccination patterns (e.g., the administration of other routine vaccines, multiple vaccinations given concurrently), confounding by indication, confounding by contraindication, underlying health states that fluctuate over time, and misclassification of exposure and outcome [5]. We also did not evaluate the effect of time-varying confounding variables on each design as the incidence of disease decreases, or with a vaccine with a strong seasonal distribution, such as influenza vaccination. Additional studies are required to explore these other factors.

Another important area requiring further research is a form of bias known as the ‘‘healthy vaccinee effect’’ [28,29]. This refers to the notion that physicians may tend to either administer vaccinations during periods in which children are at their healthiest or withhold vaccinations from children with certain illnesses [5]. Thus, the time

period immediately preceding vaccination may represent an unusually healthy period where the incidence of illness is an underestimate of the true background rate of illness. Including this period in a self-controlled analysis could add a disproportionate amount of unexposed person-time into the analysis, which would lead to overinflated relative risk estimates. Suggested remedies for the healthy vaccinee bias include censoring a time period preceding the vaccination from the analysis [29] or starting person-time follow-up after vaccination [15].

Although we used the cohort design as the gold standard, its estimates appeared to be biased when compared to the true values of β_1. To investigate this further, we increased the incidence of disease by increasing β_0 to -14.50, -13.80, and -13.10—holding β_1 fixed at 0.693 (IRR = 2.00). The mean β_1 estimates (MSE) from 250 simulations with these β_0 values were 0.676 (0.022), 0.686 (0.010), and 0.693 (0.004), respectively. The mean disease incidence rate at each successive β_0 value was approximately 17, 38, and 81 cases per 100,000 person-years, respectively. The bias, therefore, appears to diminish as the incidence of disease increases, corroborating previously published data demonstrating that logistic regression tends to underestimate the relative risk when the outcome is rare [30]. Based on these published results, Poisson regression would also underestimate the true IRR because Poisson models approximate logistic models when the incidence of disease is low ($<5\%$).

As a validation exercise, we conducted 2,000 simulations (in increments of 250) across the four designs at an IRR of 2.00 and for two case incidence levels of approximately eight cases/100,000 person-years ($n \sim 400$ cases) and 1.4 cases/100,000 person-years ($n \sim 64$ cases) per simulation. The mean estimates within and between the designs remained relatively stable across the various simulation numbers. When compared to the cohort design, the mean percent bias for each design remained within 3% as the number of simulations increased. Therefore, our final results would not have changed had we increased the number of simulations from 250 to 2,000.

Our simulation study represents a typical scenario when evaluating the safety of a vaccine. We constructed a cohort with a common exposure (MMR) and an acute exposure period (42 days), and we simulated a rare, self-limiting illness (ITP). In this setting, the risk-interval, SCCS, and case–control designs were valid methods. However, each design demonstrated several different strengths and limitations. The case–control design minimized bias in the presence of seasonality, but it was less powerful than the other designs and produced comparatively high MSEs. Its IRR estimates were also biased when the confounding was unmeasured and stable across time (i.e., fixed). The risk-interval design, in contrast, produced stable, unbiased estimates when the unmeasured confounding was fixed, but its estimates were biased when the seasonal covariates were not incorporated into the Poisson regression models. Moreover, it required 50.0% (1,387,061/2,774,122) of the total cohort for analysis. The SCCS design displayed similar characteristics to those of the risk-interval, but required only 0.01% of the total study population for analysis. Thus, the SCCS method proved to be a valid and economic design that controls for unmeasured confounding variables unaffected by the passage of time.

References

[1] Black C, Kaye J, Jick H. MMR vaccine and idiopathic thrombocytopenic purpura. J Clin Pharmacol 2003;55:107–11.

[2] Miller E, Waight P, Farrington CP, et al. Idiopathic thrombocytopenic purpura and MMR vaccine. Arch Dis Child 2001;84(3):227–9.

[3] Szklo M, Nieto M. Basic study designs in analytical epidemiology. In: Szklo M, Nieto M, editors. Epidemiology: Beyond the basics. Gaithersburg, MD: Aspen Publishers, Inc., 2000. p. 3–52.

[4] Rothman K, Greenland S. In: Rothman K, Greenland S, editors. Modern epidemiology. 2nd ed. Philadelphia, PA: Lippincott-Raven Publishers; 1998. p. 79–91.

[5] Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. Am J Epidemiol 1992;136(2):121–35.

[6] Destefano F, Mullooly JP, Okoro CA, et al. Childhood vaccinations, vaccination timing, and the risk of type I diabetes mellitus. Pediatrics 2001;108(6):112–7.

[7] Murphy TV, Gargiullo PM, Massoudi MS, et al. Intussusception among infants given an oral rotavirus vaccine. N Engl J Med 2001; 344(8):564–72.

[8] Black S, Shinefield H, Ray P, et al. Risk of hospitalization because of aseptic meningitis after measles-mumps-rubella vaccination in one- to two-year-old children: an analysis of the Vaccine Safety Datalink (VSD) project. Pediatr Infect Dis J 1997;16(5):500–3.

[9] Mullooly JP, Pearson J, Drew L, et al. Wheezing lower respiratory disease and vaccination of full-term infants. Pharmacoepidemiol Drug Saf 2002;11:21–30.

[10] Mutsch M, Zhou W, Rhodes P, et al. Use of the inactivated intranasal influenza vaccine and the risk of bell's palsy in Switzerland. N Engl J Med 2004;350(9):896–903.

[11] Farrington PC, Miller E, Taylor B. MMR and autism: further evidence against a causal association. Vaccine 2001;19:3632–5.

[12] Kramarz P, DeStefano F, Gargiullo P, et al. Does influenza exacerbate asthma? Analysis of a large cohort of children with asthma. Arch Fam Med 2000;9(7):617–23.

[13] Chen RT. Safety of vaccines. In: Plotkin SA, Orenstein WA., editors. Vaccines. 3rd ed. Philadelphia, PA: W.B. Saunders Company; 1999. p. 1144–63.

[14] Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. Am. J Epidemiol 1996;143:1165–73.

[15] Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. Biometrics 1995;51:228–35.

[16] Andrew NJ. Statistical assessment of the association between vaccination and rare events post-licensure. Vaccine 2002;20:S49–53.

[17] Farrington CP. Control without separate methods: evaluation of vaccine safety using case-only methods. Vaccine 2004;22(15–16): 2064–70.

[18] Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. Stat Med 2006. [Epub ahead of print].

[19] Chen RT, Glasser JW, Rhodes PH, et al. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. Pediatrics 1997;99:765–73.

[20] Chu YW, Korb J, Sakamoto KM. Idiopathic thrombocytopenic purpura. Pediatr Rev 2000;21(3):95–104.

[21] Langholz B, Goldstein L. Risk set sampling in epidemiologic cohort studies. Stat Sci 1996;11:35–53.

[22] Greenland S. Estimation of exposure-specific rates from sparse case-control data. J Chronic Dis 1987;40(12):1087–94.

[23] Hosmer DM, Lemethow S. Interpretation of the coefficients of the logistic regression model. In: Hosmer DM, Lemeshow S, editors. Applied logistic regression. 2nd ed. New York: John Wiley and Sons, Inc., 2000. p. 8–81.

[24] Kleinbaum DG, Kupper LL, Muller KE, et al. Confounding and interaction in regression. In: Kleinbaum DG, Kupper LL, Muller KE, et al, editors. Applied regression analysis and other multivariable methods. 3rd ed. Pacific Grove, CA: Duxbury; 1998. p. 186–211.

[25] Pang D. A relative power table for nested matched case-control studies. Occup Environ Med 1999;56:67–9.

[26] Cologne JB, Sharp GB, Neriishi K, et al. Improving the efficiency of nested case–control studies of interaction by selecting controls using counter matching on exposure. Int. J Epidemiol 2004;33(3):485–92.

[27] Schneeweiss S, Sturmer T, Maclure M. Case–crossover and case–time–control designs as alternatives in Phamacoepidemiologic research. Pharmacoepidemiology and Drug Safety 1997;3:S51–9.

[28] Virtanen M, Peltola H, Paunio M, et al. Day-to-day reactogenicity and the healthy vaccinee effect of measles-mumps-rubella vaccination. Pediatrics 2000;106(5):e62.

[29] France EK, Glanz JM, Xu S, et al. Safety of the inactivated influenza vaccine among children: A population based study. Arch Pediatr Adolesc Med 2004;158(11):1031–5.

[30] King G, Zeng L. Logistic regression in rare events data. Soc Politic Methodol 2001;12(54):137–63.