ELSEVIER

Journal of Clinical Epidemiology 59 (2006) 808–818

Journal of Clinical Epidemiology

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

Jason M. Glanz[a,*], David L. McClure[a], Stanley Xu[a], Simon J. Hambidge[a,b,c], Martin Lee[d], Margarette S. Kolczak[e], Ken Kleinman[f], John P. Mullooly[g], Eric K. France[a]

[a]*Clinical Research Unit, Kaiser Permanente, P. O. Box 378006, Colorado, Denver, CO, 80237-8066, USA*
[b]*Community Health Services, Denver Health Medical Center, Denver, CO, USA*
[c]*Department of Pediatrics, University of Colorado School of Medicine, Denver, CO, USA*
[d]*UCLA Center for Vaccine Research, Torrance, CA, USA*
[e]*Centers for Disease Control and Prevention, Atlanta, GA, USA*
[f]*Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA*
[g]*Center for Health Research, Northwest Kaiser Permanente, Portland, OR, USA*

Accepted 16 November 2005

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

* Corresponding author. Tel.: 303-636-3118; fax: 303-636-3109.
E-mail address: jason.m.glanz@kp.org (J.M. Glanz).

0895-4356/06/$ – see front matter
doi: 10.1016/j.jclinepi.2005.11.012

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:

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

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