A comparison of the test-negative and traditional case-control study designs with respect to the bias of estimates of rotavirus vaccine effectiveness

Michael Haber, Emory University
Benjamin Lopman, Emory University
Jacqueline E. Tate, Centers for Disease Control and Prevention
Meng Shi, Emory University
Umesh D. Parashar, Centers for Disease Control and Prevention

Journal Title: Vaccine
Volume: Volume 36, Number 33
Publisher: Elsevier: 12 months | 2018-08-09, Pages 5071-5076
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1016/j.vaccine.2018.06.072
Permanent URL: https://pid.emory.edu/ark:/25593/v2zb8

Final published version: http://dx.doi.org/10.1016/j.vaccine.2018.06.072

Copyright information:
© 2018 Elsevier Ltd
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed April 3, 2020 8:35 AM EDT
A Comparison of the Test-negative and Traditional Case-control Study Designs with Respect to the Bias of Estimates of Rotavirus Vaccine Effectiveness

Michael Haber\textsuperscript{a,*}, Benjamin A Lopman\textsuperscript{b,c}, Jacqueline E. Tate\textsuperscript{c}, Meng Shi\textsuperscript{a}, and Umesh D Parashar\textsuperscript{c}
\textsuperscript{a}Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, U.S.A
\textsuperscript{b}Department of Epidemiology, Emory University, Atlanta, Georgia, U.S.A
\textsuperscript{c}Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A

Abstract

Estimation of the effectiveness of rotavirus vaccines via the test-negative control study design has gained popularity over the past few years. In this study design, children with severe diarrhea who test positive for rotavirus infection are considered as cases, while children who test negative serve as controls. We use a simple probability model to evaluate and compare the test-negative control and the traditional case-control designs with respect to the bias of resulting estimates of rotavirus vaccine effectiveness (VE). Comparisons are performed under two scenarios, corresponding to studies performed in high-income and low-income countries. We consider two potential sources of bias: (a) misclassification bias resulting from imperfect sensitivity and specificity of the test used to diagnose rotavirus infection, and (b) selection bias associated with possible effect of rotavirus vaccination on the probability of contracting severe non-rotavirus diarrhea.

Our results suggest that both sources of bias may produce VE estimates with substantial bias. Particularly, lack of perfect specificity is associated with severe negative bias. For example, if the specificity of the diagnostic test is 90\% then VE estimates from both types of case-control studies may under-estimate the true VE by more than 20\%. If the vaccine protects children against non-rotavirus diarrhea then VE estimates from test-negative control studies may be close to zero even though the true VE is 50\%. However, the sensitivity and specificity of the enzyme immunoassay test currently used to diagnose rotavirus infections are both over 99\%, and there is no solid evidence that the existing rotavirus vaccines affect the rates of non-rotavirus diarrhea. We therefore conclude that the test-negative control study design is a convenient and reliable alternative for estimation of rotavirus VE.

\textsuperscript{*}Corresponding author. Tel.: +1 404 727 7698; fax: +1 404 727 1370. mhaber@emory.edu (Michael Haber).

Conflicts of interest

The authors have no conflicts of interest.

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Keywords

Case-control Study; Test-negative study; Rotavirus; Vaccine effectiveness; Bias

Background

Rotavirus is the leading cause of severe diarrhea among children worldwide and was estimated to cause 215,000 childhood deaths in 2013 [1]. To prevent the burden of severe rotavirus diarrhea (SRD), the World Health Organization recommends routine vaccination of all children worldwide. Two vaccines against rotavirus were licensed for use in 2006 and have been introduced in the national immunization programs of more than 85 countries by late 2017. In randomized clinical trials, the efficacy of these vaccines against severe rotavirus diarrhea ranged from 85%–98% in high income settings to 50%–64% in low income settings. While the exact reasons for this variable vaccine performance are not known, factors that may adversely affect the performance of these orally administered, live virus vaccines in low income settings -- such as interference by concurrent enteric infections, malnutrition, high levels of maternal antibody, and interference with concurrently administered oral polio vaccine -- likely play a role. Given this variable performance of rotavirus vaccines in clinical trials, evaluations of vaccine effectiveness (VE) in routine programmatic use in diverse range of settings are a public health priority.

As vaccination against rotavirus is now recommended globally, randomized placebo-controlled clinical trials to evaluate vaccine efficacy are challenging to conduct. Therefore, observational studies based on patients seeking care or hospitalized for SRD are the best options for obtaining estimates of rotavirus VE. Cohort studies are not widely feasible due to low incidence of SRD; therefore case-control study designs are most commonly used. In these studies, the odds of vaccination are compared between individuals who contracted SRD (cases) and control individuals. Controls should be representative of the source population that produces the cases and should be like cases with respect to covariates (potential confounders) that may affect the chances of vaccination and contracting the diseases of interest.

In traditional case-control (TCC) studies, controls are selected from the same community (e.g., children in the same neighborhood) as the cases. Over the past decade, a new type of case-control study has evolved. In these test-negative control (TNC) studies, individuals seeking care for clinical symptoms like those of the cases but who test negative for the pathogen of interest serve as controls. TNC studies have mainly been used to estimate influenza VE [2–5], but they are also being used to estimate VE against other diseases, including rotavirus. (Tate et al [6], Schwartz et al [7]). TNC studies are expected to reduce confounding bias because cases and controls have similar symptoms and therefore are likely to have similar care-seeking behaviors and may also be similar with respect to other characteristics, such as age, comorbidities or access to health care. Such studies are also logistically easier and more economical to conduct, as no external control group needs to be recruited. By collecting vaccination data prior to knowledge of the test results, bias in ascertainment of vaccination among cases and controls can be avoided through TNC studies.
However, TNC studies may still be prone to selection bias, as controls may not represent the entire population that produces the cases. In particular, VE estimates from TNC studies will be biased if vaccination affects other pathogens that cause diseases that increase the likelihood of individuals to be included in the study as controls (by testing negative for the pathogen of interest) [8]. This may happen due to virus interference or cross-protection. Both types of case-control studies are prone to bias resulting from imperfect diagnostic tests and misclassification of vaccination status, but the extent of bias may differ.

In this work, we compare the biases associated with estimates of rotavirus VE against SRD from TNC and TCC studies. We focus on two sources of bias: (a) misclassification bias resulting from errors in the test for rotavirus infection (false positives or false negatives), and (b) bias associated with the possible effect of rotavirus vaccination on the probability of severe non-rotavirus diarrhea (SNRD).

Methods

We use a simplified version of the model developed by Shi et al [9] to evaluate the bias of estimates of influenza VE from TNC and TCC studies. The general model consists of five steps: (1) assigning a binary covariate corresponding to a person’s health status (‘healthy’ or ‘frail’); (2) vaccination; (3) developing illness (severe diarrhea, in our case) resulting either from the pathogen against which the vaccine is expected to protect (rotavirus, in our case) or from other pathogens (non-rotavirus infections, in our case); (4) seeking medical care (hospitalization, in our case); and (5) Testing positive or negative to the pathogen of interest. The probabilities of the possible outcomes in each step may depend on the outcomes of the previous steps. For example, the probability of being vaccinated may depend on health status, the probabilities of illness may depend on health status and vaccination status, etc. In this work, we do not account for a patient’s health status, and we assume that the probability of seeking care (i.e., hospitalization) does not depend on the patient’s vaccination status or on the etiology of her/his severe diarrhea. Below we describe the components of the model that are relevant to the current work.

Study Designs

Children hospitalized because of severe diarrhea are tested for rotavirus infection. In the TNC study, children with severe diarrhea who test positive for rotavirus infection serve as cases while those who test negative become controls. In the TCC study, cases are again defined as children with severe diarrhea who test positive for rotavirus infection, while controls are children who did not develop severe diarrhea during the study period. We assume that eligible controls are randomly selected from the population that produces the cases.

Vaccination

A child is considered effectively vaccinated 14 days after completing a full course of the rotavirus vaccine. We assume that a child’s vaccination status does not change during the study period and that the probability of a child with severe diarrhea being hospitalized does not depend on vaccination status.
True classification

A hospitalized study participant may suffer from either SRD or SNRD. The true etiology of infection is unknown before the child is tested.

Test for rotavirus infection

Children hospitalized for severe diarrhea are tested for rotavirus infection. We assume that the test’s sensitivity and specificity do not depend on the child’s vaccination status.

Our model includes the following parameters: The probability of being vaccinated (vaccine coverage), the probabilities of SRD and SNRD among vaccinated and unvaccinated children, and the diagnostic test’s sensitivity and specificity. The true VE against SRD is defined as 100% times one minus the ratio of the risks (probabilities) of SRD in vaccinated and unvaccinated children. We consider two scenarios for the values of these parameters:

Scenario A represents a high-income setting where incidence of SRD is relatively low and VE is relatively high, such as the U.S., while scenario B represents a low-income, high incidence setting such as sub-Saharan Africa. Table 1 lists the parameters and their values under both scenarios.

For each scenario we also consider a baseline case where the probability of SNRD in vaccinated children is the same as in unvaccinated children and the test’s sensitivity and specificity are both set to 100% We only expect minimal bias (or no bias) under the baseline cases.

For each of the two case-control study designs we used the values of the parameters to calculate the probabilities that a randomly selected child is classified as either a vaccinated case, vaccinated control, unvaccinated case or unvaccinated control. The estimate of VE is then calculated as 100% times one minus the ratio of the odds of being vaccinated in cases and in controls. The bias of an estimate is the difference between the estimated VE and the true VE. Methods for calculating the bias of estimated VEs for a given array of the model’s parameters have been developed by Shi et al [9]. A SAS program for calculating the bias under our model is available from the first author upon request.

We focused on two sources of bias: (1) lack of perfect sensitivity and specificity of the test for rotavirus infection, and (2) effect of the vaccine on the probability of contracting SNRD because of virus interference or cross-protection. For the second source of bias, the effect of the vaccine is quantified by the risk ratio comparing the probabilities of SNRD in vaccinated and unvaccinated. This risk ratio will be denoted RR(SNRD). For example, if the probability of SNRD in unvaccinated children is 0.03 and RR(SNRD) = 1.5 then the corresponding probability in vaccinated children is 0.045.

Results

The baseline cases

In the baseline cases, all the systematic sources of bias are absent. The TNC-based VE estimate is unbiased, while the TCC-based estimate has a small positive bias: the bias is 0.1% and 2% in scenarios A and B, respectively. In the absence of all systematic sources of
bias, the TNC-based estimate is unbiased even though the odds ratio is used to estimate a risk ratio. This result holds even when the ‘rare disease’ assumption is not satisfied (see next paragraph and [9]). The small bias in the TCC-based VE estimate results from using the odds ratio to estimate the risk ratio and it decreases when the true VE increases (such as in scenario A).

To help explain the lack of bias of the TN-based VE estimate regardless of whether the ‘rare disease’ assumption is satisfied, let’s consider the baseline case corresponding to Scenario B. The probability of being vaccinated is 0.8, the probability of SNRD is 0.19, regardless of vaccination status, and the probabilities of SRD are 0.06 and 0.03 in unvaccinated and vaccinated children, respectively. Then the true VE is $1 - 0.03/0.06 = 0.5$. Consider now a randomly selected child. The probability of this child being a vaccinated case is $0.8 \times 0.03 = 0.024$; the probability of being an unvaccinated case is $0.2 \times 0.06 = 0.012$; the probability of being a vaccinated control is $0.8 \times 0.19 = 0.152$; and the probability of being an unvaccinated control is $0.2 \times 0.19 = 0.038$. Hence, the expected odds ratio in the ensuing 2×2 table is \((0.024 \times 0.038)/(0.012 \times 0.152) = 0.5\), and the expected VE estimate is $1 - 0.5 = 0.5$, which is the same as the true VE. It is easy to check that this result would not change if the probabilities of SRD would be increased (for example, to 0.6 and 0.3 in unvaccinated and vaccinated children, respectively), as long as their ratio remains unchanged.

**Effects of imperfect sensitivity and specificity**

Figure 1 displays the true and estimated VEs from both study designs as a function of the specificity of the test for rotavirus infection for three values of the test’s sensitivity. In both scenarios, the negative bias resulting from imperfect specificity is quite large even when the sensitivity is high. For example, when the sensitivity is 1.0 and the specificity is 0.9 then the VE estimates are more than 20 percentage points lower than the true VE (70% vs. 90%) in scenario A. In scenario B, the negative bias exceeds 10 percentage points (true VE = 50%, estimated VEs less than 40%).

We also learn from Figure 1 that (a) the TCC-based VE estimate is usually higher than the TN-based estimate, hence its (negative) bias is smaller, and (b) the effect of lack of perfect sensitivity of the diagnostic test is less pronounced than that of lack of perfect specificity. The greater importance of the lack of perfect specificity than lack of perfect sensitivity is explained by the much higher probabilities of SNRD compared with those of SRD (Table 1). This is illustrated in the following example.

**Example 1**—Consider Scenario B with a population of size 1000, and suppose that the probability of SNRD does not depend on vaccination status. Then one would expect 19% of all hospitalized children, i.e. 190 children, to be true negatives. To calculate the expected number of true positives we separately consider vaccinated and unvaccinated children. One would expect $1000 \times 0.8 \times 0.03 = 24$ true positives among the vaccinated and $1000 \times 0.2 \times 0.06 = 12$ true positives among the unvaccinated, for a total of 36 true positives. Suppose further that both sensitivity and specificity are 90%. Then we expect 3.6 true positive children to test negative and 19 true negative children to test positive. Thus, the number of misclassified true negative children is over 5 times the number of misclassified true positive children, which
explains why the bias resulting from the 10% error rate in misclassifying true negatives is expected to be more pronounced than the bias attributable to the 10% error rate in misclassifying true positives.

Effect of vaccine modifying the probability of severe non-rotavirus diarrhea

Figure 2 presents the true and estimated VEs from both study designs as a function of the risk ratio (RR(SNRD)) comparing the probability of SNRD in vaccinated and unvaccinated children for three values of the diagnostic test’s specificity. We repeated this for different values of the test’s sensitivity and we found that sensitivity has almost no effect on the association between VE estimates and RR(SNRD). Therefore, we only present the results for a single value of the sensitivity, namely 0.9.

When the vaccine modifies the probability of SNRD, the bias of VE estimates from both study designs may be quite substantial. For example, when vaccinated children are twice as likely as unvaccinated children to contract SNRD then under scenario B the estimated VEs are 74% and 36% for the TNC and TCC study designs, respectively, while the true VE is 50%. In this case, the VE estimates under scenario A are less biased (true VE 90%, estimated VEs 94.7% and 89.8%, for TN and TCC studies, respectively). In general, the absolute value of the bias increases when the specificity decreases, and it is smaller under scenario A as compared to scenario B.

Even though children with SNRD are not included in the TCC study, VE estimates from TCC studies are biased when vaccination affects the probability of SNRD. This happens because only children who did not develop severe diarrhea, regardless of etiology, qualify as controls. As a result, children with SNRD are not included as controls in the TCC study even though they did not develop SRD (i.e., they are not true cases of SRD). If vaccination affects the probability of SNRD then the proportion of children without SRD who are wrongfully excluded is not the same in vaccinated and unvaccinated individuals. This is illustrated in the following example.

Example 2—Consider scenario B and suppose that vaccinated children are twice as likely as unvaccinated to contract SNRD. Let’s look at the population of 1,000 children: 800 of them (80%) will be vaccinated and 200 will remain unvaccinated. Then 200×0.06 = 12 unvaccinated and 800×0.03 = 24 vaccinated children are expected to contract SRD. In addition, 200×0.19 = 38 unvaccinated and 800×0.38 = 304 vaccinated children are expected to develop SNRD. The estimated VE from the TNC study is therefore 100×{1 – (24×38)/(304×12)} = 75%, while the true VE is 50%. TCC controls are selected at random from those who did not contract any type of severe diarrhea. Of the 200 vaccinated, 12 and 38 children contact SRD and SNRD, respectively, therefore there are 150 unvaccinated children who do not contract severe diarrhea. Similarly, 800 – (24 + 304) = 472 vaccinees do not contact severe diarrhea. Therefore, the estimated VE from the TCC study is 100×{1 – (24×150)/(472×12)} = 36.4% while the true VE is 50%. The bias of both VE estimates is driven by the large number (304) of vaccinees who contract SNRD. If the probability of SNRD in vaccinated children was the same (0.19) as in unvaccinated children that we would...
expect only $800 \times 0.19 = 152$ vaccinees with SNRD, and the VE estimates from the TNC and TCC studies would be 50% and 51.9%, respectively.

**Discussion**

We used a previously published model [9] to evaluate the bias of estimates of rotavirus vaccine effectiveness from test-negative and traditional case-control studies. We considered two sources of bias: (1) lack of perfect sensitivity and specificity of the test for rotavirus infection, and (2) possible effect of the vaccine on the probability of contracting non-rotavirus diarrhea.

We found that both types of case-control studies may produce biased estimates of VE when either or both sources of bias are present. Even when the specificity is about 95%, estimated VEs may be substantially lower than the true VE. For example under scenario B, if the specificity of the test is 95% then the estimated VE is expected to be 38% while the true VE is 50%. The effect of the sensitivity of the test on the bias is less pronounced. The bias of estimates from TNC studies when vaccination affects the probability of developing SNRD is not surprising as cases of SNRD serve as controls in the TNC design. The TNC design relies on the assumption that the proportion of vaccinated children among children with SNRD is close to the corresponding population proportion, and this assumption is violated when rotavirus vaccination modifies the probability of non-rotavirus infections [8]. We found that this source of bias also affects VE estimated from TCC studies, as children with SNRD may not serve as controls in TCC studies. Vaccine status-related differences in the proportion of potential controls who are excluded (because they have SNRD) results in biased VE estimates.

We only considered two sources of bias in this work. Other sources, such as bias due to confounding (i.e., the presence of factors that may be associated with both vaccination and the probabilities of SRD and SNRD) can be adjusted for in the analysis. We also examined the effect of other factors on the magnitude of the biases resulting from the two sources we considered. We found (results not shown) that variations in the incidence of SRD have minimal effect on the bias, while variations in the true effectiveness of the vaccine do affect the magnitude of the biases: as the true VE increases, the biases decrease. Our model is static and does not account for the dynamics of outbreaks or for temporal changes in the probabilities of severe diarrhea. We plan to develop and apply stochastic agent-based dynamic simulation models to study the bias of rotavirus VE estimates in more realistic settings.

While both sources of bias examined in this study can affect rotavirus VE estimates from case-control studies, the influence of these biases are likely to be small, if any. Rotavirus infections are typically diagnosed by enzyme immunoassay (EIA), with reported sensitivity and specificity between 99% and 100% [14, 15]. Additionally, while the data are sparse, no solid evidence exists that rotavirus vaccines affect the probability of SNRD [16, 17]. We therefore conclude that in reality, the biases resulting from the sources we considered are likely to be small. Based on our findings, we believe that the test-negative design, which offers a particularly useful design to evaluate rotavirus VE post-licensure, can be expected to
provide reliable estimates. On the other hand, other vaccines may have non-specific effects that modify the incidence of symptoms that also result from the pathogen of interest. For example, the oral polio vaccine has been found to produce non-specific effects [18]. Therefore, we plan to apply the model and methods presented in the work to other vaccines.

Acknowledgments

Funding

This research was supported by the National Institute of Allergies and Infectious Diseases of the National Institutes of Health (NIH) under Award R01AI110474, and by IPA 17IPA17107 with the Centers for Disease Controls and Prevention (CDC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the CDC.

References


Vaccine. Author manuscript; available in PMC 2019 August 09.


Figure 1. Effects of Imperfect Sensitivity and Specificity on the Bias of VE Estimates.
Figure 2.
Effect of the Vaccine Modifying the Probability of Severe Non-Rotavirus Diarrhea on the Bias of VE Estimates when the Sensitivity of the Test is 0.9. RR(SNRD) is the Risk Ratio of Contracting Sever Non-Rotavirus Diarrhea in Vaccinated and Unvaccinated Children.
Table 1

Values of the model parameters under two scenarios: Scenario A corresponds to a high-income country and Scenario B to a low-income country.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scenario A</th>
<th>Scenario B</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of being vaccinated</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Probability of SRD in unvaccinated children</td>
<td>0.01&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.06&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>True vaccine effectiveness (VE)</td>
<td>90%&lt;sup&gt;c&lt;/sup&gt;</td>
<td>50%&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Probability of SRD in vaccinated children</td>
<td>0.001</td>
<td>0.03</td>
<td>Probability in unvaccinated multiplied by (1-VE/100)</td>
</tr>
<tr>
<td>Probability of SNRD in unvaccinated children</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.19&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Risk ratio comparing probabilities of SNRD in vaccinated and unvaccinated</td>
<td>0.5 – 2.0</td>
<td>0.5 – 2.0</td>
<td>Baseline value = 1.0</td>
</tr>
<tr>
<td>Probability of SNRD in vaccinated children</td>
<td>0.015 – 0.06</td>
<td>0.095 – 0.38</td>
<td>Probability in unvaccinated multiplied by RR(SNRD)</td>
</tr>
<tr>
<td>Sensitivity of diagnostic test</td>
<td>60% – 100%</td>
<td>60% – 100%</td>
<td>Baseline value = 100%</td>
</tr>
<tr>
<td>Specificity of diagnostic test</td>
<td>50% – 100%</td>
<td>50% – 100%</td>
<td>Baseline value = 100%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Tate et al. [6];

<sup>b</sup>Kotloff et al. [10];

<sup>c</sup>Vesikari et al. [11];

<sup>d</sup>Arma et al. [12], Zaman et al. [13].