Magnitude of Potential Biases in a Simulated Case-Control Study of the Effectiveness of Influenza Vaccination

Jill M. Ferdinands1,2 and David K. Shay1

1Influenza Division, Centers for Disease Control and Prevention; and 2Battelle, Atlanta, Georgia

Background. Many influenza vaccine effectiveness estimates have been made using case-control methods. Although several forms of bias may distort estimates of vaccine effectiveness derived from case-control studies, there have been few attempts to quantify the magnitude of these biases.

Methods. We estimated the magnitude of potential biases in influenza vaccine effectiveness values derived from case-control studies from several factors, including bias from differential use of diagnostic testing based on influenza vaccine status, imperfect diagnostic test characteristics, and confounding. A decision tree model was used to simulate an influenza vaccine effectiveness case-control study in children. Using probability distributions, we varied the value of factors that influence vaccine effectiveness estimates, including diagnostic test characteristics, vaccine coverage, likelihood of receiving a diagnostic test for influenza, likelihood that a child hospitalized with acute respiratory infection had influenza, and others. Bias was measured as the difference between the effectiveness observed in the simulated case-control study and a true underlying effectiveness value.

Results and Conclusions. We found an average difference between observed and true vaccine effectiveness of −11.9%. Observed vaccine effectiveness underestimated the true effectiveness in 88% of model iterations. Diagnostic test specificity exhibited the strongest association with observed vaccine effectiveness, followed by the likelihood of receiving a diagnostic test based on vaccination status and the likelihood that a child hospitalized with acute respiratory infection had influenza. Our findings suggest that the potential biases in case-control studies that we examined tend to result in underestimates of true influenza vaccine effects.

Beginning in 2006, the Advisory Committee on Immunization Practices (ACIP) has substantially expanded the target population recommended for annual influenza vaccination in the United States, first by recommending vaccination of all children aged 6 to 59 months, then expanding the recommendation to include all children aged 6 months to 18 years in 2008, and, finally, expanding the recommendation to include everyone aged 6 months and older in 2010 [1]. Because of ethical concerns surrounding the conduct of randomized controlled trials with placebo arms among individuals for whom vaccination is recommended, estimates of influenza vaccine effects in the United States have increasingly relied on observational studies, particularly case-control studies. Case-control studies are often favored for observational studies of the effectiveness of licensed vaccines for reasons of statistical power, logistics, and cost. However, case-control studies are prone to bias, which can lead to errors in estimates of vaccine effectiveness (VE).

Bias arises from numerous sources. These include selection bias, which occurs if there are systematic differences in the relationship between vaccination and disease between those who participate in the study and those who do not; confounding, which may occur when participant characteristics that are associated with both influenza vaccination and diagnosis of influenza distort the true relationship between vaccination and disease; and misclassification of study participants due to errors in measurement of exposure or outcome status [2, 3]. Although
biases in VE estimates derived from case-control studies have been recognized as potential problems, there have been only a few attempts to measure their magnitude and understand their influence on VE estimates. In this study, we sought to assess the magnitude of potential biases in VE estimates derived from case-control studies due to several factors, including bias from differential use of diagnostic testing based on influenza vaccine status, misclassification of outcome status due to use of imperfect diagnostic tests, and confounding resulting from the possibility that individuals vaccinated for influenza are more likely to suffer complications of influenza (and therefore be enrolled as cases). To do this, we examined biases arising in a simulated case-control study of influenza vaccine effects in children. We focused on children because there is less concern about the effects of unmeasured confounders (such as functional status) in this population compared with VE studies among adults—particularly studies among elderly adults, in whom the “healthy vaccinee” effect is likely important to consider [4–6].

METHODS

Decision Tree Structure
We examined potential biases by using a decision tree model to simulate an influenza VE case-control study in children (Figure 1). The decision tree model mimicked the structure and outcomes of an influenza VE case-control study in which cases are children hospitalized with acute respiratory infection (ARI) who tested positive with an influenza diagnostic test ordered at the discretion of a clinical provider (rather than through random or systematic sampling) and controls are age-matched children not hospitalized for influenza. As depicted by the branches of the decision tree, study participants at the time of hospitalization were vaccinated with probability $p_{vacc}$. In the unvaccinated subset of participants, the proportion denoted by $p_{flu}$ had influenza. In the vaccinated subset of participants, the proportion with influenza was a function of the true underlying influenza VE and was denoted by $p_{flu}*(1-VE)$. Vaccinated and unvaccinated participants had a probability of receiving a diagnostic test for influenza denoted by $p_{test_vacc}$ and $p_{test_unvacc}$, respectively. Diagnostic test results were either positive or negative depending on true disease status and sensitivity and specificity of the diagnostic test used. Each tested participant was either a true positive or a false negative (if the participant truly had influenza) or a true negative or a false positive (if the participant truly did not have influenza). A participant who tested positive, whether a true positive or a false positive, was classified as an observed case, as would happen in an actual case-control study. Similarly, a participant who tested negative or was not tested would be ineligible and excluded from the study. In this and the following sections, decision tree variables are denoted in italics for cross-reference with Figure 1.

![Decision Tree Model](image-url)
**Decision Tree Input Values**

We varied the value of important factors known to affect VE, including diagnostic test specificity and sensitivity, vaccine coverage, likelihood of receiving a diagnostic test for influenza, the likelihood that a child hospitalized with ARI had influenza, and the extent to which vaccinated participants were more likely to be hospitalized for influenza [3, 7, 8]. Each of these variables was defined by a probability distribution—a summary of the possible values of the variable—based on results from observational studies, randomized trials of influenza vaccine effects, or expert opinion (Table 1). For most variables, we used a triangular distribution, which is summarized by its minimum, maximum, and most likely values; this distribution is typically used for variables about which little is known [22]. For variables for which more information is available, we used beta distributions, which can reflect a wide range of distributional shapes and, like probabilities, are bounded by zero and one [22].

**Test Characteristics**

For test sensitivity and specificity, we conducted a literature review to define the performance characteristics of influenza diagnostic tests (details in the online supplementary material). We then defined beta distributions that matched the range of values observed in the literature. Based on 29 studies of diagnostic tests for influenza published since 2005 and including children, we found mean sensitivity (Figure 2) and specificity (Figure 3) of commercial rapid influenza antigen detection tests to be 61% and 98%, respectively, excluding tests for H5, H7, and 2009 pandemic H1N1 strains. Mean sensitivity and specificity of direct immunofluorescence assay (DFA) were assumed to be 80% and 98%, respectively [23]. In the simulation model, we used a weighted average of test sensitivity and specificity that reflected the frequency of use of the various types of influenza diagnostic tests in a recent large pediatric case-control study of influenza VE. Specifically, we assumed that 64% of tests were commercial rapid antigen tests, 21% were DFA, 7% were viral culture, 2% were reverse-transcription polymerase chain reaction (RT-PCR), and 3% were multiple test types (CDC, unpublished data), resulting in weighted average test sensitivity and specificity of 69% and 98%, respectively.

**True VE**

Mean true VE (VE) was assumed to be 67% based on the arithmetic average of estimates of vaccine efficacy against culture- or RT-PCR–confirmed, nonantigenically matched influenza infection from recent randomized trials of influenza vaccine among adults [14–20] and, because no recent trials among children are available, an earlier trial of inactivated influenza vaccine including children [21]. The distribution of true VE was modeled with a beta distribution fitted to the distribution of vaccine efficacy estimates from these trials.

<table>
<thead>
<tr>
<th>Table 1. Decision Tree Model Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td>Commercial rapid diagnostic test sensitivity (sensitivity)</td>
</tr>
<tr>
<td>Commercial rapid diagnostic test specificity (specificity)</td>
</tr>
<tr>
<td>Likelihood of receiving a diagnostic test for influenza if vaccinated (p_test_vacc)</td>
</tr>
<tr>
<td>Likelihood that a child hospitalized with acute respiratory infection is vaccinated for influenza (p_vacc)</td>
</tr>
<tr>
<td>Likelihood that a child hospitalized with acute respiratory infection has influenza (p_flu)</td>
</tr>
<tr>
<td>True vaccine effectiveness (VE)</td>
</tr>
<tr>
<td>Multiplicative increase in likelihood of hospitalization given vaccination (f_1)</td>
</tr>
</tbody>
</table>

Abbreviation: VE, vaccine effectiveness.

* Identifies model inputs expressed in the table as percentages for readability, but entered as the equivalent proportions (ranging from 0 to 1) in the simulation model.
**Vaccination Rate**

Vaccine coverage among children hospitalized with ARI ($p_{vacc}$) was based on Poehling et al [11], who found that 23% of children younger than five years old and hospitalized with fever or respiratory symptoms had received influenza vaccine during the 2002–2003 and 2003–2004 influenza seasons. We modeled this variable using a triangular distribution with a most likely value of 23% and a range from 15% to 32%.

**Receipt of Diagnostic Test**

Based on Grijalva et al [9], we estimated a likelihood of 52% that a child hospitalized with ARI received a diagnostic test for influenza, and that the most likely scenario was that there was no difference between rates of testing among vaccinated and unvaccinated children [10]. However, to explore bias from potential differential testing based on vaccination status, we allowed the difference in testing rates between vaccinated ($p_{test\_vacc}$) and unvaccinated ($p_{test\_unvacc}$) children to vary up to a relative difference of 50%. Thus, at the extreme, vaccinated children would be tested only 29% of the time while unvaccinated children would be tested 59% of the time.

**Rate of Influenza Infection**

The likelihood that a child hospitalized with ARI had influenza ($p_{flu}$) was assumed to be 10%, the arithmetic average of five rates from three studies [11–13], and to range from 1% to 20%. These rates reflect averages across the influenza season.

**Confounding**

Children and older persons with certain chronic medical conditions may be both more likely to be vaccinated for influenza and more likely to suffer serious complications of influenza infection. If not adjusted for, such confounding will underestimate VE [24–26]. To mimic this type of confounding, we added a multiplicative factor ($f_1$) to the model to represent the increase in the likelihood of being hospitalized for influenza given vaccination (independent of VE). For example, a factor of 1.10 indicated that vaccinated participants were 10% more likely to be hospitalized for influenza because of a chronic condition that increases the risk of serious influenza complications. We assumed that the most likely value for this variable was an increase of 20%, with a range from 0% to 50% (corresponding to a most likely value of $f_1$ of 1.20 with a range of 1.00–1.50).

**Correlation Between Input Variables**

Based on our literature review of 29 recent studies, we found no statistically significant correlation between the sensitivity and specificity of commercial rapid antigen tests (Spearman correlation coefficient = $-0.04$, $P = .77$). Although it seemed plausible to expect a positive correlation between the likelihood that a child hospitalized with ARI truly had influenza and the likelihood of receiving a diagnostic test for influenza, we found no data to establish an estimate of this correlation. When we
examined the influence of this correlation on the model results by running the model with progressively stronger positive correlations, we found that this correlation, even at its strongest, influenced the bias estimates by less than 0.5%. Therefore, we excluded this potential correlation from our model.

**Simulation Methods**

Using Monte Carlo methods [22], we simulated the total number of cases, the number of vaccinated cases, and the ratio of vaccinated to unvaccinated cases. This ratio multiplied by the ratio of unvaccinated to vaccinated controls gave the observed odds ratio; 1 minus the observed odds ratio gave the observed VE. Bias was the difference between true and observed VE. We examined the sensitivity of observed VE to model inputs by regressing the simulated observed VE on each input of interest, controlling for true VE. The ratio of unvaccinated to vaccinated controls was assumed to be 4.5, the value that equated true VE and observed VE assuming perfect test characteristics and no differential diagnostic evaluation. This ratio is equivalent to a probability of vaccination of 0.23 in the control population, within the range of reported estimates of vaccine coverage in 6- to 23-month-olds since the 2004–2005 influenza season (18% in 2004–2005 to 32% in 2006–2007, when vaccination for this age group was first recommended (2006–2007 vaccination coverage estimate from CDC website reporting data from the National Immunization Survey, available at http://www.cdc.gov/flu/professionals/acip/coveragelevels.htm [accessed April 8, 2011]). We used ten simulations of one million iterations each, which provided convergence of bias estimates to within 0.01%. Simulations were conducted using @RISKv5.7 (Palisade Corp, Ithaca, NY).

**RESULTS**

We found a mean observed VE of 55% compared with a mean true VE of 67% (Figure 4). The average difference between observed and true VE was −11.9%. The observed vaccine effectiveness underestimated the true effectiveness in 88% of model iterations. Overall, 57% of observed VE estimates were more than 10% below true VE, and 20% of observed VE estimates were more than 20% below true VE. The lowest observed VE was 77% lower than true VE, and the highest observed VE was 36% greater than true VE. Large underestimates of VE occurred in situations with a high true VE, low rapid test specificity, and a low likelihood of influenza in children hospitalized for ARI.

![Figure 4. Distributions of observed (solid) and true (dotted) vaccine effectiveness based on a simulated pediatric case-control study of influenza vaccine effectiveness (VE). The mean observed VE was 55% compared with mean true VE of 67%.](image)

![Figure 5. “Tornado” diagram showing the inputs that influence observed vaccine effectiveness (VE) in rank order from most influence exerted to least. The axis gives the estimated regression coefficient for each input when observed VE (%) is regressed on each input separately, controlling for true VE. These values can be interpreted as the percent change in observed VE for each 1% change in the respective model input. Abbreviation: ARI, acute respiratory infection.](image)
with ARI. Overestimation of VE occurred in situations with low true VE, high rapid test specificity, and low likelihood of receipt of a diagnostic test among vaccinated children.

Diagnostic test specificity exhibited the strongest association with observed VE, followed by the likelihood of receiving a diagnostic test based on vaccination status and the likelihood that a child hospitalized with ARI had influenza (Figure 5). Holding true VE constant, a 1% decrease in test specificity (eg, from 98% to 97%) was associated with a 4.2% decrease in observed VE. Test characteristics were positively associated with observed VE: the better the test, the less underestimation of VE. Differential diagnostic testing, in which vaccinated children are tested less frequently than unvaccinated children, was negatively associated with observed VE. The lower the probability of testing in vaccinated children, the more inflated the observed VE. Every 1% reduction in testing rate in vaccinated children was associated with 1.3% higher observed VE. The probability that a child hospitalized with ARI had influenza also influenced observed VE: for every 1% increase in this probability, observed VE increased by 1%. The association between observed VE and the confounding that we modeled was weaker but in the expected direction: for every 1% increase in the likelihood of influenza hospitalization among vaccinated children, observed VE was, on average, 0.2% lower.

DISCUSSION

By simulating a case-control study of influenza VE in children, we found that the sources of bias we included, when considered together, more often led to an underestimation rather than an overestimation of true vaccine benefit. In only 12% of model iterations did observed VE overestimate true VE. Our findings suggest that misclassification due to imperfect test specificity is an influential source of bias. It is increasingly common for case-control studies of VE to use highly sensitive and specific RT-PCR assays for diagnostic testing, making imperfect test specificity less important. However, during pandemics or other situations in which estimates of VE are desired but access to RT-PCR is limited, imperfect test specificity could be an important source of bias. Differential diagnostic testing was also a nontrivial source of bias in our simulation and, as expected, contributed to bias in an opposite direction. The confounding we modeled—that more severely ill children were more likely to be vaccinated as well as more likely to be hospitalized with influenza—was less influential than the other sources of bias we assessed.

The potential for bias in observational studies of vaccine effectiveness has long been recognized as a limitation of such studies [24, 25, 27–31], and numerous strategies to minimize bias have been proposed. These include statistical methods such as multivariate regression, use of propensity scoring methods, and restriction of sample populations to enhance homogeneity of the study sample [32, 33]. In addition, methods comparing outcomes during time periods with and without influenza circulation [4, 31, 34] and use of detailed sensitivity analyses [35] have been suggested as approaches for deriving less biased estimates of VE. Despite these methodologic advances, controversy continues about the most important biases in VE estimates from observational studies, and it remains unclear if biases, in aggregate, tend to over- or under-estimate true vaccine benefit. Our study offers insight into this debate by providing a quantitative estimate of the impact on VE from several potential biases evaluated in the same simulation model.

Our study has limitations. Because we focused on a VE study among children, our results may not generalize to other age groups; a recent study of VE among the elderly suggests that differential testing may lead to more bias than we observed [36]. Not every variable in our model was described by a probability distribution; hence, we did not represent all inherent uncertainty in model inputs. Some of the variables represented by probability distributions had little data upon which to base the distributions; the study would be strengthened by additional empirical data for these variables. There is inherent uncertainty in any model regarding the structure of the model itself. However, some features of our findings support the face validity of the model. First, our results are similar to those of Orenstein et al [8], with both studies suggesting that imperfect test characteristics can contribute to dramatic underestimation of observed VE. Secondly, we found that the likelihood of influenza infection in a child hospitalized with ARI had a positive association with observed VE. This result is consistent with well-described data that the prevalence of influenza is associated with predictive values of influenza tests and estimates of VE based on those tests.

Although we examined bias in VE estimates due to imperfect test characteristics, differential diagnostic testing based on vaccination status, and one type of confounding, other sources of bias exist. Of particular importance are sources of potential selection bias that can distort estimates of VE. The over- or underestimation of VE arising from different types of selection bias should be systematically evaluated in future research. In addition, the bias in VE estimates derived from study designs created to minimize selection bias, such as the so-called test-negative control design [37], should be assessed.

Influenza VE estimates derived from observational studies have yielded VE estimates that range from 0% to 93% [38, 39]. One factor explaining this wide range of estimates is the variation by season in the match between vaccine and circulating influenza strains, with higher effectiveness found in seasons when antigenic relatedness between vaccine and circulating strains is high [14, 19, 38–40]. Other factors known to affect influenza VE include failure to mount a protective immune
response, perhaps due to host immune compromise or immunosenescence; failure to be immunized prior to exposure; and the specific influenza vaccine type and age group studied. The recommendation by the ACIP [1] that all US residents receive annual influenza vaccination likely means that future comparisons of influenza outcomes in vaccinated and unvaccinated individuals in the United States will rely on observational studies. A better understanding of the magnitude of potential biases in observational studies, particularly case-control studies, will be essential if we seek to monitor the effectiveness of the influenza vaccine program through annual assessments of the effectiveness of licensed influenza vaccines.

**Supplementary Data**

Supplementary materials are available at *Clinical Infectious Diseases* online (http://www.oxfordjournals.org/our_journals/cid/). Supplementary materials consist of data provided by the authors and published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the authors.

**Notes**

**Financial support.** This work was sponsored by the Centers for Disease Control and Prevention.

**Potential conflicts of interest.** All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**


