Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness

Michael L. Jackson, Kenneth J. Rothman

A R T I C L E   I N F O
Article history:
Received 20 October 2014
Received in revised form 23 December 2014
Accepted 27 January 2015
Available online 7 February 2015

Keywords:
Influenza
Human
Methodology
Bias (Epidemiology)
Vaccine effectiveness

A B S T R A C T

Background: The recently developed test-negative design is now standard for observational studies of influenza vaccine effectiveness (VE). It is unclear how influenza test misclassification biases test-negative VE estimates relative to VE estimates from traditional cohort or case-control studies.
Methods: We simulated populations whose members may develop acute respiratory illness (ARI) due to influenza and to non-influenza pathogens. In these simulations, vaccination reduces the risk of influenza but not of non-influenza ARI. Influenza test sensitivity and specificity, risks of influenza and non-influenza ARI, and VE were varied across the simulations. In each simulation, we estimated influenza VE using a cohort design, a case-control design, and a test-negative design.
Results: In the absence of influenza test misclassification, all three designs accurately estimated influenza VE. In the presence of misclassification, all three designs underestimated VE. Bias in VE estimates was slightly greater in the test-negative design than in cohort or case-control designs. Assuming the use of highly sensitive and specific reverse-transcriptase polymerase chain reaction tests for influenza, bias in the test-negative studies was trivial across a wide range of realistic values for VE.
Discussion: Although influenza test misclassification causes more bias in test-negative studies than in traditional cohort or case-control studies, the difference is trivial for realistic combinations of attack rates, test sensitivity/specificity, and VE.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, the so-called “test-negative” design has become the standard approach for observational studies of influenza vaccine effectiveness (VE) [1–3]. In a test-negative design, the study population comprises patients who present to an outpatient clinic or hospital with acute respiratory illness (ARI) and who are tested for influenza infection [4]. VE is defined as one minus the ratio of the risk of influenza among the vaccinated to the corresponding risk among the unvaccinated. In case-control and test-negative studies, VE is estimated as one minus the odds ratio of influenza for vaccinated vs. unvaccinated. Relative to some other observational designs, the test-negative design offers the advantage of reduced confounding from differences in healthcare-seeking behavior between vaccinated and unvaccinated persons [5].

Misclassification of influenza leads to biased VE estimates, regardless of the study design. Assuming the misclassification is not differential by vaccination status, misclassification will introduce bias that will tend to underestimate VE. The degree of bias due to misclassification has been believed to be low in test-negative studies, primarily due to a 2007 paper by Orenstein and colleagues [6]. In that paper, the authors concluded that case-control and test-negative studies were less biased than cohort studies in populations subject to similar amounts of misclassification of influenza. That paper, however, had an important flaw. The authors based their calculations on a cumulative design [7] for the case-control and test-negative studies, in which the controls are sampled from those who did not get influenza during the follow-up period. Controls in a case-control study correspond to the denominator information in a cohort study. The cumulative sampling strategy excludes those who get influenza from the sampled denominators, biasing study results away from the null [7]; this bias is small for rare disease but is larger if the disease is common. In the paper by Orenstein et al, the bias toward the null in VE stemming from misclassification of disease was countered in the test-negative and case-control design by a bias away from the null due to the cumulative design.
We ran a series of 1000 simulations to compare the study designs. In each, we simulated a population of 50,000 subjects, which gives study sizes roughly equal to those in existing observational VE studies [1,3,11]. We assumed that 40% of subjects received influenza vaccine at the start of follow-up. Within the follow-up period, subjects could be infected (up to once) with influenza, with risk equal to IP_{flu}, and infected (up to once) with a non-influenza pathogen, with risk equal to IP_{other}. We assumed that risk of influenza is independent of the risk of non-influenza ARI, as analyses of the effects of hypothetical non-independence on test-negative VE estimates have been performed previously [12]. After running the simulation to determine the simulated disease events, we randomly allowed these simulated events to be misclassified according to the rapid test and RT-PCR values for sensitivity and specificity.

We estimated VE using three separate designs, first using the correctly classified outcomes. In the cohort design, we calculated the risk of influenza infection in the vaccinated and the unvaccinated. We then estimated VE_{cohort} as (1 – RR), where RR is the risk ratio. In the case-control design, for each detected influenza case, we randomly sampled three controls from the total study population. We estimated VE_{cc} as (1 – OR_{cc}), where OR_{cc} is the odds ratio from the case-control study, which estimates the risk ratio [7]. For the test-negative design, we included all ARI events testing positive for influenza as cases. All ARI events testing negative for influenza were selected as a comparison group. We estimated VE_{in} as (1 – OR_{in}), where OR_{in} is the odds ratio from the test-negative study. We calculated 95% confidence limits from the 2.5th and 97.5th percentiles of the simulations.

After estimating VE in the simulated population based on the true disease status, we repeated the analysis using the rapid test misclassified outcomes, and again using the RT-PCR misclassified outcomes. We calculated the bias of each design from each simulation as a percent: Bias = [(VE – VE) / VE] × 100%. For each design at each level of misclassification, we calculated the mean bias with 95% confidence limits.

We further assessed the independent effects of influenza test sensitivity and test specificity on VE estimates in the test-negative design, using the “all ages” scenario. For this, we ran 1000 simulations assuming IP_{flu} = 5%, IP_{other} = 10%, and VE = 50%. In each simulated population, we calculated VE_{in} at a range of test sensitivities (from 0.8 to 1.0, keeping specificity at 1.0) and at a range of specificities (from 0.8 to 1.0, holding sensitivity at 1.0). Finally, we assessed the degree to which bias in the cohort and test-negative designs varies with varying VE. For this, we ran 1000 simulations, assuming PCR sensitivity and specificity, IP_{flu} = 5%, and IP_{other} = 10%, while varying VE between 10 and 70%.

Finally, we conducted sensitivity analyses of the “young children” and “all ages” scenarios, where we allowed subjects to have multiple influenza and non-influenza ARI events during follow-up. Instead of incidence proportions, the number of events per person was randomly sampled from a Poisson distribution with mean equal to IP_{flu} (for influenza ARI) and IP_{other} (for non-influenza ARI). Results were trivially different from the main analyses for all study designs in both scenarios (less than one percentage point difference in estimated VE at PCR levels of misclassification) and are not further presented here. Analyses were conducted using SAS Version 9.3 (SAS Institute, Cary NC) and R Version 3.0.2 (The R Foundation for Statistical Computing, Vienna, Austria).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>VE_{true}</td>
<td>Vaccine effectiveness</td>
<td>All ages: 50%, Young children: 70%</td>
</tr>
<tr>
<td>IP_{flu}</td>
<td>Incidence proportion of ARI due to influenza</td>
<td>All ages: 5%, Young children: 15%</td>
</tr>
<tr>
<td>IP_{other}</td>
<td>Incidence proportion of ARI due to other pathogens</td>
<td>All ages: 10%, Young children: 30%</td>
</tr>
<tr>
<td>Sens (RT)</td>
<td>Rapid test sensitivity</td>
<td>All ages: 80%, Young children: 90%</td>
</tr>
<tr>
<td>Spec (RT)</td>
<td>Rapid test specificity</td>
<td>All ages: 90%, Young children: 95%</td>
</tr>
<tr>
<td>Sens (PCR)</td>
<td>RT-PCR sensitivity</td>
<td>All ages: 90%, Young children: 95%</td>
</tr>
<tr>
<td>Spec (PCR)</td>
<td>RT-PCR specificity</td>
<td>All ages: 97%, Young children: 97%</td>
</tr>
</tbody>
</table>

* ARI = acute respiratory illness.
Table 2
Estimated vaccine effectiveness (VE) and bias from 1000 “young children” simulations under different study designs and levels of misclassification, assuming 70% VE.

<table>
<thead>
<tr>
<th>Design</th>
<th>Misclassification</th>
<th>Estimated VE</th>
<th>Percent bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>None</td>
<td>70% (68–72%)</td>
<td>0%</td>
</tr>
<tr>
<td>Cohort</td>
<td>RT-PCR</td>
<td>66% (63–68%)</td>
<td>–6%</td>
</tr>
<tr>
<td>Cohort</td>
<td>Rapid test</td>
<td>56% (53–58%)</td>
<td>–20%</td>
</tr>
<tr>
<td>Case-control</td>
<td>None</td>
<td>70% (68–72%)</td>
<td>0%</td>
</tr>
<tr>
<td>Case-control</td>
<td>RT-PCR</td>
<td>66% (63–68%)</td>
<td>–6%</td>
</tr>
<tr>
<td>Case-control</td>
<td>Rapid test</td>
<td>56% (53–58%)</td>
<td>–20%</td>
</tr>
<tr>
<td>Test-negative</td>
<td>None</td>
<td>70% (68–72%)</td>
<td>0%</td>
</tr>
<tr>
<td>Test-negative</td>
<td>RT-PCR</td>
<td>65% (63–68%)</td>
<td>–7%</td>
</tr>
<tr>
<td>Test-negative</td>
<td>Rapid test</td>
<td>53% (50–56%)</td>
<td>–24%</td>
</tr>
</tbody>
</table>

Table 3
Estimated vaccine effectiveness (VE) and bias from 1000 “all ages” simulations under different study designs and levels of misclassification, assuming 50% VE.

<table>
<thead>
<tr>
<th>Design</th>
<th>Misclassification</th>
<th>Estimated VE</th>
<th>Percent bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>None</td>
<td>50% (45–55%)</td>
<td>0%</td>
</tr>
<tr>
<td>Cohort</td>
<td>RT-PCR</td>
<td>47% (42–52%)</td>
<td>–6%</td>
</tr>
<tr>
<td>Cohort</td>
<td>Rapid test</td>
<td>40% (34–45%)</td>
<td>–20%</td>
</tr>
<tr>
<td>Case-control</td>
<td>None</td>
<td>50% (44–55%)</td>
<td>0%</td>
</tr>
<tr>
<td>Case-control</td>
<td>RT-PCR</td>
<td>47% (40–52%)</td>
<td>–6%</td>
</tr>
<tr>
<td>Case-control</td>
<td>Rapid test</td>
<td>40% (33–46%)</td>
<td>–20%</td>
</tr>
<tr>
<td>Test-negative</td>
<td>None</td>
<td>50% (44–55%)</td>
<td>0%</td>
</tr>
<tr>
<td>Test-negative</td>
<td>RT-PCR</td>
<td>46% (39–52%)</td>
<td>–8%</td>
</tr>
<tr>
<td>Test-negative</td>
<td>Rapid test</td>
<td>37% (29–43%)</td>
<td>–26%</td>
</tr>
</tbody>
</table>

at the level of RT-PCR tests, all three designs slightly underestimated VE (mean bias = –6% for cohort and case-control and –7% for test negative). Underestimation of VE was more pronounced in the presence of rapid test misclassification, and this underestimation was greater in the test-negative design (mean bias = –26%) compared with the case-control and cohort designs (mean bias = –20% for each).

Similar results were seen in the “all ages” scenario where true VE was 50%, with mean VE across the simulations of 50% in the absence of outcome misclassification (Table 3). Underestimation of VE was slight in the presence of RT-PCR misclassification (–6% for cohort and case-control, –8% for test-negative) and more pronounced in the presence of rapid test misclassification (–20% for cohort and case-control and –20% for test-negative).

In the test-negative design, low specificity caused greater bias in VE estimates than did low sensitivity. Assuming true VE of 50%, in simulations holding specificity constant at 1.0, VE_{ln} ranged from 50% when sensitivity was 1.0 to 48% when sensitivity was 0.8 (Fig. 1). In simulations holding sensitivity constant at 1.0, VE_{ln} ranged from 50% when specificity was 1.0 to 36% when specificity was 0.8.

When using RT-PCR for diagnosing influenza, the difference in mean VE between cohort and test-negative designs was trivial across a wide range of VE values (Fig. 2); VE from the test-negative design tended to be one percentage point lower than VE from the cohort design.

4. Discussion

Although first introduced in 1980 [13], the test-negative design was little used until it began to be applied to influenza vaccine studies less than 10 years ago [14]. Much remains unknown about the relative impact of potential biases on VE estimates from test-negative studies. One source of bias is misclassification of disease status due to the imperfect sensitivity and specificity of influenza tests. 

A priori, we would expect disease misclassification to cause greater bias in a test-negative study than in a cohort or a case-control study. In the cohort and case-control designs, persons falsely testing positive only contribute to the study in proportion to the rate of false positive tests in the entire population, which is a product of test sensitivity and incidence of non-influenza ARI. In contrast, in the test-negative design, persons falsely testing negative contribute to the study in proportion to the rate of false positive tests among persons with non-influenza ARI. Our findings confirm this expectation: in the presence of disease misclassification, VE estimates from the test-negative design are more biased than VE estimates from case-control or cohort designs. However, when highly sensitive and specific tests such as RT-PCR are used, the difference in bias between these designs is small and does not lead to meaningful differences in estimated VE between study designs.

We found that imperfect specificity of influenza testing caused greater bias in the test-negative design than did imperfect sensitivity. This finding agrees with the expected effects of imperfect sensitivity and specificity of traditional case-control studies [15]. In selecting influenza tests for test-negative studies, high specificity is much more important than high sensitivity for avoiding bias. For individual test results that are inconclusive (such as RT-PCR results with high cycle times), erring on the side of considering these to be negative results is expected to cause less bias than considering these to be positive.

Multiple influenza types or subtypes can co-circulate in a given season, and therefore people may remain at risk for a second influenza infection within a single season even after being infected. In such settings, an individual who was previously infected with influenza is still at risk for influenza and should still be eligible to be sampled into the control group of a case-control study, including the test-negative group in a test-negative case-control study. Sampling subjects into the control group if they have previously had influenza may bias VE estimates if sampling of such subjects is associated with vaccine status. Our primary simulation analysis
assumed that only one influenza infection occurred, and did not address this potential bias.

Although most commonly used for influenza VE studies, the test-negative design has also been used for rotavirus VE studies, with subjects being children with diarrheal illness (e.g., [16, 17]). Enzyme immunoassay tests for rotavirus have sensitivities of 75–80% and specificities of 98% or more [18, 19]. Rotavirus test-negative studies have found positive rotavirus tests for 11–17% of subjects tested [16,17]. Given the high proportion of subjects with diarrhea due to non-rotavirus causes, test specificity will be more important than sensitivity for causing biased rotavirus VE estimates. Given the high specificity of rotavirus tests, our results suggest that rotavirus VE estimates from test-negative designs will also not be meaningfully biased relative to VE estimates from other designs in the presence of imperfect diagnostic tests.

Several limitations of our study are worth considering. We assumed that misclassification was non-differential by vaccination status, i.e., that the accuracy of influenza test results was not affected by vaccination. Serologic testing for influenza infection has been shown to have reduced sensitivity in vaccinated persons, but data are less clear for RT-PCR [20]. Our results will not hold true if sensitivity or specificity is differential by vaccination status. We also assumed that influenza test results (either correct or misclassified) are known for all persons with ARI, and that patients with prior influenza infection can be excluded from the study population. In practice, potential study subjects may have had prior ARI that was not tested for influenza, and so persons with prior influenza infection may continue to contribute person time (in a cohort study) or be included as controls (in a case-control or test-negative study). The degree to which this phenomenon may bias VE results is uncertain, and the issue is further complicated by the possibility that persons may be infected with influenza multiple times during a single season (from infection with different virus types or subtypes). Finally, we did not examine the effect of other sources of bias, including selection bias, inaccurate exposure measurement, or confounding.

5. Conclusions

Influenza VE estimates from test-negative studies are biased in the presence of outcome misclassification, and this bias is greater than in corresponding VE estimates from cohort or case-control studies. However, the relative increase in bias is slight when using RT-PCR for diagnosing influenza. Given that observational studies of influenza VE can be dramatically confounded by differences in care-seeking and disease risk between vaccinated and unvaccinated persons [21,22], the gains in control of confounding by use of the test-negative design are likely to outweigh the bias due to misclassification.

Conflicts of interest statement

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by grant U01 IP000466 from the Centers for Disease Control and Prevention.

References