Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2007; all rights reserved. Advance Access publication 2 April 2007

International Journal of Epidemiology 2007;**36**:623–631
doi:10.1093/ije/dym021

THEORY AND METHODS

Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness

Evan W Orenstein,[1] Gaston De Serres,[2]* Michael J Haber,[3] David K Shay,[4] Carolyn B Bridges,[4] Paul Gargiullo[5] and Walter A Orenstein[6]

Accepted 30 January 2007

Background Influenza causes substantial morbidity and annual vaccination is the most important prevention strategy. Accurately measuring vaccine effectiveness (VE) is difficult. The clinical syndrome most closely associated with influenza virus infection, influenza-like illness (ILI), is not specific. In addition, laboratory confirmation is infrequently done, and available rapid diagnostic tests are imperfect. The objective of this study was to estimate the joint impact of rapid diagnostic test sensitivity and specificity on VE for three types of study designs: a cohort study, a traditional case-control study, and a case-control study that used as controls individuals with ILI who tested negative for influenza virus infection.

Methods We developed a mathematical model with five input parameters: true VE, attack rates (ARs) of influenza-ILI and non-influenza-ILI and the sensitivity and specificity of the diagnostic test.

Results With imperfect specificity, estimates from all three designs tended to underestimate true VE, but were similar except if fairly extreme inputs were used. Only if test specificity was 95% or more or if influenza attack rates doubled that of background illness did the case-control method slightly overestimate VE. The case-control method usually produced the highest and most accurate estimates, followed by the test-negative design. The bias toward underestimating true VE introduced by low test specificity increased as the AR of influenza- relative to non-influenza-ILI decreases and, to a lesser degree, with lower test sensitivity.

Conclusions Demonstration of a high influenza VE using tests with imperfect sensitivity and specificity should provide reassurance that the program has been effective in reducing influenza illnesses, assuming adequate control of confounding factors.

Introduction

Influenza viruses are constantly evolving antigenically, and vaccine strains must be selected annually based on global surveillance of circulating viral strains.[1] Therefore, past evidence of influenza vaccine effectiveness (VE) does not necessarily predict VE for a subsequent season.[2–4] Methods to permit rapid, simple and affordable assessments of influenza VE would permit annual measures of the impact of current vaccination strategies.[5] Laboratory confirmation of cases is critical when trying to assess the true effectiveness of influenza vaccine against influenza because non-specific outcomes bias VE estimates toward the null.[6] The gold standard for diagnosis

[1] Yale University, New Haven, CT, USA.

[2] Quebec National Institute of Public Health and Laval University, Quebec, Canada.

[3] Rollins School of Public Health, Emory University.

[4] Influenza Division, National Center for Immunization and Respiratory Diseases (proposed), Centers for Disease Control and Prevention.

[5] Epidemiology and Surveillance Division, National Immunization Program, Centers for Disease Control and Prevention.

[6] School of Medicine, Emory University, Atlanta, GA, USA.

*Corresponding author. Quebec National Institute of Public Health, 2400 d'estimauville, Quebec, PQ, Canada G1E 7G9.
E-mail: gaston.deserres@ssss.gouv.qc.ca

of influenza infection is viral culture and, increasingly, reverse transcription polymerase reaction (RT-PCR).[1,7,8] However, culture methods are expensive, results take 2 or more days and thus cannot guide patient care, and are not perfectly sensitive. RT-PCR is more sensitive, but is also more expensive and not widely available.[8] A number of easy-to-use, rapid diagnostic tests have become available in recent years and provide results in <30 min, although their sensitivities and specificities are lower than viral culture or PCR.[1,9–14] Given their lower cost and rapid results, use of these tests is increasing. The bias that might be associated with the use of tests with imperfect sensitivity or specificity to define outcomes in VE studies has not been evaluated carefully.

The objective of this study was to estimate the joint impact of test sensitivity and specificity on VE for three types of study designs: a cohort study, a traditional case-control study and a case-control study that used individuals with influenza-like illness (ILI) who tested negative for influenza virus infection as controls.

Methods

To focus the simulations on the effects of test sensitivity and specificity, we assumed that all other factors which could bias these estimates (e.g. differential exposure of vaccinees and non-vaccinees to the virus, differential access to care and diagnostic evaluations or differential rates of vaccination due to unmeasured factors) were equally distributed among vaccinated and unvaccinated individuals and that sampling was representative of the population with respect to vaccine coverage, influenza-ILI and non-influenza-ILI.

Five parameters were included in all simulation models:

VE_{true} = true vaccine effectiveness
AR_{flu} = true attack rate of medically attended influenza among unvaccinated
AR_{nonflu} = true attack rate of medically attended non-influenza ILI
sens = sensitivity of rapid influenza test
spec = specificity of the influenza test

For our base-case assumptions, we used data collected from young children aged 6–24 months because the ACIP recommendation for universal vaccination in this population[20] has rendered placebo-controlled trials in young children unethical. Therefore, VE data for young children must be obtained through observational studies, and VE is likely to be assessed during upcoming influenza seasons because of the increasingly recognized health burden of influenza in this age group.[15,16]

We assumed a true VE of 70%, based on the demonstration of a vaccine efficacy of 66% in year 1 of a placebo-controlled randomized study of trivalent inactivated influenza vaccine among children aged 6–24 months.[17] The sensitivity of the rapid diagnostic test was set at 80%, approximating the median sensitivity (79%) found among 35 studies of these tests among children.[18] It was assumed there was no differential test sensitivity between vaccinees and non-vaccinees. The specificity of the tests was set at 90%, again approximately the median (91%) found in 33 studies conducted among children.[18]

The true attack rate (AR) of influenza in the unvaccinated was set at 15%, representing the AR over an entire 6–8 week influenza season, consistent with an AR of 16% among placebo recipients 6–24 months of age in one study[17] and an AR of 18% among a cohort of 2–24-month old Finnish children.[19] The true AR of non-influenza-ILI was set at 30%, similar to the rates of background illness seen during a respiratory season in a randomized placebo-controlled trial among young children (30%),[17] and among a cohort of young children in Finland (32%).[19]

The case definition of ILI utilized for influenza surveillance by the Centers for Disease Control (CDC) and Prevention includes a fever ≥100°F and a cough or sore throat.[20,21] It was assumed that non-influenza- and influenza-ILIs were independent of each another (i.e. during an influenza season, an individual who developed influenza was equally likely compared with individuals without influenza to develop a non-influenza-ILI).

Estimation of VE using three observational study designs

Figure 1 demonstrates what proportion of the population with ILI tested positive and negative for influenza. No factors affected the probability of contracting a non-influenza-ILI; only vaccination affected the risk of contracting an influenza-ILI in this theoretical example. We assumed that each ill individual contracted influenza only, a non-influenza-ILI only, or a case of each illness (i.e. two episodes). The figure describes the proportion of the population in each category, and the multiplicative factor used to move down each pathway leads to the population testing positive or negative. For example, $(AR_{flu})^*(AR_{nonflu})$ defined the proportion of the unvaccinated population which had both an influenza- and a non-influenza-ILI. Because members of this group had influenza, the sensitivity determined what fraction tested positive. While the test initially misclassified the remainder (1 − sensitivity), these individuals also had a non-influenza-ILI. The test for this second episode yielded a fraction test-positive determined by (1 − specificity).

Cohort design

VE in the cohort design was calculated as 1 minus the relative risk $(1-(AR_{vac}/AR_{unvac}))$ where AR_{vac} and AR_{unvac} represented the observed ARs of influenza in the vaccinated and unvaccinated populations, respectively. The observed ARs were expressed as functions of test sensitivity, which determined the proportion of true influenza cases detected, and specificity, which described the proportion of true negatives that test negative.

True influenza cases were detected at a rate determined by the sensitivity of the test, $(sens*AR_{flu})$. Some non-influenza-ILI cases were misclassified as influenza, because the test was not 100% specific. $AR_{nonflu}*[1-(sens*AR_{flu})]$ was the proportion of the population with ILI minus those true flu cases detected based on test sensitivity. 1 minus the specificity determined the fraction of this group who were test positives: $(1-spec)*AR_{nonflu}*[1-(sens*AR_{flu})]$.

Calculated AR in Unvaccinated Population (CAUP):

$$sens * AR_{flu} + (1 - spec) * AR_{nonflu} * [1 - (sens * AR_{flu})]$$

The calculated AR in the vaccinated population was defined similarly, except that the true AR of influenza in the vaccinated was the original AR of influenza times 1 minus the true VE: $(AR_{flu}*(1-VE_{true}))$.

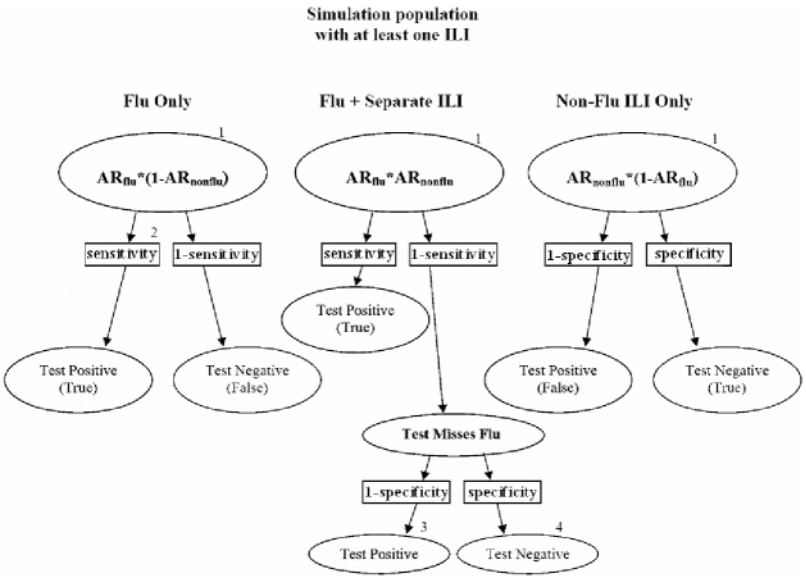

Figure 1 Flowchart to determine the proportion of the unvaccinated simulation population that tests positive and negative for influenza. The three mutually exclusive groups (influenza, influenza plus non-influenza-ILI or non-influenza-ILI only) assume that each person could contract one episode of influenza, one episode of non-flu ILI, or one episode of each during the study period. Each individual received one test for each episode. Once a person tested positive for influenza, that person would always be classified as test-positive regardless of future episodes. If an individual contracted both influenza and non-influenza-ILI at the same time, it was assumed that the test would behave as if the individual only had influenza and the non-influenza-ILI was not detected as part of the AR_{nonflu}. For the corresponding values in the vaccinated population, replace the value 'ARflu' with 'ARflu*(1-VE)'

Calculated AR in Vaccinated Population (CAVP):

$$\text{sens} * \text{AR}_{\text{flu}} * (1 - \text{VE}_{\text{true}}) + (1 - \text{spec}) * \text{AR}_{\text{nonflu}} * \left\{1 - [\text{sens} * \text{AR}_{\text{flu}} * (1 - \text{VE}_{\text{true}})]\right\}$$

The cohort estimate of VE was calculated as 1 minus the calculated AR in the vaccinated divided by the calculated AR in the unvaccinated.

Cohort Estimate of Vaccine Effectiveness:

$$\text{cohorteff} = 1 - \frac{\text{CAVP}}{\text{CAUP}}$$

Case-control design

VE in the case-control design was 1 minus the odds ratio, which was calculated as the number of vaccinated cases times the number of unvaccinated controls divided by the number of unvaccinated cases times the number of vaccinated controls. We allowed any individual who was not influenza test-positive to be a control. Because we assumed that sampling would reflect the actual proportions in the population, we calculated the proportion of controls as the total proportion of the population that did not test positive for influenza. We used 1 minus the calculated AR of influenza in each population as the proportion of non-cases to be used as controls.

Case-Control Estimate of VE:

$$\text{ccontroleff} = 1 - \frac{\text{CAVP} * (1 - \text{CAUP})}{\text{CAUP} * (1 - \text{CAVP})}$$

Case-control design, using those who test negative for influenza (test-negatives) as controls

VE in the test-negative design was defined similarly as for the case-control design, except that $(1 - \text{CAUP})$ and $(1 - \text{CAVP})$ were replaced with the proportions of test negatives in each population.

Using Figure 1, we determined that among the population with individuals contracting both influenza- and non-influenza-ILIs, a percentage determined by [sensitivity + (1 − sensitivity)*(1 − specificity)] tested positive, and another percentage [(1 − sensitivity)*specificity] tested negative.

Thus, the proportions of the vaccinated and unvaccinated populations that tested negative were calculated as follows:

Proportion of Unvaccinated Population that Tests Negative (PUPTN):

$$\text{AR}_{\text{flu}}(1 - \text{sens})\,(\text{AR}_{\text{nonflu}} * \text{spec} - \text{AR}_{\text{nonflu}} + 1) + \text{AR}_{\text{nonflu}}(1 - \text{AR}_{\text{flu}})\,\text{spec}$$

Proportion of Vaccinated Population that Tests Negative (PVPTN):

$$\text{AR}_{\text{flu}}(1 - \text{VE}_{\text{true}})\,(1 - \text{sens})(\text{AR}_{\text{nonflu}} * \text{spec} - \text{AR}_{\text{nonflu}} + 1) + \text{AR}_{\text{nonflu}}[1 - \text{AR}_{\text{flu}}(1 - \text{VE}_{\text{true}})] * \text{spec}$$

Test Negative Estimate of VE:

$$\text{testnegeff} = 1 - \frac{\text{CAVP} * \text{PUPTN}}{\text{CAUP} * \text{PVPTN}}$$

The main outcomes were differences between true and estimated VE using each of the three study designs. In addition to determining the magnitude of this error, we performed a slope approximation analysis for each input variable under different conditions by calculating VE after changing the value of one variable by a 1% absolute value. We then subtracted the new estimate from the original estimate and divided the difference by 1%. The greater the absolute value of the approximated slope, the more sensitive the estimate was to changes in that variable. If the sign was negative, then increasing the value of that variable

decreased the error; if the sign was positive, increasing the value of that variable increased the error.

Results

Using the base-case assumptions, including a true VE of 70%, all influenza VE estimates were lower than the true VE, ranging from 56% in the cohort design to 60% in the case-control design (Table 1). In general, the higher the true VE, the greater the absolute differences between true and estimated VE (Table 1). Relative errors for the cohort and the test-negative methods were within ~20% of the true VE, no matter the value used to represent the true VE. However, the relative error between the case-control VE estimate and the true VE decreased with decreasing true VE.

The case-control method consistently produced a VE estimate that was closer to the true VE than either of the other two methods. For example, if true VE was 90%, the case-control method yielded an estimate of 75%, 15% lower than the true effectiveness. In contrast, the cohort and test-negative methods produced estimates 18 and 17% lower than the true effectiveness, respectively. When true VE was 30%, the estimated effectiveness from the case-control method was only 3.2% lower than the true effectiveness, while the cohort and test-negative methods were 6.1 and 5.2% lower. Overestimation by any method was rare. Even with 100% specificity and an influenza to non-influenza attack rate ratio as high as 2–3 : 1, which might be seen during the peak weeks of an influenza epidemic,[22] the case-control method only overestimated effectiveness by 7% when true VE was 70%. The test-negative estimate never exceeded that of the case-control method and the cohort method never overestimated true VE.

We examined the biases generated by using each study method under different assumptions. We plotted test specificity and sensitivity as they varied from 70 through 100% against the estimated VE, while varying the ARs of influenza- and non-influenza-ILIs, for a range of true VE from 25 through 70%. For each method, at both high and low true VE, the absolute values of the ARs of influenza- and non-influenza-ILIs were less important than their ratio. For example, if the base-case ARs of 15% for influenza-ILI and 30% for non-influenza-ILI were changed to 1.5 and 3%, respectively, or to 40 and 80%, the cohort study VE estimate remained between 55 and 56%.

Table 1 Differences between true vaccine effectiveness and calculated vaccine effectiveness by using three observational study methods as true vaccine effectiveness varies

True vaccine effectiveness (%)	Calculated vaccine effectiveness		
	Cohort (%)	Case-control (%)	Test negative (%)
90	71.6	74.7	72.6
70[a]	55.7	59.5	57.0
50	39.8	43.6	41.1
30	23.9	26.8	24.8
10	8.0	9.2	8.4
5	4.0	4.6	4.2

[a]Base-case assumptions: $VE_{true} = 70\%$, $AR_{flu} = 15\%$, $AR_{nonflu} = 30\%$, sensitivity = 80%, specificity = 90%.

We graphed test specificity and sensitivity against VE estimates from each study design by using three sets of ARs for influenza-and non-influenza-ILIs: 30 and 30% (ratio 1 : 1), 15 and 30% (ratio 1 : 2), 5 and 30% (ratio 1 : 6). At 25% true VE, all three methods tended to bias the effectiveness estimates towards the null (Figure 2). A lower ratio for the AR of influenza-ILI to the AR of non-influenza-ILI, combined with lower test specificity, tended to increase the error between the estimated and the true VE. Using 70% sensitivity and specificity, with a 1 : 1 ratio of ARs of influenza- to non-influenza-ILIs, the cohort design estimated VE at 17%, an error of 8%. Changing the ratio of AR to 1 : 6 yielded an estimated effectiveness of 6.5%, or 18.5% from the true value. The case-control method yielded the highest estimates in all situations; however, the estimated VE was never more than 5% greater than VE. The cohort and test-negative designs gave comparable effectiveness estimates in most cases. Low test specificity played a more important role than low sensitivity in determining these errors. The test-negative design produced higher estimates than the cohort design at high sensitivity, while the reverse was true at low sensitivity.

When using a true VE of 70%, similar trends were documented. Most estimates were lower than the true VE because of biases introduced by using an imperfect diagnostic test (Figure 3). However, while relative errors in estimates were similar for given conditions with either a low or a high true VE, the absolute error was much larger when the true VE was 70%. The greatest errors occurred when both specificity and the ratio of ARs of influenza to non-influenza-ILI were low. When specificity and sensitivity were set to 70%, and the ratio of ARs was 1 : 1, the cohort design estimated effectiveness at 48%, an error of 22%. Using a ratio of ARs of 1 : 6, the effectiveness estimate was 18%, a 52% error. The case-control method yielded higher estimates than the other methods, but never overestimated the true VE by more than 7%. The cohort and test-negative designs again produced comparable estimates; however, with high test sensitivity the test-negative method estimate exceeded the cohort estimate, and at low sensitivity the cohort estimate exceeded the test-negative estimate. In general, test specificity was more important than sensitivity in producing these errors.

If specificity was 100%, the cohort method VE estimate was equivalent to the true VE, regardless of test sensitivity or ARs of influenza- and non-influenza-ILIs. However, as specificity decreased, the cohort method began to underestimate the true VE. This error increased as the test sensitivity and the ratio of ARs for influenza- to non-influenza-ILIs were decreased. With the other two methods, there were minor differences between true and estimated VE at 100% specificity.

The slope approximation analysis was performed using values of true VE of 70 and 25%, and three different ratios of ARs for influenza- and non-influenza-ILIs (Table 2). Using the base-case assumptions (VE = 70%, influenza-ILI AR = 15%, non-influenza-ILI AR = 30%), specificity was the most influential variable on the VE estimates. At a true VE of 70% and using an AR ratio of 1 : 6, the influenza AR produced the largest slope approximation. A 1% increase in the influenza AR decreased the AR ratio from 1 : 6 to 1 : 5. At 25% true VE and a 1 : 1 AR ratio, specificity and the

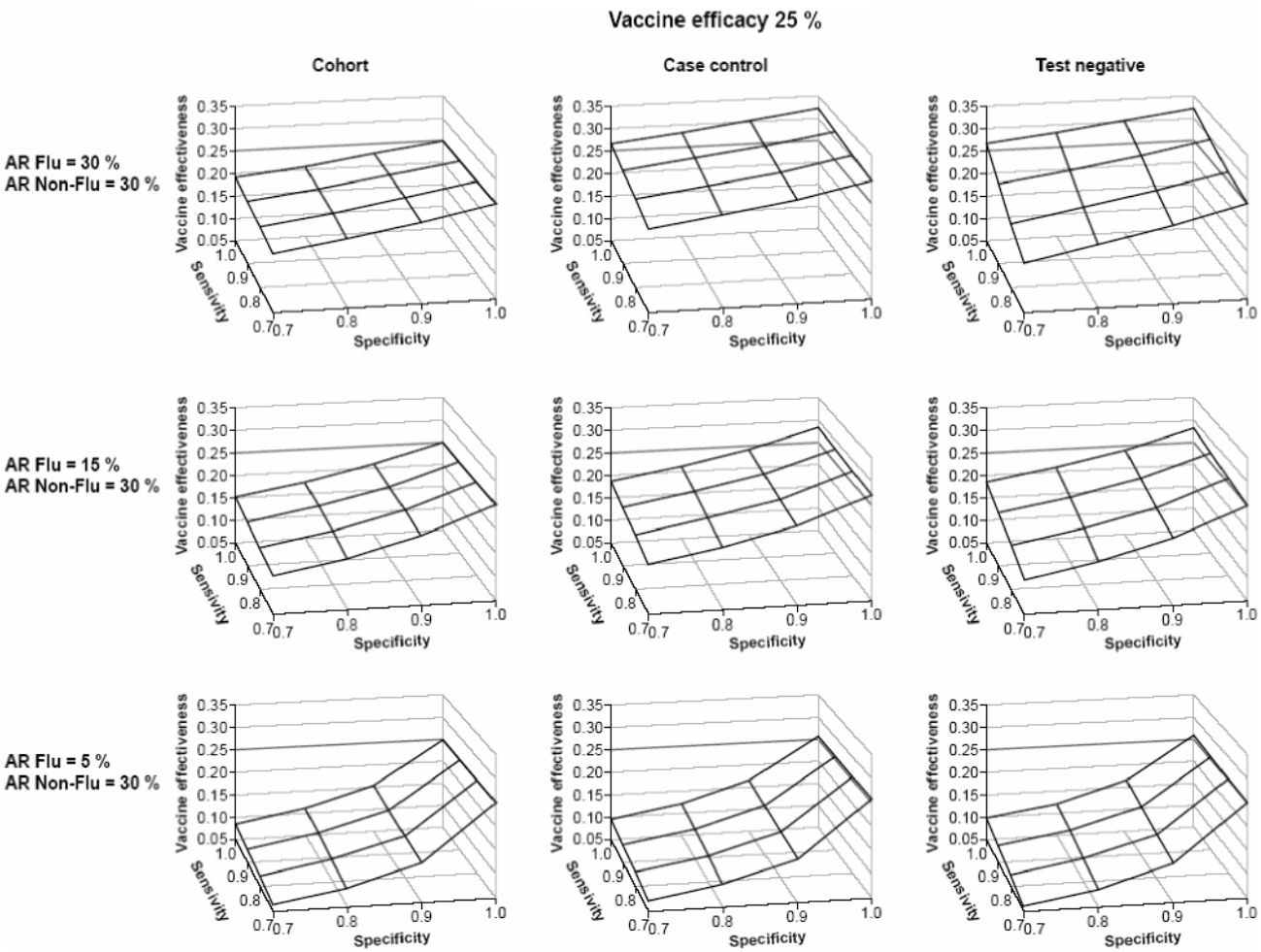

Figure 2 Vaccine effectiveness estimates by three methods as specificity and sensitivity vary. Various conditions with a true VE of 25%

influenza AR were the two most influential variables. However, in the test-negative design, the slope approximation analysis demonstrated that sensitivity was nearly as important as specificity. Under all assumptions, the slope approximation for sensitivity was highest in the test-negative design.

Discussion

In these simulations, the use of an imperfect test to diagnose influenza infection biased VE estimates from all three study designs towards the null. Although the use of a case-control study design generally provided the most accurate results, the three designs provided similar VE estimates in the base case and most scenarios. The major determinants of bias were test specificity and the ratio of the AR of influenza and non-influenza-ILIs. With a relatively low ratio of influenza compared with non-influenza illness (e.g. a ratio of 1 : 2), even using a diagnostic test that is 90% specific results in a large number of non-influenza-ILI cases being attributed to influenza, and in equal proportions among the vaccinated and unvaccinated populations. These false positives lower the VE estimates from all three study types. Use of even less specific outcomes, such as clinical ILI, would generate even less accurate VE estimates. We have not estimated the quantitative consequences of using clinical ILI as an influenza 'test'. However, other studies have demonstrated that when influenza cases are defined using only less specific clinical definitions, the calculated health benefits of vaccination are lower.[2,23,24]

In most of our simulations, the three methods produced comparable results within 10% of each other, suggesting that all three methods perform well. In particular, this analysis provides evidence that using test-negatives as controls is a valid design. While the differences were small, the case-control method produced the highest and the most accurate VE estimates. Because influenza is not a rare disease (e.g. the AR in young children may be >10% during a single season) and vaccination decreases the risk of disease, the odds ratio (OR) underestimates the relative risk (RR). Increasing the AR leads to a greater decrease in the proportion of controls in the unvaccinated population relative to the decrease in the proportion of controls in the vaccinated population. Thus, the case-control method generates higher VE estimates than the cohort method. The bias introduced from a case-control design generally opposes the bias introduced by a diagnostic test with <100% specificity and produces more accurate overall VE estimates.

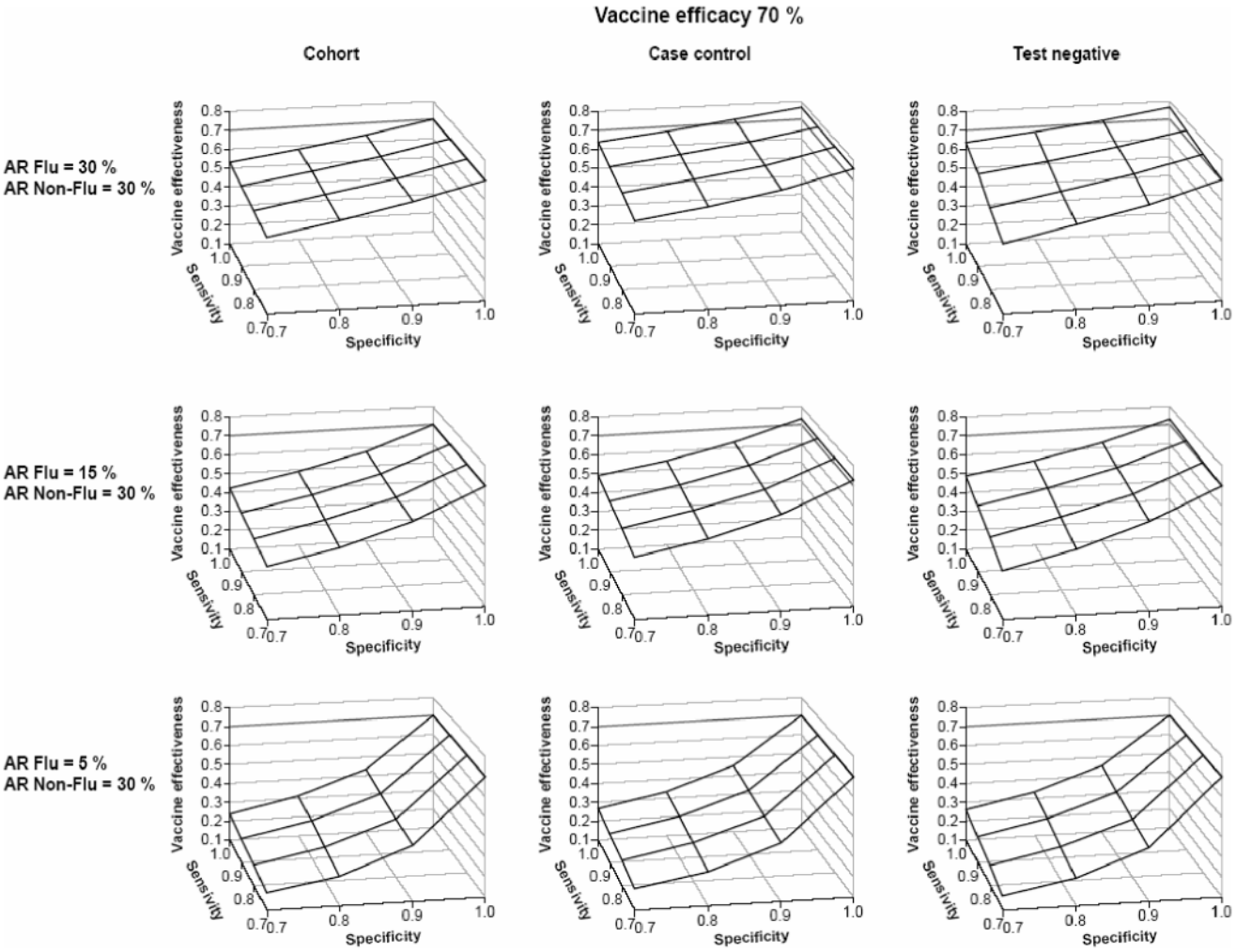

Figure 3 Vaccine effectiveness estimates by three methods as specificity and sensitivity vary. Various conditions with true VE of 70%.

Only in rare situations did the case-control estimate exceed the true VE: even with assumptions of a 3 : 1 ratio of attack rates and 100% test specificity, VE was overestimated by less than 7%. The test-negative method tends to produce VE estimates slightly higher than those made by using the cohort method because the OR underestimates RR, and slightly lower than those from the case-control method because imperfect sensitivity enriches the unvaccinated population with false negatives relative to the vaccinated population. Using realistic assumptions, the test-negative design provides relatively accurate VE estimates. This design may be easier to implement than the other two for VE studies, particularly among young children with high ILI attack rates.

Previous discussions of vaccine effectiveness have stated that in cohort studies specificity, and not sensitivity, is the important determinant of accurate VE estimates.[25] However, this is correct only when specificity is 100%. When specificity is <100%, false positive cases distribute themselves equally among vaccinees and non-vaccinees, lowering the VE estimate. The error is compounded by low sensitivity because true cases, which would tend to be distributed more in the unvaccinated population, are not detected and do not counterbalance the false positives.

This study has several limitations. We did not address issues of bias introduced by differences in exposure, access to care, healthcare seeking behaviours, ability to mount a protective response after vaccination, decisions to obtain diagnostic tests or other factors that may relate both to being vaccinated and to being diagnosed with an influenza infection. Thus, simulations with the cohort design describe results that would be obtained in a perfectly randomized placebo-controlled clinical trial. In contrast, actual observational studies of influenza VE must attempt to adjust for such potential differences between vaccinated and unvaccinated individuals. This study evaluated only the effects of systematic and not sampling errors. We did not address the issue of chance variation as illustrated, for example, by wide confidence intervals. Confidence intervals depend on the statistical power of the study, determined by influenza attack rates and the sample size, which we did not examine.

Our analyses suggest that use of an imperfect test to diagnose influenza infection in studies of influenza VE among

Table 2 Slope approximation analysis[a] of the equation defining the difference between true vaccine efficacy and calculated vaccine effectiveness with respect to each input variable by three methods with varying true vaccine efficacy and ratio of attack rates

	Cohort	Case-control	Test negative
[b]VE_{true} = 70%, ARflu = 30%, ARnonflu = 30%, Sens = 80%, Spec = 90%			
VE_{true}	0.11	0.19	0.14
ARflu	−0.24	−0.45	−0.31
ARnonflu	−0.09	−0.17	−0.39
Sens	−0.72	−0.59	−0.73
Spec	0.24	0.20	0.20
VE_{true} = 70%, ARflu = 15%, ARnonflu = 30%, Sens = 80%, Spec = 90%			
VE_{true}	0.20	0.22	0.21
ARflu	−0.80	−1.00	−0.87
ARnonflu	0.39	0.36	0.26
Sens	−0.18	−0.22	−0.37
Spec	−1.15	−1.05	−1.15
VE_{true} = 70%, ARflu = 5%, ARnonflu = 30%, Sens = 80%, Spec = 90%			
VE_{true}	0.44	0.43	0.43
ARflu	−3.88	−4.14	−3.97
ARnonflu	0.60	0.58	0.54
Sens	−0.22	−0.23	−0.29
Spec	−1.70	−1.66	−1.70
VE_{true} = 25%, ARflu = 30%, ARnonflu = 30%, Sens = 80%, Spec = 90%			
VE_{true}	0.11	−0.04	0.06
ARflu	−0.09	−0.31	−0.15
ARnonflu	0.09	0.08	−0.08
Sens	−0.03	−0.12	−0.27
Spec	−0.26	−0.24	−0.29
VE_{true} = 25%, ARflu = 15%, ARnonflu = 30%, Sens = 80%, Spec = 90%			
VE_{true}	0.20	0.13	0.18
ARflu	−0.29	−0.47	−0.34
ARnonflu	−0.05	−0.09	−0.17
Sens	−0.41	−0.40	−0.43
Spec	0.14	0.14	0.06
VE_{true} = 25%, ARflu = 5%, ARnonflu = 30%, Sens = 80%, Spec = 90%			
VE_{true}	0.44	0.41	0.43
ARflu	−1.39	−1.56	−1.44
ARnonflu	−0.08	−0.09	−0.12
Sens	−0.61	−0.60	−0.61
Spec	0.21	0.21	0.19

[a]The values given approximate the slope of the difference $VE_{true} - VE_{estimate}$ in the direction specified by each variable. The approximation is determined by changing the specified variable by 1% (absolute), subtracting the new estimates' error from the original error and dividing by 0.01. A negative sign indicates that increasing the corresponding variable decreases the difference between actual vaccine effectiveness and calculated vaccine effectiveness using the corresponding method.

[b]VE_{true} = true vaccine efficacy; ARflu = attack rate of influenza; ARnonflu = attack rate of non-influnza ILIs; sens = test sensitivity; spec = test specificity.

young children, in whom the ARs of influenza- and non-influenza-ILIs are high, produce estimates of VE likely substantially lower than the 'true' values. These results can be extrapolated to other age groups, as the same methodological principles apply and ARs for both influenza and non-influenza respiratory illnesses decrease with increasing age. Thus, obtaining a VE estimate of 60–70% in an observational study that used tests with imperfect sensitivity and specificity should provide reassurance that the program has been effective in reducing influenza illnesses, assuming adequate control of possible

confounding factors. It may be reasonable to use rapid diagnostic tests as a screen in VE studies and to bank respiratory specimens for more specific testing (including viral culture or PCR) if a low VE is calculated. Field investigators may be able to approximate the magnitude and direction of VE errors by determining the ratio of the ARs of influenza- and non-influenza-ILIs as well as the specificity and sensitivity of the diagnostic test for influenza that is being considered.[26] Low test specificity and a low ratio of ARs would suggest that the observed effectiveness represents a lower bound for the actual VE. Further research is needed to determine if the biases demonstrated in this simulation study could be adjusted for in the analysis stage of an actual field assessment of influenza VE.

Declaration of interest

I receive research funding from Novartis to study ways to improve influenza vaccine coverage among healthcare workers.

I receive research funding from Sanofi Pasteur to study the impact of simultaneous immune globulin on viremia following yellow fever vaccination.

I am on a data and safety monitoring board for GlaxoSmithKline evaluating pneumococcal conjugate vaccine.

Walter A. Orenstein, Director, Emory Vaccine Policy and Development, Associate Director, Emory Vaccine Center

Acknowledgements

Dr Walter Orenstein's participation was supported in part by Grant Number 1P20RR020735 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

KEY MESSAGES

- With imperfect specificity, VE is generally underestimated with all three designs.
- Case-control followed by the test-negative design are those giving the highest and most accurate estimates.
- The bias toward underestimating true VE introduced by low test specificity increases as the AR of influenza- relative to non-influenza-ILI decreases and, to a lesser degree, with lower test sensitivity.

References

[1] Fukuda K, Levandowski RA, Bridges CB, Cox NJ. Inactivated Influenza Vaccines. In: Plotkin S, Orenstein W, (eds). *Vaccines*. Philadelphia, PA: Elsevier; 2004. p. 339–70.

[2] Beyer WEP. Heterogeneity of case definitions used in vaccine effectiveness studies – and its impact on meta-analysis. *Vaccine* 2006;**24**:6602–4.

[3] Carrat F, Tachet A, Rouzioux C, Housset B, Valleron AJ. Evaluation of clinical case definitions of influenza: detailed investigation of patients during the 1995 – 1996 epidemic in France. *Clin Infect Dis* 1999;**28**:283–90.

[4] Monto AS, Ohmit SE. The evolving epidemiology of influenza infection and disease. In: Brown LE, Hampson AW, Webster RG, (eds). *Options for the Control of Influenza III,*. Amsterdam, The Netherlands: Elsevier Science BV; 1996. p. 45–49.

[5] Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney P. Does this patient have influenza? *JAMA* 2005;**293**:987–97.

[6] Scharfstein DO, Halloran ME, Chu H, Daniels MJ. On estimation of vaccine efficacy using validation samples with selection bias. *Biostatistics* 2006;**7**:615–29.

[7] Hedrick JA, Barzilai A, Behre U, *et al*. Zanamivir for treatment of symptomatic influenza A and B infection in children five to twelve years of age: a randomized controlled trial. *Pediatr Infect Dis J* 2000;**19**:410–17.

[8] Weinberg GA, Erdman DD, Edwards KM *et al*. New Vaccine Surveillance Network Study Group. Superiority of reverse-transcription polymerase chain reaction to conventional viral culture in the diagnosis of acute respiratory tract infections in children. *J Infect Dis* 2004;**189**:706–10.

[9] Poehling KA, Griffin MR, Dittus RS *et al*. Bedside diagnosis of influenza virus infections in hospitalized children. *Pediatrics* 2002;**110**:83–88.

[10] Esposito S, Marchisio P, Morelli P, Crovari P, Principi N. Effect of a rapid influenza diagnosis. *Arch Dis Child* 2003;**139**:321–29.

[11] Bellei N, Benfica D, Perosa AH, Carlucci R, Barros M, Granato C. Evaluation of a rapid test (QuickVue) compared with the shell vial assay for detection of influenza virus clearance after antiviral treatment. *J Virol Methods* 2003;**109**:85–88.

[12] Pregliasco F, Puzelli S, Mensi C *et al*. The Collaborative Group Influchild. Influenza virological surveillance in children: the use of the QuickVue rapid diagnostic test. *J Med Virol* 2004;**74**:1190–95.

[13] Rawlinson WD, Waliuzzaman ZM, Fennel M, Appleman JR, Shimasaki CD, Carter IW. New point of care test is highly specific but less sensitive for influenza virus A and B in children and adults. *J Med Virol* 2004;**74**:127–31.

[14] Poehling KA, Zhu Y, Tang YW, Edwards K. Accuracy and impact of a point-of-care rapid influenza test in young children with respiratory illnesses. *Arch Pediatr Adolesc Med* 2006;**160**:713–18.

[15] Jefferson T, Smith S, Demicheli V, Harnden A, Rivetti A, Di Pietrantonj C. Assessment of the efficacy and effectiveness of influenza vaccines in healthy children: systematic review. *Lancet* 2005;**365**:773–80.

[16] Poehling KA, Edwards KM, Weinberg GA *et al*. The underrecognized burden of influenza in young children. *N Engl J Med* 2006;**355**:31–40.

[17] Hoberman A, Greenberg DP, Paradise JL *et al*. Effectiveness of inactivated influenza vaccine in preventing acute otitis media in young children: a randomized controlled trial. *JAMA* 2003;**290**:1608–16.

[18] Uyeki TM. Influenza diagnosis and treatment in children: a review of studies on clinically useful tests and antiviral treatment for influenza. *Pediatr Infect Dis J* 2003;**22**:164–77.

[19] Vesa S, Kleemola M, Blomqvist S, Takala A, Kilpi T, Hovi T. Epidemiology of documented viral respiratory infections and acute otitis media in a cohort of children followed from two to twenty-four months of age. *Pediatr Infect Dis J* 2001;**20**:574–81.

[20] Centers for Disease Control and Prevention. Prevention and control of influenza: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR* 2006;**55**(No. RR-10):1–42.
[21] Rothberg MB, Fisher D, Kelly B, Rose DN. Management of influenza symptoms in healthy children. *Arch Pediatr Adolesc Med* 2005;**159**:1055–62.
[22] Meara MO, Brien AO, Feely E, Conlon M. Influenza A outbreak in a community hospital. *Ir Med J* 2006;**99**(6):175–7.
[23] Nichol KL, Mendelman P. Influence of clinical case definitions with differing levels of sensitivity and specificity on estimates of the relative and absolute health benefits of influenza vaccination among healthy working adults and implications for economic analyses. *Virus Res* 2004;**103**:3–8.
[24] Ohmit SE, Monto AS. Symptomatic predictors of influenza virus positivity in children during the influenza season. *Clin Infect Dis* 2006;**43**:564–8.
[25] Orenstein WA, Bernier RH, Hinman AR. Assessing vaccine efficacy in the field: further observations. *Epidemiol Rev* 1988;**10**:212–41.
[26] Halloran ME, Longini IM, Gaglani MJ *et al*. Estimating efficacy of trivalent, cold-adapted, influenza virus vaccine (CAIV-T) against influenza A (H1N1) and B using surveillance cultures. *Am J Epidemiol* 2003;**158**(4):305–11.

 International Journal of Epidemiology 2007;**36**:631–632
Advance Access publication 22 June 2007 doi:10.1093/ije/dym084

Commentary: Observational studies and the art of accurately measuring influenza vaccine benefits

Lone Simonsen

Accepted 22 March 2007

As Orenstein and colleagues point out in an elegant paper in this issue,[1] observational studies that compare the incidence of influenza-like illness in vaccinated and unvaccinated groups are theoretically prone to underestimating the vaccine's true effectiveness (VE). This is because many other respiratory pathogens cause similar symptoms; these other infections form a sizeable background of influenza-like illness cases in both case and control groups that are not preventable by influenza vaccination. Orenstein *et al*. reasonably call for laboratory confirmed endpoints, which have rarely been obtained in observational studies of influenza vaccine effectiveness in populations targeted for vaccination. These theoretical effects of low endpoint specificity are perhaps not so novel, but this paper and its careful quantification of the issue is timely because the problems of interpreting results from studies that use low-specificity end-points seem to have been all but forgotten in the contemporary literature.

In a perfect world, there would be plenty of 'Gold standard' evidence from randomized placebo-controlled clinical trials (RCTs) that measure vaccine efficacy using highly specific laboratory-confirmed influenza endpoints. But these are scarce in the influenza literature. A placebo control group is simply not an option when studying vaccine benefits in populations that are already recommended for vaccine (such as seniors and persons with high-risk conditions). For that reason, observational studies have long made up the largest part of the evidence base, especially for influenza vaccine benefits in seniors.

Assuming a near-perfect world, Orenstein *et al*. theoretically explore the expected performance of cohort and case-control study designs that use laboratory-confirmed endpoints, the latter with two different approaches to control selection. They focused on the consequences of less-than-perfect sensitivity and specificity of the rapid laboratory tests (an increasingly popular choice over culture-confirmation as the price of these kits falls) and on the prevalence of influenza relative to other respiratory pathogens. However, their simulations *did not* explore the possibility of selection bias leading to various degrees of mismeasurement. Orenstein *et al*. first explored a base-case scenario of a paediatric population, assuming realistic parameters of attack rates of influenza (15%) and other respiratory pathogens (30%), and rapid tests with 80% sensitivity and 90% specificity. In this scenario, all observational study designs performed about the same, although case-control studies using

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA.

Corresponding author. National Institute of Allergy and Infectious Diseases, National Institutes of Health, 6700B Rockledge Drive, room 1107, Bethesda 20892, Maryland, USA. E-mail: lsimonsen@niaid.nih.gov