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

Lone Simonsen

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...
non-lab-positive controls had a slight advantage. All underestimate the true VE by about 15%; for example, a true VE of 70% would be measured at 56–60%. But as can be seen from results of the sensitivity analyses (Figure 2 in ref.11), under a scenario in which the influenza attack rate is low relative to other respiratory illness (5% vs 30%), the VE would be substantially underestimated; for example, a VE of 70% would be measured at ~40% or lower.

So the good news from Orenstein et al. is that observational studies can generate reasonably accurate VE estimates, as long as laboratory tests are highly (>90%) specific and as long as the incidence of influenza is relatively high. But as the authors point out, the bad news is that the use of endpoints with low specificity—such as clinical influenza-like illness without laboratory confirmation—is not appropriate. The consequences can be seen by perusing their Figure 2. If the specificity of the endpoint is 70%, for example, the underestimation is profound (measured VE at 10%). With extrapolation of Orenstein’s findings to a lower level of sensitivity of the outcome, it would naturally produce far worse results. Halloran and Longini2 previously developed a classy framework for cohort studies to analytically adjust measured VE for less specific outcomes in order to provide true VE estimates; however, these authors also did not consider the possible complications of selection bias.

In the end, however, we must come back to the not-even-close-to-perfect world in which we all live. Here, observational studies often report on a variety of endpoints that include unconfirmed influenza-like illness, pneumonia hospitalization and death from any cause (see reviews1–3,5). Had Orenstein et al. extended their considerations to include the state-of-the-art in observational studies of VE of seniors, it would give them pause. Since the mid-1990s such studies have typically been set in electronic managed care organization databases, and laboratory confirmation has largely gone out of fashion.1–6 These studies often assess VE by looking for a relative decline in mortality from any cause during winter months among vaccinated people. But this endpoint has a specificity of only about 5%, in that excess mortality studies have shown that only about 5% of all winter deaths among seniors (~36,000 excess deaths/~600,000 winter deaths) can be attributed to influenza in an average season.7,8 Extrapolating Orenstein’s findings, such a low-specificity endpoint would be expected to profoundly underestimate the true VE; indeed, the measured VE should theoretically be in the range of 0–5%. Paradoxically, those observational studies have consistently reported VE estimates of about 50%6,8 for mortality. Given that influenza-related deaths account for only 5% of the total, this is a substantial over-estimation of influenza vaccine’s mortality benefit. It is remarkable that few have even questioned at the consistent reports that influenza vaccination could halve all winter deaths among seniors, even though that prediction clearly exceeds any realistic expectation of the true magnitude of vaccine-preventable deaths.

Recently, however, someone finally did. Lisa Jackson and her colleagues published two papers in IJE last year, demonstrating that the senior cohort study findings are largely a result of systematic mismeasurements due to the combined effect of residual selection bias, counter-productive adjustment efforts, and low-specificity endpoints.9,10 Jackson et al. employed two clever strategies to make their point. First, they showed that the greatest mortality reductions occurred in early winter before influenza ever circulated, and were not associated with the peak influenza period. Second, they showed that the analytical adjustment techniques typically used in cohort studies actually magnified the mismeasurement rather than reducing it.

Although the Orenstein base case scenario focused on a paediatric setting, I bring this ‘senior’ example to attention here because it is important to remember that while Orenstein’s focused analysis importantly endorses the use of laboratory-confirmed (and highly specific) endpoints, their study should not be taken to mean that observational studies always underestimate VE when less specific endpoints are used. I wish the Orenstein model scenarios could have somehow incorporated the real-life complications of unadjusted selection bias, because there is a great need for analytic studies to address the current confusion in the influenza literature, above and beyond endpoint specificity issues. Orenstein’s endorsement of laboratory-confirmed endpoints is a good beginning as it hopefully will convince authors of observational studies to consider laboratory-confirmed outcomes; at least ‘toxic’ enhancement of unadjusted selection bias can be avoided. But future efforts should go further, and rise to the challenge of identifying unadjusted selection bias in studies of influenza vaccine benefits by trying to quantify the degree of mismeasurement attributable to bias in observational studies.

References