Postlicensure Evaluation of COVID-19 Vaccines

After rigorous clinical trials of vaccines for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have established safety and efficacy, and after vaccines are deployed, evaluating vaccine performance in actual clinical settings will be essential for understanding the risks and benefits of vaccination programs. However, unique aspects of coronavirus disease 2019 (COVID-19) may pose challenges for these postmarketing approaches for evaluating vaccines. This Viewpoint describes potential methodologic challenges with using the commonly applied "test-negative" case-control design for evaluating COVID-19 vaccines and proposes potential solutions for consideration.

Postlicensure COVID-19 Vaccine Evaluations

The widespread morbidity, mortality, and societal ramifications of the COVID-19 pandemic have motivated the testing and development of vaccines at an unprecedented pace, with possible introduction in the population in the near future. More than 200 SARS-CoV-2 vaccines using multiple (some novel) platforms are being developed, and several candidates have entered phase 3 clinical trials. Vaccines are likely to be licensed as soon as safety and efficacy can be demonstrated. However, many unanswered questions will need postlicensure assessment, including the magnitude of protection across population subgroups defined by age, underlying conditions, and race/ethnicity; the duration of protection; comparative evaluations of vaccine types; assessments of 1 vs 2 doses of vaccination; and a more comprehensive understanding of safety, including rare complications.

The Test-Negative Design

During the past 2 decades, countries worldwide have introduced many vaccines and have amassed a wealth of practical and methodologic experience with postmarketing vaccine evaluations. Data generated by these evaluations have motivated global introductions, vaccine uptake, and policy modifications and informed benefit-risk considerations when unexpected safety concerns emerged.

A standard approach to postlicensure vaccine evaluation involves observational studies, such as cohort studies, or, more commonly, case-control studies, especially for diseases with low incidence. Although the findings of these observational studies have traditionally been described as estimating “vaccine effectiveness,” these estimates are not the same as “vaccine efficacy.” Efficacy is determined by prelicensure randomized clinical trials that are a direct method for establishing cause-and-effect relationships between vaccination and disease outcome through comparative incidence of disease in the 2 groups. In contrast, in postlicensure case-control studies vaccine performance is assessed by comparison of the odds of antecedent vaccination among cases (ie, people who developed the disease the vaccine is designed to prevent) and controls (ie, people who did not develop the disease the vaccine is designed to prevent) and by calculation of the corresponding odds ratio. Vaccine effectiveness is estimated as 1 minus the odds ratio and is expressed as a percentage. However, despite the conventional use of language in the vaccine literature that suggests causation (ie, vaccine effectiveness), protection is inferred from differences in vaccination rates between cases and controls. Thus, the design of these studies establishes associations between disease outcome and prior vaccine exposure. Compared with randomized trials, establishing causation in case-control studies is not possible, and any inference about potential causal relationships in observational investigations requires careful consideration of numerous additional factors, such as confounding, biological plausibility, collateral evidence, consistency in findings, strength of association, specificity, and biological gradient, as well as other factors specific to the study methodology, such as disease status misclassification or limitations of completeness or quality of variables in administrative data sets. The goals of “efficacy” and “effectiveness” studies are different and provide complementary evidence to understand the benefits of vaccination.

One of the standard approaches to postlicensure evaluations has been the test-negative design, a method that has become the de facto standard for influenza and rotavirus and has been used in most published postlicensure vaccine studies (approximately 300) in the past decade. The test-negative design is a variant of the case-control study in which all participants meet a prespecified syndrome definition, such as acute respiratory illness, and a diagnostic assay is used to differentiate cases from controls. The test-negative design is popular because it offers 2 advantages: simplified logistics, because controls are already identified when identifying cases, and reduced confounding due to similar health care use patterns between cases and controls, similar to a nested case-control study.

Biases With the Test-Negative Design for Severe COVID-19

Despite its use in some situations, the test-negative design is not infallible and requires disease-specific considerations of assay validity and enrollment case definitions. As with any observational study, test-negative studies can have bias due to misclassification of disease status using diagnostic assays with imperfect sensitivity and specificity, which appears to be particularly important with respect to COVID-19. However, low sensitivity and specificity are likely to introduce more bias in a test-negative design than in a traditional cohort or case-control study. In a traditional cohort or case-control study, imperfect assays only affect members of the source population that develop symptoms and are tested for illness; the majority of the source population is not tested, so only a small proportion of noncases can be misclassified by imperfect diagnostic tests. In contrast, because a test-negative study solely focuses on the...
subset of the population that meet some testing criteria, all non-cases in a test-negative study could potentially be misclassified.

COVID-19 has unique challenges, particularly among inpatients, that pose obstacles for the validity of the test-negative design. Analytic validation studies for SARS-CoV-2 reverse transcriptase-polymerase chain reaction assays indicate that sensitivity and specificity are high when validated against material containing viral RNA. In contrast, when reverse transcriptase-polymerase chain reaction is used clinically among persons who may have severe COVID-19, sensitivity has ranged from 33% to 80% due to factors such as delayed presentation when virus shedding is lower, inadequate swabbing, and site of sampling (nasal, oral, nasopharyngeal, tracheal). Specificity is likely high for detecting viral RNA in specimens, although it has not been directly evaluated in the clinical setting where cross-contamination or human error may occur. Moreover, clinical findings of severe COVID-19 overlap with common causes of severe respiratory illness, such as influenza, pulmonary edema, and asthma exacerbation. Patients with these conditions may coincidentally have test results positive for SARS-CoV-2 due to a concurrent mild infection or recovered infection with prolonged RNA shedding. Such scenarios may be more common when community prevalence of SARS-CoV-2 is high. Although misclassifying such patients is problematic for any study design when vaccines attenuate disease severity, bias may potentially be greater in test-negative studies. In addition, pragmatic challenges of standardized testing and case definitions could pose issues for the test-negative design during the evolving pandemic.

Potential Solutions for Reducing Bias

Although observational studies are susceptible to bias, anticipating potential flaws can improve study design. Imperfect specificity substantially affects test-negative and case-control studies. A case definition for severe COVID-19, an outcome of primary interest to the global community, can be strategically defined to improve specificity. Most patients with severe COVID-19 have a distinct phenotype of primary viral pneumonia with ground-glass opacities on chest radiographic imaging; low oxygen saturation; and markedly elevated markers of hyperinflammation, including C-reactive protein, D-dimer, and neutrophil-lymphocyte ratio. Thus, coupling SARS-CoV-2-positive test results with a clinical syndrome could reduce erroneous diagnoses of severe COVID-19. Clinically adjudicating test-positive cases and test-negative cases by independent expert clinicians blinded to the study design may help reduce bias. Test sensitivity remains a concern for diagnosing severe COVID-19, although it is less likely than specificity to bias vaccine effectiveness. Clinical sensitivity of assays may vary by specimen type, time since exposure, and specimen quality. Studies are needed to identify the optimal conditions for assay performance for detecting SARS-CoV-2 in the inpatient setting. These studies could inform enrollment criteria and standardize specimen collection to optimize sensitivity and improve the validity of the test-negative design. If sensitivity is demonstrated to be higher closer to onset of illness, restricting enrollment of participants meeting the case definition to that period would reduce bias related to lower sensitivity. However, this restriction would have to be balanced against reduced power related to missing cases of severe COVID-19 that present outside of the onset period or less generalizability of the findings to the subgroup with delayed clinical presentation.

Two additional approaches could be considered to validate the test-negative design for SARS-CoV-2 vaccine evaluation during this period of uncertainty and evolving knowledge on COVID-19. First, a traditional case-control study with disease-free controls from the emergency department, hospital, or community, in addition to the test-negative design, would be valuable until further evidence emerges on optimal choice of controls. Second, investigators could evaluate for potential bias by simultaneously conducting a “bias-indicator” study on “protection” afforded by SARS-CoV-2 vaccine against a sham non-COVID-19 outcome. All other aspects of the analysis and the controls would remain similar to the primary postlicensure vaccine evaluation study design. Any divergence of estimates from null against a sham outcome would suggest bias in the primary study of vaccine performance in protecting against severe COVID-19. Early studies that incorporate these types of checks and balances could strengthen the results and inform the future direction of vaccine evaluations.

Conclusions

Postlicensure vaccine evaluations will be a crucial component of an evidence-based SARS-CoV-2 vaccine program as it is used around the world. Anticipating potential pitfalls of these evaluations and identifying creative solutions to meet challenges posed by the complex biological effects of COVID-19 will improve study validity. With such high stakes, reflexively relying on the test-negative design as a fail-safe approach for vaccine evaluations may not be prudent unless modified to address the potential weaknesses of assay validity for diagnosing severe COVID-19. Immediate investments today in efforts to understand aspects of COVID-19 that pose methodologic challenges for “vaccine effectiveness” case-control studies under conditions in routine clinical settings could have dividends for the global vaccine rollout.