The necessity of validity diagnostics when drawing causal inferences from observational data

James Weaver[1], Erica A Voss[1], Guy Cafri[2], Kathleen Beyrau[3], Michelle Nashleanas[3], Robert Suruki[1]
[1]Janssen, Global Epidemiology, [2]Johnson & Johnson MedTech Epidemiology and Real-World Data Sciences, [3]Johnson & Johnson Global Medical Safety

Background

Autoimmune disorders may have primary manifestations such as joint pain and bowel inflammation but can also have secondary manifestations such as non-infectious uveitis (NIU). A regulatory authority communicated a potential safety concern from spontaneous reports for NIU following exposure to Remicade®, a biologic therapy with multiple indications for which alternative therapies are available. Thus, to assess the risk of NIU among patient exposed to Remicade® we conducted four analyses, each restricted to an indicated population to reduce confounding by underlying disease state. We applied four pre-specified diagnostics to assess causal effect estimate validity.

Methods

Separately among patients with inflammatory bowel diseases, psoriatic conditions, rheumatoid arthritis, and ankylosing spondylitis, we compared new users of Remicade® to indication-specific alternative therapies for the risk of NIU under four analysis strategies intended to address limitations of causal estimation using observational data. We developed and evaluated candidate NIU phenotypes using comprehensive clinical characterization. We fit Cox proportional hazards models, conditioned on propensity score (PS)-matched sets, to estimate the on-treatment risk of NIU among Remicade initiators versus alternatives under four analysis strategies, in four US administrative claims databases and one US electronic health record database. This produced 80 analyses, 20 of which we designated primary. We reviewed estimates from analyses that passed the following four validity diagnostics: 1) >0 outcomes in each treatment time-at-risk, 2) >35% patients in equipoise between 0.3 and 0.7 of the PS distribution, 3) standardized mean difference<0.1 for all covariates, 4) residual systematic error<0.25. We conducted meta-analyses within indication groups where >1 database passed diagnostics.

Results

Our phenotype development resulted in a NIU definition that returned patients with either [2 occurrences of a NIU code where the 2nd occurrence must be observed between 31 and 365 days relative to the 1st occurrence] OR [1 occurrence of a NIU code observed in an ophthalmology care setting]. Of all analyses, 19/80 (24%) passed diagnostics, 4/20 (20%) of which were primary. Of the 34 analyses that passed three diagnostics, most failed to achieve covariate balance. Of the 15 that passed two diagnostics, all failed covariate balance and most failed equipoise. Among patients with inflammatory bowel diseases, our primary, meta-analytic result indicated no evidence of an increased risk for NIU (pooled hazard ratio [HR] 0.75, 95% confidence interval [CI] 0.38-1.40). We observed no increased risk among patients with rheumatoid arthritis in one database that passed diagnostics, although results were uncertain (HR: 1.23, 95% CI 0.14-10.47).

Conclusions

We applied validity diagnostics to a study addressing a specific research question in heterogenous, real-world settings. Validity diagnostic results indicated that safety effect estimates from many analyses would be unreliable and inappropriate to interpret as causal, given the data available and methods employed. Rigorous phenotype development and evaluation and validity diagnostics application should always be used to determine if observational data source(s) and analysis design are conducive to producing reliable evidence for making causal inferences. Clinically, if an increased risk exists, it is unlikely to be greater than 40%.