Framework for Assessing the Reproducibility of Observational Comparative Effectiveness Researc: DOACs and Ischemic or Hemorrhagic Events Case Study

Asieh Golozar, Anna Ostropolets, Martin LaVallee, Adam Black, Ines Neves, Christian Reich

BACKGROUND:
The value of observational research lies in producing clear evidence on risks and benefits associated with medication uses in a real-world setting. However, there are concerns about the reliability of such evidence. Moreover, increasing numbers of research questions cannot be realistically investigated in clinical trials. Healthcare decisions are routinely informed by findings from a single study, often examining a specific population at a single point in time, with little consideration of how those findings fit within the greater body of evidence or generalize to different populations. Therefore, there is a need to develop systematic assessments of the reliability of existing evidence ensuring the meaningful application of the findings in different clinical settings. Recognizing this need, the OHDSI Center at Northeastern University launched the Reproducibility Service to provide a framework for systematic assessment of reliability of evidence generated from real-world data. In its first attempt, the OHDSI Center has begun evaluating the reliability of a recently published observational research findings (1) on the association of anticoagulants rivaroxaban and apixaban with major ischemic or hemorrhagic events in patients with atrial fibrillation. Here, we summarize the approach and some of the initial findings.

METHODS:
The general approach for this study is to accumulate a body of evidence to examine the reliability of the original Ray et al. findings (figure 1) using the OHDSI methodologies and toolkits. Specifically, we applied the new user active comparator cohort design in the IQVIA Open Claims database to replicate as closely as possible the analysis described by Ray et al who used a comparable database. To adjust for confounders, we used 1:1 propensity score matching. A series of sensitivity analys will be performed to assess the robustness of finding:

1) Changes in the definition of the exposure and outcome phenotypes (ex. different age groups, different pairs of DOACs, different timeframe)
2) Empirical calibration of effect estimates using negative and positive control outcomes
3) Using observational study diagnostics (ex. different time at risk, different propensity score methods for confounding control) to assess the robustness of their findings and risk of systematic bias before and after calibration

Finally, the generalizability of the findings will be evaluated in additional data sources from the US, EU and Asia Pacific. Target, Comparator, and Outcome cohorts are created according to the descriptions of the phenotypes in the paper and the appendix. Given the lack of death information in Open Claims, initial reliability assessment will focus on the following four outcomes: 1) hemorrhagic stroke, 2) ischemic stroke, 3) intracranial bleed, and 4) extracranial bleed. These outcomes, however, will be assessed in other data source with available death information.

RESULTS:
Details on the study can be found on the study GitHub page. Using Ray's definitions for exposure cohorts, we identified 852,413 patients with atrial fibrillation who were new users of rivaroxaban or apixaban between 2013 and 2018 with similar comorbidities and age distribution. The preference score distribution as shown in Figure 2 shows a considerable overlap between the Target and Comparator cohorts indicating that selection of comparable groups is possible through propensity score matching. Additional diagnostic results showed no imbalance between the two groups after propensity score matching (Figure 3). The preliminary results indicate similar effect estimates to the original paper with

overlapping confidence intervals (except for extracranial bleeding) (Figure 4). The current results will be updated after comprehensive evaluation of the cohort diagnostics. Empirical calibration and other sensitivity analysis results will be presented at the symposium along with findings from other data sources.

CONCLUSIONS:

It appears that the OHDSI methodologies and toolkits are well equipped to replicate and reproduce the findings of research that might otherwise be limited in reliability and generalizability. A more comprehensive view of findings in the literature is critical for informing clinical practice and for establishing targets for drug development.

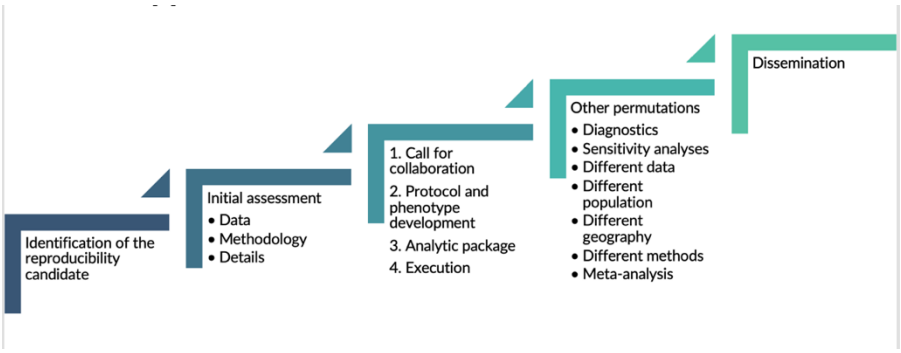

Figure 1. Systematic and sequential steps in replicating and reproducing a published study

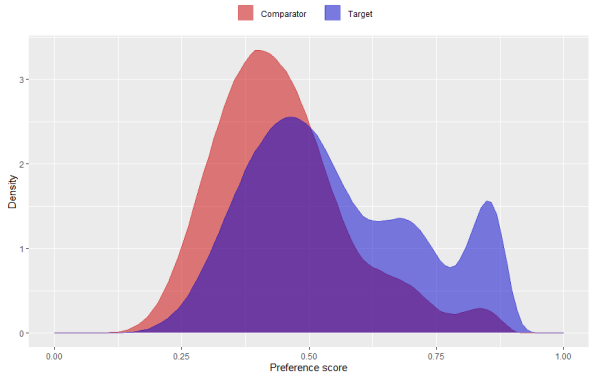

Figure 2. Preference score distribution

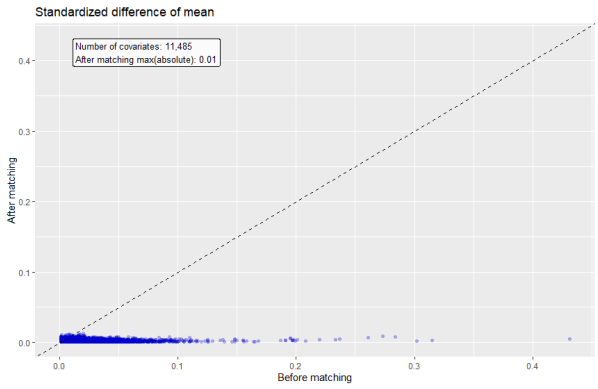

Figure 3. Covariate balance showing absolute standardized difference of mean before and after propensity score matching.

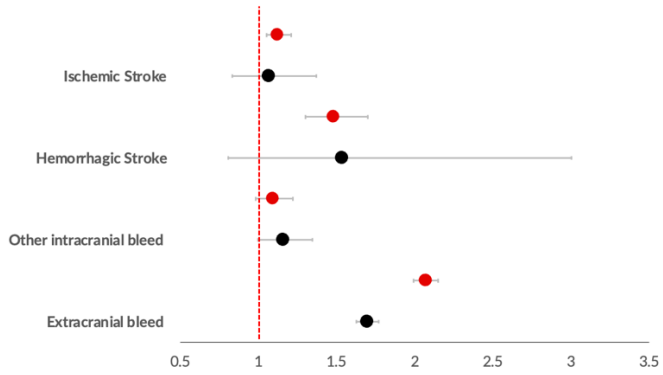

Figure 4. Hazard ratio (HR) (95% confidence intervals (CI)) for rivaroxaban versus apixaban on selected outcomes. Red: HR (95% CI) in the original study and black: reproducibility results

References:

1. Ray WA, Chung CP, Stein CM, Smalley W, et al. Association of Rivaroxaban vs Apixaban With Major Ischemic or Hemorrhagic Events in Patients With Atrial Fibrillation. JAMA. 2021 Dec 21;326(23):2395-2404. doi: 10.1001/jama.2021.21222.