

## Appendix 1 (as supplied by the authors): Illustrative example of the practice of testing group imbalances for confounder selection

We conducted a systematic search within the Canadian Medical Association Journal, a high-impact general medical journal, of published articles in 2018 that focused on non-randomized studies comparing two or more groups to assess the impact or effect of some exposure. Using the keywords "observational", "cohort", "quasi-experimental", "historical control" and "case control", we identified 34 eligible studies. Two independent raters (N.S and M.L.B.) assessed the articles for use of statistical testing for confounder selection and resolved any disagreements through consensus.

Among the 34 studies, 8 (24%) (95% CI: 11% to 41%) employed a form of statistical testing to examine group imbalances, commonly reported through p-values or confidence intervals in the patient characteristics table. Authors either directly or indirectly reported results from these tests to inform the choice of confounders to include in their adjusted models. For example, one article reported "We selected potential confounding variables based on the literature and on p-values (< 0.2) after univariate comparisons." Based on this example, the practice of confounder selection based on observed group differences remains present in the published literature.