ICSDS 2025

Causal Inference

Causal Record Linkage

Linked data sets present a valuable resource for research. In causal inference, record linkage methods grant researchers access to the covariates needed for unbiased estimation and offer the means to explore long-term outcomes. This paper examines how record linkage conflicts with the assumptions required for identifying causal effects. Our investigation reveals that linkage errors result in inconsistencies and alter exchangeability in the causal framework, leading to attenuation and opposite contribution biases in the inference. In attempting to address these biases by being more stringent on the linkage, positivity is curtailed and sampling bias inadvertently emerges. We demonstrate how to generalise the effect estimated on linked data and discuss how linkage decisions should be informed accordingly. Importantly, we identify when existing and novel solutions support valid causal inference on linked data and when inference should be treated with caution or even abandoned. We propose strategies to report on the estimated effect uncertainty and we illustrate the challenges raised and the potential solutions using a simulation study and real data from a Study on Women's Health.