ROBUSTNESS OF PREVALENCE ESTIMATES DERIVED FROM MISCLASSIFIED DATA FROM ADMINISTRATIVE DATABASES

M.Ladouceur¹, E.Rahme², C.Pineau², L.Joseph¹⁻²

¹McGill University, Montréal, Canada ²Montréal General Hospital Research Institut, Montréal, Canada

Email: martin.ladouceur@mcgill.ca

Since primary data collection can be expensive, researchers are increasingly using information collected in medical administrative databases for scientific purposes. This information, however, is typically collected for reasons other than research, and many such databases have been shown to contain substantial proportions of misclassification errors. For example, many administrative databases contain fields for patient diagnostic codes, but these are often missing or inaccurate, in part because physician reimbursement schemes depend on medical acts performed rather than any diagnosis. Errors in ascertaining which individuals have a given disease biases not only prevalence estimates, but also estimates of associations between the disease and other variables, such as medication use. We attempt to estimate the prevalence of osteoarthritis (OA) among elderly Quebeckers using a government administrative database. We compare a naive estimate relying solely on the physician diagnoses of OA listed in the database, to estimates from several different Bayesian latent class models which adjust for misclassified physician diagnostic codes via use of other available diagnostic clues. We find that the prevalence estimates vary widely, depending on the model used and assumptions made. We conclude that any inferences from these databases need to be interpreted with great caution, until further work estimating the reliability of database items is carried out.