

International Journal of Population Data Science

Journal Website: www.ijpds.org



Using linked cohort data to help address residual confounding in analyses of population administrative data

Richard Silverwood¹, Gergo Baranyi¹, Lisa Calderwood¹, Bianca De Stavola¹, George Ploubidis¹, Ian White¹, and Katie Harron¹

¹University College London, London, United Kingdom

Objectives

Analyses of population administrative data can often only be minimally adjusted due to the unavailability of a full set of control variables, potentially leading to bias due to residual confounding. We aimed to use linked cohort data to help address residual confounding in analyses of population administrative data.

Methods

We propose a multiple imputation-based approach, introduced through application to simulated data in three different scenarios related to the structure of the datasets. We then apply this approach to a real-world example – examining the association between pupil mobility (changing schools at non-standard times) and Key Stage 2 (age 11) attainment using data from the UK National Pupil Database (NPD). The limited control variables available in the NPD are supplemented by multiple measures of socioeconomic deprivation captured in linked Millennium Cohort Study (MCS) data.

Results

The proposed approach is observed to perform well when using simulated data across the different scenarios. The association between pupil mobility and Key Stage 2 attainment was attenuated after supplementing the NPD analysis with information from linked MCS data, though with a decrease in precision.

Conclusion

We have demonstrated the potential of the proposed approach. More work is required to understand how it can be applied more broadly. The principles underlying this innovative approach are widely applicable: any analysis of administrative data could potentially be strengthened by linking a subset of individuals into richer cohort data.



Temporary page!

$\LaTeX$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because $\LaTeX$ now knows how many pages to expect for this document.