International Journal of Population Data Science

Journal Website: www.ijpds.org

Effects of Data Quality on Linkage Bias: Results from a Central Register of Pupils

Severin V. Weiand[1], Rainer Schnell[1], and Tobias Brändle[2]

[1]University of Duisburg-Essen, Duisburg, Germany

[2]Institut für Bildungsmonitoring und Qualitätsentwicklung, Hamburg, Germany

In record linkage, quasi-identifiers (QIDs), such as personal names or addresses, are used to link entities. Although it is generally known that QIDs are subject to data quality issues and may thus affect linkage quality, there is little empirical evidence on the extent of such effects. Hence, except for medical applications, empirical data on differences in linkage quality across population subgroups are rare. To study linkage bias in a non-medical context, we linked two school years of the central pupil register of the German city of Hamburg using a unique ID ($N1 = 246{,}472$ and $N2 = 252{,}230$, with an overlap of 182,385 records). We compared the results with record-linkage procedures based on QIDs. Furthermore, we tested whether geo-locations might be used as additional QIDs to improve linkage results. The results show linkage bias among migrants, as they exhibit larger errors on QIDs and distinct regional mobility patterns. Finally, pupils living in large residential buildings are less likely to be linked over time; therefore, linkage bias resulting in selective longitudinal attrition would be observed in record-linkage studies using QIDs.

https://doi.org/10.23889/ijpds.v11i5.3714