

International Journal of Population Data Science

Journal Website: www.ijpds.org



Assessing markers of data quality in routinely-collected primary care data

Emma Whitfield¹, Matthew E Barclay¹, Nadine Zakkak², Meena Rafiq³, Georgios Lyratzopoulos¹, and Becky White¹

¹University College London, London, United Kingdom

²University College London, London, United Kingdom; Cancer Research UK, London, United Kingdom

³University College London, London, United Kingdom; University of Melbourne, Melbourne, Australia

Routinely-collected primary care data are widely used for research. However, clinical recording likely varies between GP practices. It is therefore important that researchers consider practice-level data quality and, if necessary, take steps to exclude data of insufficient quality. One approach is to assign each practice an 'up-to-standard' (UTS) date after which its data is considered acceptable for research. This study will propose a methodology for deriving practice UTS dates in routinely-collected primary care data. Our methodology will be developed on a sample extracted from CPRD Aurum for the purpose of examining risk of cancer or symptomatically-similar diseases in patients presenting in primary care. We will consider four components likely to reflect practice-level data quality: the registration of new patients, and the recording of

observations, consultations, and deaths. Researchers experienced in primary care data analysis will review time series of practice-level metrics based on these components for a random sample of 30 practices. This will inform the development of criteria to determine up-to-standard dates. These criteria will then be applied to the full extract to generate a UTS date for each practice. We will report descriptive statistics to examine how the use of these dates affects our sample. Our findings will be ready to present at the time of the conference. It is important that research data are of sufficient quality to minimise bias. The use of UTS dates will help ensure robust analyses, and facilitate comparison and replication between primary care data sources.

