

## Agreement and completeness of routine versus trial-specific patient outcome data : a systematic review

Mr Saïam Ahmed<sup>1</sup>, Prof. Matthew R. Sydes<sup>1</sup>, Ms. Sharon B. Love<sup>1</sup>, Prof. Nicholas D. James<sup>2</sup>, Prof James Carpenter<sup>1</sup>

<sup>1</sup>MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, <sup>2</sup>Institute of Cancer Research

There is growing interest in enhancing the conduct of randomised controlled trials (RCTs) by using routinely-collected healthcare data (RCHD). However, few studies have formally compared the suitability of these healthcare systems datasets, collected through interactions between patient and the health service. This review identified studies that compared RCHD to trial-specific data collection to assess the quality, challenges, and suitability for use by trialists. It provides an overview of the outcomes that have been investigated and the conclusions drawn whilst identifying key gaps in the evidence.

The review searched MEDLINE and EMBASE for papers published between January 2017 – April 2021. The search strategy follows on a previous unpublished review by G. Powell (University of Liverpool). A manual search was also performed in conference proceedings of the International Clinical Trials Methodology Conference and Society for Clinical Trials. A study was considered eligible if at least one routine data source (e.g. hospital episode statistics (HES) or data from a national data provider) was compared to trial-specific data collection in the UK. The review protocol was prospectively registered on PROSPERO (CRD42020186048).

1977 records were identified of which 1945 were excluded on title and abstract screening. Of 32 papers considered in more detail for eligibility, 26 papers were excluded. The 6 eligible papers represent 6 studies that compared RCHD between 2002-2015 to trial-specific data. Data assessments varied from comparing outcomes of interest to trialists, such as incidence of hospital admission and overall survival, with comparisons of clinical characteristics. Authors' methods used to assess agreement varied, with half the studies using Kappa statistics and the remainder using a combination of frequency, proportions, sensitivity, and specificity. Two studies assessed death data: one reported no evidence of a difference in 5-year and 8-year survival rates and the other reported substantial agreement however, highlighted some disagreement of confirmed deaths between both sources. Two studies assessed hospital admission: one reported better sensitivity and specificity of HES compared to a reference and the other reported an underestimation of events by HES.

Surprisingly, few studies have presented formal assessment of the relationship between trial-specific data collection and RCHD. Recent publications represent older assessments with limited recent evidence of the suitability of current RCHD that are available through national data providers like NHS Digital. This highlights the need of carrying out these data assessments within ongoing clinical trials to map the potential of RCHD and utilise effectively in clinical trials.