Submission Id: 2907

Title

Multimorbidity in a selected cohort compared to a representative sample: Does selection bias influence outcomes?

Priority 1 (Research Category)

Big Data

Presenters

Peter Hanlon, , MBChB, MSc, Bhautesh Jani, PhD, MB ChB, MRCGP, Frances Mair, MD, FRCGP, David McAllister

Abstract

Context: UK Biobank is increasingly used to study causes, associations, and implications of multimorbidity. However, UK Biobank is criticised for lack of representativeness and 'healthy volunteer bias'. Selection bias can lead to spurious or biased estimates of associations between exposures and outcomes.

Objectives: To compare association between multimorbidity and adverse health outcomes in UK Biobank and a nationally representative sample.

Design: Cohorts identified from linked routine healthcare data from UK Biobank and from the Secure Anonymised Information Linkage (SAIL) databank.

Setting: Community.

Participants: UK Biobank participants (n=211,597, age 40-70) with linked primary care data and a sample from a nationally representative routine data source (SAIL) (n=852,055, age 40-70).

Main outcome measures: Multimorbidity (n=40 long-term conditions [LTCs]) was identified from primary care Read codes and quantified using a simple count and a weighted score. Individual LTCs and LTC combinations were also assessed. Associations with all-cause mortality, unscheduled hospitalisation, and major adverse cardiovascular events (MACE) were assessed using Weibull or Poisson models and adjusted for age, sex, and socioeconomic status.

Results: Multimorbidity was less common in UK Biobank than SAIL. This difference was attenuated, but persisted, after standardising by age, sex and socioeconomic status. The effect of increasing multimorbidity count on mortality, unscheduled hospitalisation, and MACE was similar between UK Biobank and SAIL at LTC counts of ≤3, however above this level UK Biobank underestimated the risk associated with multimorbidity. Absolute risk of mortality, hospitalisation and MACE, at all levels of multimorbidity, was lower in UK Biobank than SAIL (adjusting for age, sex, and socioeconomic status).

Both cohorts produced similar hazard ratios for some LTCs (e.g. hypertension and coronary heart disease) but underestimated the risk for others (e.g. alcohol problems or mental health conditions). Similarly hazard ratios for some LTC combinations were similar between the cohorts (e.g. cardiovascular, respiratory conditions), UK Biobank underestimated the risk for combinations including pain or mental health conditions.

Conclusions:

UK Biobank accurately estimates risk of outcomes associated with LTC counts ≤3. However, for counts ≥4 estimates of magnitude of association from UK Biobank are likely to be conservative.