NIHR Statistics Group – Routine Data resources

Below are useful links for designing and analysing studies using routine health data (predominantly primary and secondary care databases from the UK such as the Clinical Practice Research Datalink).

Background on UK routine databases

- **CPRD.** Information, including data dictionaries, available on their website https://www.cprd.com. Herrett et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015 Jun; 44(3): 827–836. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4521131/
- **THIN.** Blak et al. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care.* 2011;19(4):251-5. https://hijournal.bcs.org/index.php/jhi/article/view/820
- **HES.** Information, including data dictionaries, available on the website https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics. Herbert et al. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol.* 2017 Aug; 46(4): 1093–1093i. https://academic.oup.com/ije/article/46/4/1093/3072145
- **CALIBER.** Denaxas et al. Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol.* 2012 Dec; 41(6): 1625–1638. https://academic.oup.com/ije/article/41/6/1625/748070

Checking a similar study is not already underway

- The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) registered protocols http://www.encepp.eu/encepp/resourcesDatabase.jsp
- CPRD's approved studies https://www.cprd.com/protocol-list

Study design and reporting guidelines

- "Developing a Protocol for Observational Comparative Effectiveness Research": http://www.effectivehealthcare.ahrq.gov/ehc/products/440/1166/User-Guide-to-Observational-CER-1-10-13.pdf
- The REporting of studies Conducted using Observational Routinely-collected Data (RECORD) – http://record-statement.org/
- The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) checklist for study protocols http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml

Coding

Routine data studies require time to define code lists to identify patients of interest for the study. Below are resources to help select code lists and code lists used by other researchers.

The following resources are guides to creating code lists:

- Watson et al. Identifying clinical features in primary care electronic health record studies: methods for codelist development. *BMJ Open* 2017;7:e019637. https://bmjopen.bmj.com/content/7/11/e019637
- Dave S and Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf.* 2009 Aug;18(8):704-7 https://onlinelibrary.wiley.com/doi/pdf/10.1002/pds.1770

The following are freely available Read code lists created by other researchers:

- The University of Manchester published Read codes repository – https://clinicalcodes.rss.mhs.man.ac.uk/.
- CPRD at Cambridge https://www.phpc.cam.ac.uk/pcu/cprd_cam/codelists/
- CALIBER Read Code repository https://www.caliberresearch.org/portal. This includes Read code lists for the following publication: Kuan et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *The Lancet Digital Health (2019)* https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30012-3/fulltext

Validity of code lists reviews

- Khan et al. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract*. 2010 Mar; 60(572): e128–e136. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2828861/
- Herrett et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol.* 2010 Jan; 69(1): 4–14. https://bpspubs.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2125.2009.03537.x
- Lewis et al. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf.* 2007 Apr;16(4):393-401. https://onlinelibrary.wiley.com/doi/abs/10.1002/pds.1335