* Medcodes
  + CPRD unique code for the medical term selected by the GP (GP systems often swap between using read codes and SNOMED codes)
  + To be used with the observation table
    - This contains the medical history including:
      * Symptoms
      * Clinical measurements
      * Laboratory test results
      * Diagnoses
      * Demographic information
    - Long format (i.e., multiple rows per subject)
    - From the CPRD Aurum Data Specification document:

A close-up of a document

AI-generated content may be incorrect.

* Prodcodes
  + These are for the drug issue table
* Read codes
  + Read codes are coded thesaurus of clinical terms
  + They provide a standard vocabulary for clinicians to record patient findings and procedures
  + Basis of the medcode coding system used in the CPRD gold event files
  + Historically used by GP systems before switching to SNOMED
* SNOMED codes
  + Used by GPs like read codes
  + Most comprehensive coding system used worldwide
* ICD-10 and OPCS-4 codes are used during hospital visits
  + Linked data
  + In patients
  + ICD-10 are diagnoses
  + OPCS-4 are procedures
  + Surgeries can be useful for assessing a diagnosis
    - For example if you are looking at retinopathy and a patient has had a laser treatment then you can tell it was a severe case
* Different tables
  + Observation is the one most relevantA diagram of a patient structure

    AI-generated content may be incorrect.
* NHS classifications browser (ICD-10 5th edition)
  + Volume 1 – Tabular list
  + Tabular list of inclusions and four character subcategories
* EMIS Medical Dictionary text file (shared by katie)
  + Allows you to look up Med codes and what their Snomed code is
* CPRD Aurum
* Sourcing published code lists
  + Open safely – OpenCodelists
  + HDRUK Phenotype library
  + Hopkins et al. (2025)
* A reproducible open-source framework for defining type 1 and type 2 diabetes reseaarch cohorts in routinely collected electronic health record data
  + EHR records are observational data that are generated and collected as part of routine clinical care (from GPs or hospital admissions)
  + Need to find a robust and reproducible framework for defining standardised type 1 and type 2 diabetes cohorts due to several current issues with data transformation from raw EHR to a ready dataset:
    - Complex and time-consuming
    - Often repeated for each individual study published, and with different rules to define same population
    - Unstandardised, resulting in heterogenity
    - Full descriptions rarely provided in papers
    - Guides rarely available on defining specific cohorts
  + Methods
  + A framework to define diabetes research cohorts using EHR data
    - Gather code lists (source published code lists, generate own clinically reviewed code lists using standardised pipeline)
    - Define diabetes population (using clinical codes for diabetes)
    - Define diabetes diagnosis rate
    - Classify diagnosis types
    - Define index date (incident (date of diagnosis), prevalent (01/02/2020), or treatment initiation cohort)
    - Define baseline features (e.g., sociodemographics, comorbidities, biomarkers, medications)
    - Define key outcomes (first incidence, mortality, treatment response)
  + CPRD Aurum is a large database of longitudinal, routinely collected medical records from primary care practices in the UK and contains information on patients’ demographics, diagnoses, prescriptions, lifestyle factors, and test results
    - Covers ~ 13% of the UK population and is largely representative
  + Code lists: EHR data are commonly stored in the form of codes.
  + To define each variable of interest in our dataset, we need to generate a list of the codes that could be used to record that variable
  + Published codelists are available from online repositories
  + Or can be generated using a standardised pipeline:
    - Collect codes
    - Combine and de-duplicate
    - Final code list
    - Map to CPRD medcodes using CPRD Medical dictionary
    - Final medcode list to use in CPRD data
  + When existing codelists are not available, or to ensure comprehensiveness, term searching the target coding system or mapping from one coding system to another can also provide further codes
  + Codes from different sources are then combined and reviewed by a clinician
  + Quality and outcomes framework (QOF)
    - Disease prevalence and care quality achievement rates
  + Hospital Admissions Statistics (HES)
    - Contains data on all admissions to the NHS secondary care providers
  + CPRD have their own quality standards for data which include identifying patients with non-continuous follow-up or poor data recording issues wose records are deemed not “acceptable” for research
  + Valid dates
    - No earlier than the patient’s DoB
    - No later than the patient’s end of practice registration, last collection date from the practice, or date of death
  + Classifying diabetes
    - Need to use robust and valid classification approaches
    - Look at clinical codes
    - Prescriptions
    - Features such as age at diabetes diagnosis