Primary Care Database practical

NCRM e-Methods festival 7-9 November 2023

In this short practical session, you will get a chance to look at some artificial primary care database data to get an idea of the structure of this type of data, and what sort of coded information is available. We will look at just three types of files:

1. The patient file – which contains some basic demographic information and dates
2. The diagnosis file – which contains information on coded diagnoses and clinical findings
3. The prescriptions file - which contains records for each electronically issued prescription in primary care

The names of these files and variables contained within will be slightly different in different primary care databases (e.g. CPRD Gold and Aurum each contain very similar information, but in differently named files).

You can download the three csv files from this GitHub link:

<https://github.com/UCL-CHIG/intro_to_pcd_course>

For this practical, we will be looking at some diagnoses coded using Read codes, and prescriptions using drug codes. For the purpose of the practical, the descriptions these codes stand for have been added to the files.

The practical can be done using excel (answers will be provided using excel functions), but you can also use the statistical software of your choice.

1. Open the patient file (PCD\_patient\_file.csv) and count how many patients are in the file, how many men and women there are, and how many patients have died.  
   *(Hint: sex is coded as 1 for men and 2 for women)*

Answer: 400 patients, 191 men (sex=1) and 209 women (sex=2), 16 patients have died

1. Look up patient with ID number 216. How old were they when they registered with the practice, and how old were they at the last data collection date?   
   (*Hint: use lastdate, birth\_date and registration\_date)*

This woman (sex = 2) was 19 when she registered, 33 at the last data collection date and no transfer out or death date so she is still registered with the same practice. There should be about 14 years of primary care data for her.

1. Open the diagnosis file (PCD\_diagnosis\_file.csv) and just take a look at the file itself. You’ll see each patient can have multiple diagnoses recorded (using *read codes*), over time or on the same day. How many diagnoses are recorded, and what is the most common diagnosis code?   
   (*Hint: if you are using excel, you can use UNIQUE() and COUNTIF() functions)*

Answer: 12,412 diagnoses for 372 patients, and 9N11.00 (555 times): Seen in GP’s surgery

1. Now try to look up a specific condition, for instance non-insulin dependent diabetes mellitus (you might want to start with Read code C109.00). How many times has a diagnosis for this condition been entered? Would you want to use any other Read codes to enhance the phenotype? (Hint: look up code lists for diabetes on the HDR UK phenotype library <https://phenotypes.healthdatagateway.org/>

Answer: 24 times (for 7 individual patients). The phenotype library has a code list for diabetes ([PH152](https://phenotypes.healthdatagateway.org/phenotypes/PH152/version/304/detail/)) which includes 135 different Read codes.

1. Coming back to patient with ID 216, what does their diagnosis history look like? Does this patient have any underlying chronic health conditions?   
   (*Hint: make sure the data are sorted chronologically according to diagnosis\_date)*

Answer: The patient has asthma, which is recorded soon after she registers. From the diagnosis codes it also seems like she has delivered at least one child and has had several more pregnancies which were high risk and several ended in miscarriages.

1. Now open the prescription file (PCD\_prescription\_file.csv). Same as the diagnosis file, start by just having a look a the layout of the file and see how many prescriptions were issued for this group of patients.

Answer: 26,879 prescriptions for 311patients

1. Which is the most common prescription?   
   (*Hint: if you are using excel, you can use UNIQUE() and COUNTIF() functions)*

Answer: aspirin with 434 prescriptions

1. Look at patient 216 again, they only had 1 diagnosis code for an underlying condition. Does their prescription history give you more information on the severity or duration of this condition?

Answer: looking at their prescription history, regular prescriptions for asthma inhalers (salbutamol) over the years show the patient continued to need medication. Plus prescriptions for oral corticosteroids (prednisolone) and antibiotics (amoxicillin) could indicate exacerbations.