Coded data from Primary Care Electronic Extraction: August 2023

In the initial manual extraction of data for Exceed we had to make a choice about the range of coded data to extract because of file size limits applicable at the time. The data extracted was diagnostic data from all the Read hierarchy diagnostic chapters, codes on cancer tumour stage, symptom codes, occupation codes, examination findings, procedures and operations, referrals, respiratory related exacerbations and administration, smoking, exercise and alcohol, adverse drug reactions, family history, main and secondary spoken language and ethnicity.

In addition, a limited range of test codes were extracted along with their numeric values, where applicable. This included: - BP, BMI, height, weight, waist ratio, spirometry, Full blood counts, U+E, creatinine, EGFR, Thyroid function, cardiac enzymes, liver function, plasma viscosity, BNP, HbA1C, Imaging Chest x-ray, CT, glucose, calcium, cholesterol and CRP.
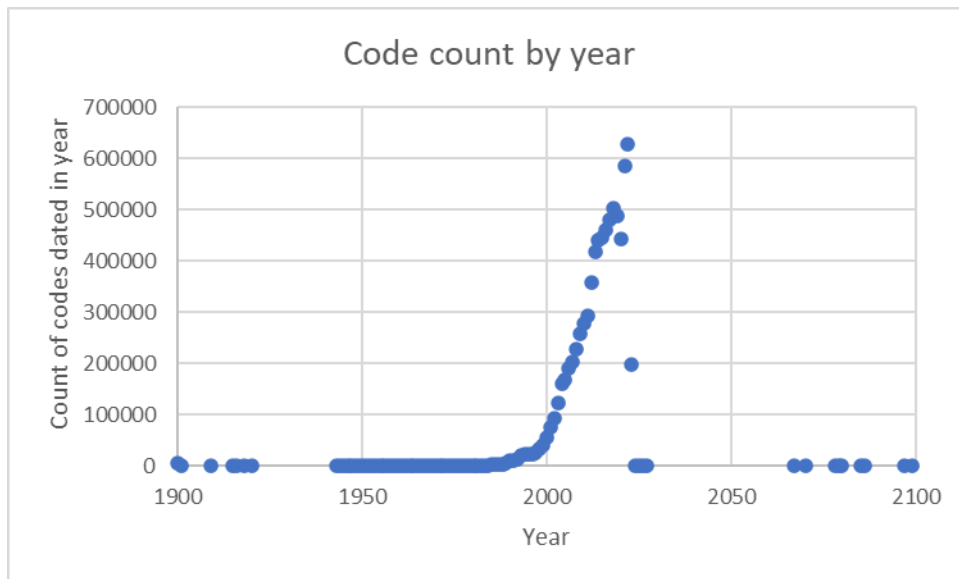
The new electronic extraction consists of all coded entries in the electronic heath record. The range is therefore more comprehensive.

There are some differences to be aware of: -

- In CTV3 there are Y codes, that is codes starting with a Y, not the usual X. These are codes created by SystmOne that are not standard CTV3 codes. Many of which have no SNOMED equivalent. Fortunately, the vast majority are minor administrative codes.
- The SNOMED codes are given alongside the equivalent CTV3 codes
- We now have the date of the event as well as the date it was entered to the GP system. (Previously we did not have the latter)
- Numeric values on some codes have the units of measurement, which we did not have before
- Some fields in the SRCode table link by rowidentifier to tables which we do not have ie SROrganisation, SRVisit, SRPatient, SREvent, SRQuestionnaire, SRTemplate, SRStaffMemberProfile, SRStaffMember
- Most of these are for admin functions in the SystmOne they came from or would be inappropriate for us to have.
- None of the other tables in the download appear to link to SRCode by row.

There are about 3.4m numeric codes and 4.4m non numeric codes in the extract.

There are over 9,000 unique study identifiers with 7.8m codes, averaging 840 codes per participant (range 1 to 6923).

The majority of codes have been dated from the turn of the century, 6m dated since 2010, 1.6m dated between 2000 and 2009 inclusive, with a few (65) future dates, some of which may be future recall dates.

There are 84 different GP organisations identified as sources for the coded data, averaging 186,300 codes each. (range 1 to 1.1m!)

The range of codes now fully covers that typically found in GP records.

Here is a list of the sort of categories of code with some actual examples to give an idea of what is available.

Diagnosis codes eg XE0eW Endometriosis, H33..Asthma, E200z Anxiety state NOS,

Injury codes eg SH2.. Burn of trunk, S2301 Closed fracture olecranon, extra-articular

Symptom codes eg 1991. No vomiting, R081. [D]Dysuria, Xa35W Pain in toe,

Clinical finding codes eg 2431.   O/E - pulse rhythm regular, 24F6. O/E - L.post.tib.pulse present, XaIwz Blanching rash

Administration codes eg XaIhQ Scanned document, 9344. Notes summary on computer, Xa1kG Did not attend, XaIgc Telephone call to a patient

Test codes eg 4682. Urine ketone test negative done at the practice, and lab results eg XE2px Serum alkaline phosphatase level, 4381. Syphilis titre test negative

Procedure codes eg X20Th Sigmoidoscopy, X70lf Ultrasound scan, 537.. Mammogram

Referral codes eg XaAZQ Referral by GP

Ethnicity codes eg XaJR2 Indian or British Indian - ethnic category 2001 census

Language codes eg XaG5t Main spoken language English

Hospital attendance codes eg 9N1y6 Seen in fertility clinic, 9N19. Seen in accident and emergency department

Care delivery codes eg XaIfK Asthma medication review, XaIoW Medication review done by pharmacist, 62R.. Postnatal visit

Family history codes eg 12A1. FH: Glaucoma

Social history codes eg XaCJ0 Diet average, XE0oh Never smoked tobacco, 13Z8. Social problem, XaMz4 Divorced/person whose civil partnership has been dissolved, 1361. Teetotaller

Medical risk assessment codes eg 14O8. At risk of diabetes mellitus, XaLDN Patient health questionnaire (PHQ-9) score, XM0cp Dundee Coronary Risk Disk, XaQVY QRISK2 cardiovascular disease 10 year risk score, XaORP Alcohol use disorder identificatn test consumptn questionnre

Encounter codes eg 8CB.. Had a chat to patient

In terms of formulating research questions or devising phenotype definitions, the above information should give researchers an idea if the data they need is likely to be covered. The next step would be to go into an online SNOMED browser to create a list of the kinds of codes needed and then we can search the database to show what is actually available. I am happy to advise on peoples' code lists as I may well be able to suggest codes that people won't have thought of either as specific codes or useful proxies / alternative search strategies.

DJS Oct 2023