Healthcare Records Data Simulation

This dataset represents a simulated electronic medical record designed to track patients, staff, and their interactions in a healthcare setting.

There are intentionally ‘dirty’ data elements included in some of these tables that should be examined and cleaned (illogical date of births, etc).

There are multiple tables included:

* Patient table:
  + Every patient is included in this table with one patient per row.
  + Demographics about the patient are captured as different columns.
  + This table links to many other tables using PatientID, including OutpatientVisit, PatientInsurance
* PatientInsurance table:
  + This table captures patient insurance status over time. If a patients insurance status changes, then a new row is generated (one patient is potentially in multiple rows).
  + We do not have ‘insurance end date’ in this table – we only have insurance start date (CoverageStartDate). A reasonable assumption is that patients can only have one insurance active at a time – so if we observe a new insurance start, the old one might end.
  + This table can be linked to the patient table (PatientID) and the insuranceProvider table (InsuranceProviderID)
* InsuranceProvider table:
  + This table has one row for each insurance provider (insuranceProviderID), along with the type of insurance.
  + This can be linked to the PatientInsurance table
* Staff table:
  + This table has one row for each staff member along with demographics information about the staff member.
  + This table can be joined to the OutpatientVisit table (staffID), as well as self-joined using ‘StaffReportsTo’ and ‘StaffID’
* OutpatientVisit table:
  + This table represents visit that occur in the healthcare system (a single visit to a doctor for example). Each row is one visit.
  + The patient and staff member participating in the visit are documented.
  + One or more appropriate disease codes (ICD10) are documented. For example, if the visit was to treat diabetes, then a diagnosis code representing diabetes would be present. If both diabetes and depression were treated or diagnosed during the visit, then there would be one ICD Code present for depression (ICD10\_1), and another present for diabetes(ICD10\_2).
  + The type of clinic is indicated by the clinicCode
* Clinic table:
  + Every clinic is listed in this table and if it is specialty care, primary care, or emergency department.
  + This can be linked to the OutpatientVisit table to identify if visits occurred in primary care, specialty care, or emergency departments.
* Mortality table:
  + If a patient died, their date of death is in this table. This can be linked to the patient table.
* DiseaseMap table:
  + This tables maps the ICD10 codes to the different types of disease (condition).
  + This can be linked to the visit table to understand which conditions (diseases) were address during the visit (ICD10\_1, ICD10\_2, ICD10\_3 can all be joined to ICD10)