NHS
DATA

Utilizing SQL for Clinical Data Analysis

By Lateefat Tobun & Ibrahim Oshinubi

Overview

Introduction

- Utilizing SQL for querying clinical data in a research and development environment. Three distinct scenarios are presented, each requiring the extraction of specific data from the provided database schema:

- **Analysing anaemia prevalence and trends.**
- **Analysing impact of Multiple Myeloma incidences**
- **Identifying at-risk patients for urgent medical advice.**

Dataset

PATIENTS Table

containing IDs and demographic data for each unique patient

PATIENT_ID:	[int]
GENDER:	[varchar]
POSTCODE:	[varchar]
CURRENT_AGE:	[int]
DATE_OF_DEATH:	[date]

BLOOD_TEST Table

containing the results of all pathology blood tests undertaken for patients.

RESULT_ID:	[int]
PATIENT_ID:	[int]
SAMPLE_DATE:	[date]
RESULT_DATE:	[date]
TEST:	[varchar]
RESULT:	[int]
UNIT:	[varchar]
RANGE_FLAG:	[varchar]

INPATIENT_SPELLS Table

containing IDs and demographic data for each unique patient

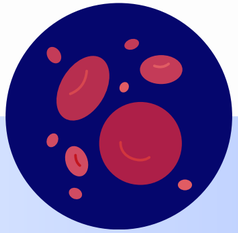
SPELLS_ID:	[int]
PATIENT_ID:	[int]
ADMISSION_DATE:	[date]
DISCHARGE_DATE:	[date]

DIAGNOSIS Table

containing ICD-10 coded diagnoses made by clinicians, recorded during inpatient spells. Due to comorbidity recording, multiple diagnoses may be made for each spell.

SPELLS_ID:	[int]
DIAGNOSIS_CODE:	[int]
DIAGNOSES:	[date]

Anaemia & Haemoglobin Levels



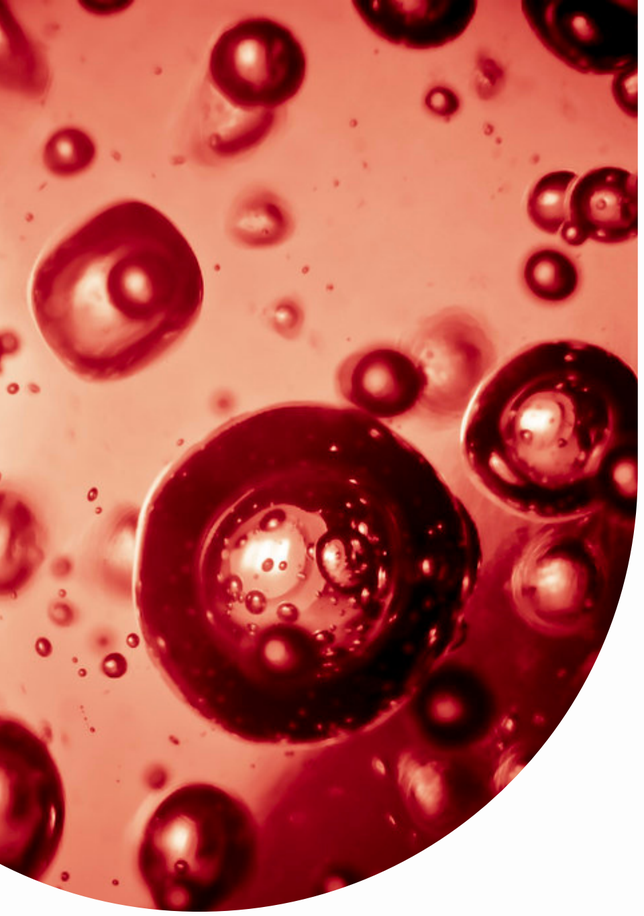
Haemoglobin is a protein found in red blood cells that carries oxygen throughout the body. Abnormal levels of haemoglobin can indicate various conditions, including anaemia.

Testing

A blood sample is collected from the patient.

Target

This analysis is relevant for assessing anaemia prevalence and trends.



SQL Query

- **Objective**:
  Investigate fluctuating levels of anaemia over time by examining haemoglobin levels in blood test data for female patients aged 18 and over in the ‘SO’ area.

- **CODE:**

```
SELECT * from PATIENTS
JOIN BLOOD_TESTS
ON PATIENTS.PATIENT_ID = BLOOD_TESTS.PATIENT_ID
WHERE PATIENTS.CURRENT_AGE >=18
AND PATIENTS.GENDER = ’Female’
AND BLOOD_TESTS.TEST LIKE’Haemoglobin’
AND PATIENTS.POSTCODE LIKE ’SO%’
AND BLOOD_TESTS.SAMPLE_DATE BETWEEN ’2015-01-01’ AND ’2022-12-31’
```

- **Result**: This query retrieves all relevant blood test data for female patients aged 18 and over in the ‘SO’ area, specifically focusing on haemoglobin levels between January 1, 2015, and December 31, 2022

Multiple Myeloma

Multiple Myeloma is a cancer of plasma cells, a type of white blood cell present in the bone marrow. It leads to the overproduction of abnormal plasma cells, which can cause bone pain, anaemia, kidney damage, and other complications.

Testing

Diagnosis typically involves a combination of blood tests, urine tests, imaging studies, and bone marrow biopsy.

Target

Analysing the incidence of Multiple Myeloma and its impact on hospital admissions over time.

SQL Query

- **Objective**:
  Extract inpatient spells for patients diagnosed with Multiple Myeloma and provide a total count of distinct spells per year.

- **CODE:**

```
--Total Multiple Myeloma diagnosis
SELECT * FROM DIAGNOSES
JOIN INPATIENT_SPELLS
ON DIAGNOSES.SPELL_ID = INPATIENT_SPELLS.SPELL_ID
WHERE DIAGNOSIS LIKE 'Multiple Myeloma';

--Yearly multiple Myeloma diagnosis
SELECT YEAR(INPATIENT_SPELLS.ADMISSION_DATE) AS YEAR, COUNT(INPATIENT_SPELLS.PATIENT_ID) AS Cases FROM DIAGNOSES
JOIN INPATIENT_SPELLS
ON DIAGNOSES.SPELL_ID = INPATIENT_SPELLS.SPELL_ID
WHERE DIAGNOSIS LIKE 'Multiple Myeloma'
GROUP BY YEAR(INPATIENT_SPELLS.ADMISSION_DATE);
```

- **Result**:
  The first query retrieves all inpatient spells with a diagnosis of Multiple Myeloma.
  The second query provides a yearly count of distinct spells including this diagnosis, offering insights into the incidence of the disease over time.

Crohn’s Disease

Crohn’s Disease is a chronic inflammatory bowel disease that affects any part of the gastrointestinal tract. It leads to inflammation, ulcers, and other complications, causing symptoms such as abdominal pain, diarrhoea, weight loss, and fatigue.

Testing

Diagnosis involves a combination of medical history, physical examination, laboratory tests, imaging studies, and sometimes biopsy of affected tissues.

Target

Analysis to identify immunocompromised Crohn’s Disease patients to provide them with urgent medical advice regarding COVID-19 precautions.

SQL Query

- **Objective**:
  Generate a list of immunocompromised Crohn’s Disease patients for urgent medical advice, based on specific ICD-10 codes (K500, K501, K508, K509)

- **CODE:**

```
SELECT * FROM(
(INPATIENT_SPELLS JOIN DIAGNOSES ON DIAGNOSES.SPELL_ID = INPATIENT_SPELLS.SPELL_ID)
LEFT JOIN PATIENTS ON PATIENTS.PATIENT_ID = INPATIENT_SPELLS.PATIENT_ID)
WHERE DIAGNOSES.DIAGNOSIS_CODE IN (’K500’, ’K501’, ’K508’, ’K509’);
```

- **Result**: This query retrieves data of patients diagnosed with Crohn’s Disease, allowing the medical team to identify and advise immunocompromised patients regarding COVID-19 precautions.

Further Analysis

Excel

Utilise Excel to import and organize clinical data, conduct exploratory analysis, and create visualizations for initial insights.

Power BI

Leverage Power BI for advanced analytics and interactive dashboards, enabling dynamic exploration and presentation of key findings.

For Health Data Analysis

Connect with us.

ibrahim.oshinubi@gmail.com

tobunlateefat@gmail.com