BCDE103 Design Assessment

# Data Dictionary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Patient Table | | | | | |
| Column Type | **Column Name** | **Data Type** | **FK Relationship** | **Constraints** | **Description** |
| PK | NHI | CHAR(7) |  |  |  |
|  | FirstName | VARCHAR(20) |  |  |  |
|  | LastName | VARCHAR(30) |  |  |  |
|  | DOB | DATE |  | Format: YYYY-MM-DD |  |
|  | Gender | VARCHAR(6) |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Department Table | | | | | |
| Column Type | **Column Name** | **Data Type** | **FK Relationship** | **Constraints** | **Description** |
| PK | DepartmentID | CHAR(2) |  |  |  |
|  | DepartmentName | VARCHAR(20) |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Referee Table | | | | | |
| Column Type | **Column Name** | **Data Type** | **FK Relationship** | **Constraints** | **Description** |
| PK | RefereeID | CHAR(4) |  |  |  |
|  | FirstName | VARCHAR(20) |  |  |  |
|  | LastName | VARCHAR(30) |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Origin Table | | | | | |
| Column Type | **Column Name** | **Data Type** | **FK Relationship** | **Constraints** | **Description** |
| PK | OriginID | CHAR(2) |  |  |  |
|  | Origin Name | VARCHAR(20) |  |  | Where the referral was generated |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Surgeon Table | | | | | |
| Column Type | **Column Name** | **Data Type** | **FK Relationship** | **Constraints** | **Description** |
| PK | SurgeonID | CHAR(6) |  |  |  |
|  | FirstName | VARCHAR(20) |  |  |  |
|  | LastName | VARCHAR(30) |  |  |  |
| FK | DepartmentID | CHAR(2) | Department |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Referral Table | | | | | |
| Column Type | **Column Name** | **Data Type** | **FK Relationship** | **Constraints** | **Description** |
| PK | ReferralID | INT |  | Starts at 1000 |  |
| FK | NHI | CHAR(7) | Patient |  |  |
| FK | SurgeonID | CHAR(6) | Surgeon |  |  |
| FK | OriginID | INT | Origin |  |  |
| FK | RefereeID | CHAR(2) | Referee |  |  |
|  | ReferralDate | DATE |  |  |  |
|  | WaitlistDate | DATE |  |  |  |
|  | FSA | DATE |  |  |  |
|  | AgeAtReferral | INT |  |  |  |
|  | DaysOnWaitlist | INT |  |  | Time between referral date and FSA date |

# Design Report

Removed the Year-Month column

Removed the health target column

Null values set to a date 80 days in the future

## Data Cleansing

* NHI numbers are used as a unique identifier for a patient, so rows with duplicate NHI numbers are removed from the data. The reason for removing all duplicates rather than just one is because there is no certain way to determine which of the duplicates is more legitimate and therefore able to be kept.
* DOBs are sanity checked and must fall between 1903, the birth year of the oldest person currently alive, and the current year and month, August 2021. 2 birth dates fall outside of this range, one in 2027 and one in 1756. These rows are removed from the data set.
* There is an unnamed column on the end of the data set containing dates. As there is no heading for the column the purpose of the dates is unknown and therefore the column has been removed.
* Some department names have been entered with trailing whitespace, which has been trimmed.
* The column Year-Month is redundant, as the data is a duplicate of the data in Referral Date and is thus removed.
* Patients who are not eligible for the health target are removed.
* Given the data set and business rules stating that patients who are ineligible for the health target are ignored, the Health Target Eligible column is removed from the final data set, as the column is made redundant by the business rule.

## Entities, Attributes, and Keys

I initially settled on 5 entities:

* Patient
* Department
* Referee
* Surgeon
* Referral

This seemed to be the smallest number of entities I could use, but later Referee was split into 2 entities, being Referee and Origin. This is because a referee is not locked in to one and only one referral origin location. For a given Referee, they could refer one Patient from GP and another from Internal, meaning that the referral origin cannot be a part of the Referee entity.

I chose to use the NHI as the primary key for the Patient Table, and “generate” unique keys for the other entities. The SurgeonID and RefereeID are generated from initials of the person’s name. The DepartmentID is generated using the initials of the name of the department. The OriginID and ReferralID are both incrementing integers, starting from 1 and 1000, respectively.