



We have 5 total entities in our ER model: User, Past and Current Conditions, Biometrics, Diagnosis, and Lifestyle. User is the only strong entity in the dataset, with everything else being a weak entity dependent on User because they explain different things about the user. The entities are separate from each other based on their domains and categories. For example, User stores a user's ID as its primary key (listed as PatientId in the ER, but essentially any distinguishing ID will suffice), age, gender, and country make up a user's "personal" profile. The Diagnosis entity is made of risk and diagnosis attributes for the user; this is in a 1-1 relationship with user because we expect each user to have only one diagnosis at a time. ID acts as a foreign key from Diagnosis to User. There is no use in having records of past diagnostics. The Conditions entity is made up of mainly boolean valued attributes that signify whether the user has certain conditions or not. Lifestyle also mainly contains boolean attributes that detail whether a user has certain lifestyle choices. The Biometrics entity tracks biometric values for a user (BMI, blood glucose levels, etc) that take positive float values. EACH of these entities is in a 1-Many relationship with User because we expect each user to have multiple Biometrics, Lifestyle, and Conditions records that they update over time. Having multiple records for each user should add predictive value to the application. We add a date attribute to each that, together with ID as the foreign key to User, make up a composite key uniquely identifying records in each entity table.

Relational Schema:

User(PatientId:VARCHAR(64)[PK], Country:VARCHAR(64), Age:INTEGER, Gender:BIT(1))

Lifestyle(PatientId:VARCHAR(64)[PK,FK to User.PatientId, Date:VARCHAR(16)[PK], Smoking History:BIT(1), Smoking Frequency:BIT(1), Cholesterol Checked:BIT(1), Physical Activity:BIT(1), Fruit Consumption:BIT(1))

Diagnosis(PatientId:VARCHAR(64)[PK, FK to User.PatientId), Diabetic:BIT(1), Diabetes Risk:INT)

Biometrics(PatientId:VARCHAR(64)[PK, FK to User.PatientId), Date:VARCHAR(16)[PK], BMI:REAL, HbA1c Level:REAL, Blood Glucose Level:REAL)

Past/Current Conditions(PatientId:VARCHAR(64)[PK,FK to User.PatientId, Date:VARCHAR(16)[PK], Hypertension:BIT(1), Heart Disease:BIT(1), High Blood Pressure:BIT(1), High Cholesterol:BIT(1), Stroke:BIT(1), Heart Attack:BIT(1))

This schema is in 3NF because PatientId acts as a primary key and in some cases PatientId and Date together act as a primary key. All non-key attributes such as "BMI" and "Hypertension" are dependent on PatientId/Date (dependent on single primary key or composite key). Therefore there are no transitive dependencies.