**United States Drug Data and Transparency: Database Schema Construction**

Group 2

Gabriel Fernandez, Ashleigh Hillebrand, and Joshua Main-Smith

*Department of Computer Science, University of Florida*

*Gainesville, FL*

Table of Contents

[I. Introduction 2](#_Toc21879361)

[II. Conceptual Database Design 2](#_Toc21879362)

[III. Relational Database Schemas 3](#_Toc21879363)

[IV. SQL Table Schemas 4](#_Toc21879364)

[References 10](#_Toc21879365)

I. Introduction

We intend to create an application that will make it simple for users to easily compare varying parameters of drugs in the United States. This application will provide transparency to the relationships between drug information from the Food and Drug Administration (FDA) National Drug Code (NDC) Database [1], drug pricing based on the National Average Drug Acquisition Cost (NADAC) [2], and state insurance programs’ drug reimbursement from the CMS State Drug Utilization Data from 2014-2018 [3]. Now that we have the overall design, we must transform the conceptual database design into a relational database schema and transfer this to SQL table schemas.

II. Conceptual Database Design

A picture containing text, map

Description automatically generated

Figure 1: ER Diagram

Figure 1 shows the updated ER diagram. The NDC dataset contains information about the type, proprietary and nonproprietary names, ingredients of the drug to include strength, dosage, and administration information, the proprietary company, and the packaging information for each drug. We have divided company, ingredient, and packaging into separate entities with direct and specific relationships comprising all information about a drug. The NDC code or primary key identifying a drug in its own entity and in both State Drug Utilization and NADAC datasets is an eleven digit code made up of the five digit label code from “Company”, four digit product code from “Ingredient”, and two digit package size from “Package.” Because the State Drug Utilization datasets contain information about state insurance program’s payment of a drug, it is its own entity labeled as “Insurance.” The entity “Pharmacy” represents attributes derived from NADAC. It is labeled “Pharmacy” because NADAC’s data is based on the random survey of acquisition cost data from pharmacies across all 50 states and the District of Colombia. We believe that this division of entities and their relationships prove most logical and applicable for the web application we intend to develop.

III. Relational Database Schemas

Listed below is the complete drug data schema.

* company(label-code : integer, labeler-name : string, start-market : string, end-market : string, market-category-name : string)
* ingredient(product-code : integer, sub-name : string, strength-no : string, strength-unit : string, dosage : string, route : string, pharm-class : string)
* package(package-size : integer, package-desc : string, sample : integer)
* drug(ndc : integer, product-type : string, prop-name : string, non-prop-name : string)
* pharmacy(ndc : integer, pharm-type : string, prop-price-per-unit : decimal, pricing-unit : string, prop-price-effective-date : date, non-prop-price-per-unit : decimal, non-prop-price-effective-date : date)
* insurance(ndc : integer, util-type : string, state : string, year : string, quarter : string, suppression-used : character, units-reimbursed : decimal, no-of-prescriptions : integer, total\_reimbursed : decimal, medicaid-reimburse : decimal, non-medicaid-reimburse : decimal)
* markets(ndc : integer, label-code : integer)
* forms(ndc : integer, product-code : integer)
* contains(ndc : integer, package-size : integer)
* prices(ndc : integer)
* pays(ndc : integer)

IV. SQL Table Schemas

A screenshot of a social media post

Description automatically generated

Figure 2: Oracle Create Table Command (Company)

A screenshot of a social media post

Description automatically generated

Figure 3: Empty Table (Company)

A screenshot of a social media post

Description automatically generated

Figure 4: Oracle Create Table Command (Ingredients)

A screenshot of a computer

Description automatically generated

Figure 5: Empty Table (Ingredients)

A screenshot of a social media post

Description automatically generated

Figure 6: Oracle Create Table Command (Package)

A screenshot of a computer

Description automatically generated

Figure 7: Empty Table (Package)

A screenshot of a social media post

Description automatically generated

Figure 8: Oracle Create Table Command (Drug)

A screenshot of a social media post

Description automatically generated

Figure 9: Empty Table (Drug)

A screenshot of a social media post

Description automatically generated

Figure 10: Oracle Create Table Command (Pharmacy)

A screenshot of a social media post

Description automatically generated

Figure 11: Empty Table (Pharmacy)

A screenshot of a social media post

Description automatically generated

Figure 12: Oracle Create Table Command (Insurance)

A screenshot of a computer

Description automatically generated

Figure 13: Empty Table (Insurance)

# References

|  |  |
| --- | --- |
| [1] | FDA, "National Drug Code Directory," 09 November 2017. [Online]. Available: https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory. [Accessed 21 September 2019]. |
| [2] | Centers for Medicare and Medicaid Services, *NADAC (National Average Drug Acquisition Cost),* 2019. |
| [3] | Centers for Medicare and Medicaid Services, "State Drug Utilization Data," [Online]. Available: https://data.medicaid.gov/browse?q=state%20drug%20utilization%20data&sortBy=relevance. [Accessed 21 September 2019]. |
|  |  |