

The overall goal in the whole mini-project is to create and work with a medical database in several incremental steps. The first step is to identify the information that is going to be represented as well as special requirements. Then, the information is modeled in an ER diagram and mapped to a relational schema. The database schema is then refined and normalized. Later on, database dumps will be provided and queried.

Database Modeling

The database should store information about drugs, their categories, the diseases they can treat, their side effects, their commercial products, their interactions with other drugs, and possibly information about related clinical trials. In general, you may use the WebMD website (<https://www.webmd.com/>) for inspiration. But be aware that WebMD stores a large variety of information and that some information there is quite verbose. You should not try to incorporate all the information available in WebMD into your database design! For most aspects, it will be sufficient to have 5 attributes instead of 20. Please choose carefully.

Sample Information

To provide you with an idea of how the information to be loaded in the database might look like, consider the following sample information available in the given excel file. Please notice that the sample information does not correspond to what would be a database table (or relation).

Your database model should answer the following queries:

- a) Find the number of drugs that have nausea as a side effect
- b) Find the drugs that interact with butabarbital
- c) Find the drugs with side effects cough and headache
- d) Find the drugs that can be used to treat endocrine diseases
- e) Find the most common treatment for immunological diseases that have not been used for hematological diseases
- f) Find the diseases that can be treated with hydrocortisone but not with etanercept
- g) Find the top-10 side effects that drugs used to treat asthma related diseases have
- h) Find the drugs that have been studied in more than three clinical trials with more than 30 participants
- i) Find the largest number of clinical trials and the drugs they have studied that have been active in the same period of time

- j) Find the main researchers that have conducted clinical trials that study drugs that can be used to treat both respiratory and cardiovascular diseases
- k) Find up to three main researchers that have conducted the larger number of clinical trials that study drugs that can be used to treat both respiratory and cardiovascular diseases
- l) Find the categories of drugs that have been only studied in clinical trials based in United States

Report

In Task 1, you have already designed this database based on what you already knew back then. Now, try it again based on what you have learned so far in the course – please use the notation introduced in the course (slides) and solve the following tasks:

1. Identify relevant data (consider the sample queries included in this document).
2. Create an ER diagram with additional information about cardinality (Chen and [min,max]), roles if necessary, primary, candidate, and foreign keys. Do not blindly add ID attributes to all entity types. Instead, think about native attributes that could be used as primary keys. Please add a (short!) description explaining non-trivial design choices that you have made.
3. Create an appropriate relational model – you do not yet have to consider functional dependencies and normalization.
Do not create extra relations if an alternative solution with fewer relations is possible!
You may ignore attribute domains but do not forget to mark the primary keys and foreign keys. You have to include:
 - Names of relations
 - Attribute names (optionally also data types)
 - Primary keys and candidate keys
 - Foreign keys
4. Compare your solution to the results of self study 1 and reflect on the differences. Examples of questions you might want to consider are:
 - What are the differences to your initial design?
 - Why did these differences occur?
 - What advantages does the new design have over the old one?

The created relational schema will be the basis for the following self study slots.

Course goals covered by this self study

- Conceptually design a database (ER model, conceptual design)

