There will be an SQL exercise on a small subset of the Children's Hospital database, from about five years ago. These data have been "scrubbed", so that specifically identifying information has been altered or removed. Nevertheless, please treat the data with sensitivity, because you will know from our discussion of medical data confidentiality that even innocuous-seeming data can help re-identify.

The database is not in a perfect shape. The purpose of using such a database is to get a sense of dealing with the real-world problems: in many cases, you won't see neat and organized data. Rather, it can be fumbled, distorted, and redundant. It will be your job to dig everything out.

The database has 19 tables (including table pat_test_histv_new, which we derived from the original table to separate numerical and text test values). The tables can be grouped into the following categories:

Document tables:
- DOC_ATTRIBUTES: visit-time, dictation-time, and comments of master documents.
- CHILD_DOCS: this table links a master document to its sub-documents.
- DOC_DESCRIPTION: basic information of sub-documents, including date/time, related provider and patients, etc.
- DOC_STORE: full content of sub-documents.

Patient-provider tables:
- PAT_DEMOGRAPHY: basic biographical information of patients.
- PERSNL_PUBLIC: basic information of providers.
- PPR: describes the relationships between providers and patients.

Data tables:
- PAT_FIN_ACCT: detailed information of patient financial encounter.
- CLINICAL_DATA: results of certain exams
- PROBLEMS: medical problems of patients
- PHARMACY_TABLE: meds of patients
- PAT_TEST_HISTV: detailed information of lab tests, including test name, patient, results, etc.
- REMOTE_TEST: information on remote tests
- PAT_TEST_HISTV_NEW: this table was derived from PAT_TEST_HISTV. All other fields are the same except 3 fields RSLT_VAL, REF_HIGH_VAL, REF_LOW_VAL were split into numerical and text values, and thus 6 fields: fields RSLT_VAL_NUM, REF_HIGH_VAL_NUM, REF_LOW_VAL_NUM, RSLT_VAL_TXT, REF_HIGH_VAL_ TXT, REF_LOW_VAL_ TXT. We split these values to make it a easier job comparing the test results with the references (if they are numbers) to judge if a test result is within normal range, because the field ABN_STATUS doesn't always give the correct indication.

Other tables:

- ICD9: ICD9 diagnosis code
- ICD9_PROCDR: ICD9 procedure code
- PROBLEMS_NOSOLOGY: problem nosology
- CPT_CODE: CPT code
- CLINICAL_DATA_TYPE: data types in CLINICAL_DATA

Here are a few simple examples of extracting information from this database:

Example 1:
Suppose we want to look up the information of a patient Bart Park. First, we can find the patient number in PAT_DEMOGRAPHY by performing a query:

Select * from pat_demograph where last_name = 'park';

From the result, we see the patient number is 200, as well as other demographic information of this patient. Now we can look for the provider(s) of this patient:

Select * from ppr where pat_num = '200';

This query returns 10 rows, and we can see there are at least 3 different providers were associated with this patients.
We can also look at the medical problems of the patient and find he has X-LINKED-HYPOPHOS.-RICKETS:

Select * from problems where pat_num = '200';

Now we can look for the related documents:

Select * from doc_description where pat_num = '200';

Here we find 4 documents signed by the three providers, while the full content of the documents can be found in table DOC_STORE.

Example 2:
Suppose we want to look at the list of Dr. Shields' patients. We can find this information by joining three tables:

Select pat_demograph.last_name, pat_demograph.first_name
from (persnl_public inner join ppr on persnl_public.persnl_id=ppr.provider_id)
inner join pat_demograph on ppr.pat_num=pat_demograph.pat_num
where persnl_public.last_name = 'shields';