Assignment 1: SQL

Welcome to the first assignment of BMI 701! This assignment is designed to give you a chance to extend your SQL skills and also explore some interesting questions. A few notes on the assignment:

- You need to work on this assignment on your own. Feel free to discuss general SQL questions and issues with your classmates, but do the work by yourself. Definitely don't share any SQL code with others. If you have any questions, try to explore some of the many SQL resources on the web, or ask Adam.
- All the questions here are designed to be answered using SQL queries and your own thinking. Resist the temptation to paste a lot of data into Excel or another program to get to the answers. You're welcome to use other programs to extend your answers (e.g. to make a graph) but the main part of your computation should be done in SQL.
- SQL is a standard, but there are some quirks across implementations. The class server is running MySQL 5.5 The documentation for this version is here: http://dev.mysql.com/doc/refman/5.5/en/.
- For each question, please give the query you used, the results and your interpretation of them. Your answers don't need to be long but they should be thoughtful.
- The assignment has for questions – each question is weighted equally, and you can earn partial credit on a question if you're on the right track, so please show your work and explain your thinking.
- You can access the class server at http://sql.adamwright.com

Question 1:
This question has some basic queries to help you warm up. For these questions, you can just give the query you used and the result. For later questions, you'll need to give some more interpretation.

a) How many female patients are in the database?
b) How many patients have diabetes in their past medical history?
c) How many patients have arthritis in their past medical history and are taking the drug celecoxib?
d) What is the most common visit reason?
e) How many patients got both their glucose and the HbA1c checked?

Question 2:
In 2003, Elizabeth McGlynn published a widely cited paper that showed that adults in the US only received 54.9% of guideline-recommended care[1]. Let's see if this has improved by looking at the following guideline recommendations. For each one, calculate the proportion of patients that have the problem in their past medical history who received the recommended intervention.

- Diabetic patients should have their HbA1c checked
- Patients with hypertension should have their systolic blood pressure assessed
- Patients with hyperlipidemia should be on a lipid lowering drug (for this, use the drugclasses table, looking for 'Metabolic agents; antihyperlipidemia agents' in level 2)

It's probably easiest to use one SQL query for each measure, and a separate one to compute the measure's denominator – there's no need to try and so simultaneously.

Then, pick a demographic characteristic (age, sex, region) and see how it influences these quality measures – in this case, please report the quality measures again, broken out by your characteristic.

[1] McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. "The quality of health care delivered to adults in the United States." N Engl J Med 348(26): 2635-45.

Question 3:
Blood pressure is measured on a continuous scale, but several studies have found that providers demonstrate an "end digit preference", preferring blood pressure numbers that end in zero[2] and that this preference can result in misclassification – causing patients to receive unneeded therapy or miss needed therapy. Is there evidence of an end-digit preference in systolic blood pressure measurements in this dataset?

Hint: there are at least two ways to solve this problem – one using a string function (section 12.5 of the MySQL manual) and one using a numeric function or operator (section 12.6 of the MySQL manual).

Question 4:
Think of an interesting question that you could explore using this data and SQL. Explain the question you're trying to answer, why it's important and how you decided to explore it. Please pick a question that involves at least a couple of tables and some thinking – more like Questions 2 or 3 than Question 1.

[2] Broad, J., S. Wells, R. Marshall, and R. Jackson. 2007. "Zero end-digit preference in recorded blood pressure and its impact on classification of patients for pharmacologic management in primary care - PREDICT-CVD-6." Br J Gen Pract 57(544): 897-903.