**Background:** [CMS](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DESample01) publishes synthetic claims and beneficiary data available for the public. We are providing 3 beneficiary sample files and 1 outpatient claims file for analysis. In the outpatient claims file, the HCPCS CPT Codes for ***office visits*** are identified by the ranges: 99211-99215. The last digit of the code represents the *office visit severity*. Office visits tend to be $50-75 more expensive per level of severity. There’s a deidentified patient ID that connects claims with beneficiaries.

**Goal:** “Upcoding” occurs when a provider submits codes for more serious services than are actually necessary or performed. As a health plan, we want to understand if there are any providers that tend to consistently upcode on the level of severity of an office visit. Or, at the very least, to estimate our exposure to this type or *waste or abuse,* which is a hundred-billion-dollar problem for the nation.

**Content:** Propose and implement a plan to, on an ad hoc basis, analyze this set of data to understand the problem above. What is the exposure to upcoding risk in this data set? Can we hone-in on individuals? What analytics/data science techniques can help us build a business case to stakeholders? Spend no more than 3-4 hours working on this, it’s not expected to be a final solution or production caliber.

**Deliverable:** Provide code (R or Python or anything) that gives insight into how you structured your analysis and then put together 3-4 slides that would explain how you would build your case to the business to invest in the analysis and a program to deal with it. You’ll be asked to walk through your presentation in your panel interview. Code files and slides should be sent 24 hours in advance.

Please feel free to send clarifying questions to [sjaeger@healthfirst.org](mailto:sjaeger@healthfirst.org) as you’re working on the project.

Necessary but not sufficient files:

[Beneficiary 1](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/DE1_0_2008_Beneficiary_Summary_File_Sample_1.zip)

[Beneficiary 2](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/DE1_0_2009_Beneficiary_Summary_File_Sample_1.zip)

[Beneficiary 3](https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/SynPUFs/Downloads/DE1_0_2010_Beneficiary_Summary_File_Sample_20.zip)

[Claims](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/DE1_0_2008_to_2010_Outpatient_Claims_Sample_1.zip)

[CMS Data User Manual](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/SynPUF_DUG.pdf)