



# Understanding Statistical Sampling and Extrapolation

John L. Adams, Ph.D.



# Outline

- What is statistical sampling?
- What are the steps in the sampling process?
- How is extrapolation done?



# What is statistical sampling?

- Statistical sampling is a name used to underscore that a sample is scientifically designed to support estimation of the original population
- Statistical sampling uses random sampling to ensure that we can extrapolate to the original population
  - Specifically excludes convenience sampling and purposive sampling
  - It doesn't have to be simple random sampling
- Common uses include opinion polls, quality control, medical studies



# Chapter 8 of the PIM

The main source of sampling information for Medicare is:

Medicare Program Integrity Manual

Chapter 8 – Administrative Actions and Statistical Sampling for Overpayment Estimates



# Step #1: Define the universe and the sampling frame

- The universe is the population for which you would like to estimate the overpayment
  - It could be beneficiaries, particular types of claims, or visits
- The sampling frame is the specific and detailed list of elements we will use to draw the sample
  - This should be well documented and saved



## Step #2: Design the sampling plan

- Pick a design
  - Simple random sample
  - Stratified random sample
  - More advanced samples
- Determine the sample size (and allocation if needed)



## Step #3: Draw the sample

- It is essential that this be reproducible
- You need software to do this, SAS and RAT-STATS are popular
- It is important to save the computer code and the random number seed(s)
- The sample is then used to determine the overpayments



# How is extrapolation done?

- In the case of a simple random sample (overpayment):
  - Calculate the mean (average) overpayment in the sample ( $X$ )
  - Calculate the standard error of the mean (SE)
  - Calculate the lower confidence limit (LCL) of a one sided confidence interval:  $X - t * SE$  (typically we use the  $t$  for a 90% one sided interval)
  - This is the LCL for the average so multiply by the size of the population to get the LCL for the population
- For more complicated sampling designs you need to use statistical software but the logic is the same



# HOLD THAT THOUGHT—



# QUESTIONS COME LATER