

Medicaid Audits: Challenging Statistical Sampling and Extrapolation

WEDNESDAY, AUGUST 29, 2018

1pm Eastern | 12pm Central | 11am Mountain | 10am Pacific

Today's faculty features:

Dr. Patricia L. Maykuth, Ph.D, President, **Research Design Associates**, Decatur, Ga.

David R. Ross, Senior Shareholder, **O'Connell and Aronowitz**, Albany, N.Y.

The audio portion of the conference may be accessed via the telephone or by using your computer's speakers. Please refer to the instructions emailed to registrants for additional information. If you have any questions, please contact **Customer Service at 1-800-926-7926 ext. 1.**

TERMS Used in the “MEDICARE & MEDICAID AUDITS USING STATISTICAL SAMPLING & EXTRAPOLATION” Webinar June 2018

These are a list of terms and definitions to explain some of the statistical concepts in this webinar. They are working definitions to aid in understanding the statistics used in CMS audits. There is no attempt to provide the most statistically authoritative or mathematically precise definition.

MPIM CMS Medicare Program Integrity Manual Chapter 8 is 19 page of verbal guidance for using statistical sampling and extrapolation.

RAT-STATS – free DHHS provided statistical package to calculate sample size, random number tables, and extrapolations

Numbers can readily be manipulated and outcomes understood through the use of simple math: addition, subtraction, multiplication and division, e.g., percentages, differences, sums, averages and back of the envelope estimation.

Statistics is a branch of applied math concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate universe parameters, e.g. correlations, *t*-tests and point estimates

INFERENCE STATISTICS is a branch of applied statistics drawing conclusions about a universe/frame from a random sample drawn from it. These mathematical analyses move beyond mere description of research data to make inferences about the larger population from which the sample was drawn.

Probability statistics Statistical analysis that uses probability theory to generate and properly interpret inferences. Probability theory is the mathematical basis of those distributions tested repeatedly and compared to random outcomes.

Parameter refers to a DISTRIBUTIONAL characteristic of a frame

- **Parametric Statistics** are probability estimates based on the parameters of a normal distribution. Parametric tests make specific assumptions about the population parameters that characterize the underlying distributions for that test
- **Non Parametric Statistics** tests make few or no assumptions about the underlying distribution of the and parameters of the population
- **Parametric tests** make specific assumptions about the population parameters that characterize the underlying distributions for that test
- **Non parametric** statistics make few or no assumptions about the underlying distribution of the parameters of the population

Samples

Sampling unit – the unit of measurement for the study: claim, beneficiary, specific payment codes. The auditor must stick with the chosen unit and not switch back and forth

Universe – claims paid to a Provider in a specific timeframe

Pat Maykuth, PhD
Research Design Associates
721 E Ponce de Leon
Decatur, GA 30030

June 2018
www.researchdesignassociates.com
(404) 373-4637

Sampling Frame – subset of the universe defined as variables of interest from which the sample will be randomly selected and over which the sample will be extrapolated

Sample – purportedly a randomly selected subset of sampling frame to be audited for overpayments
Unit of Analysis (Sampling Unit) – what is measured in the audit: claim line, claim, beneficiary, provider (must be invariant throughout the audit)

Random sample In mathematics and statistics random means: having a value which cannot be determined but only described probabilistically as a random variable, chosen without regard to any characteristics of the individual members of the population so that each has an equal chance and known probability of being selected randomly. It is not the lay meaning of “haphazard”.

Representative – an unbiased sample that accurately reflects the numerical membership of the entire universe and its distribution. It must cover all salient features of the universe (without overlapping segments) to be a true picture of the universe from which it was selected and over which projections will be made. The distribution of the sample should represent the distribution of the frame.

SVRS Statistically Valid Random Sample selected from a frame of paid Medicare or Medicaid claims guards against cherry picking or any bias by the audit and will be an accurate estimator it (and only if): it meet the requirements the methodology – especially assumptions about distribution; proper statistics are used to measure it; it meets chosen sampling error; it is of sufficient size to accurately measure the variable;
is random; representative (without bias); and address the impact of non-sampling errors

Extrapolation

Extrapolation takes the results of an audited sample and projects the dollar amount of the claims in the universe of claims paid.

Null Hypothesis a statistical hypothesis that is tested then accepted or rejected; *specifically* : the hypothesis that an observed overpayment amount is due to chance alone and not due to a systematic or biasing cause

Distribution the measurement yardstick of statistics. It is a description of the relative numbers of times each possible outcome will occur in a number of trials. The frequency outcome forms a frequency distribution that is compared to a theoretical distribution. A mathematical function describing the probability that a given value will occur is called the probability function

Normal Distribution is the bell-shaped distribution necessary for use of parametric statistics. It is continuous probability distribution (a function that tells the probability of a number in some context falling between any two real numbers). The normal distribution is symmetric around the mean. The mean, median and mode are the same number. Its use allows two key theories of probability to be used (the theory of large numbers and the Central Limit Theorem).

Measures of central tendency In a normal distribution these three are the same number

- **Mean** (average) the arithmetic sum of all scores divided by the number of cases
- **Median** – the middle most real score in the data set
- **Mode** the score that occurs most frequently in the data set (does not have to be unique – sometimes more than one value is equally likely)

Independence Two events are independent if the occurrence of one event makes it neither more nor less probable that the other occurs.

Point Estimate uses sample data to calculate a single point (mean) which serves as the best estimate of a universe parameter

Confidence level – upper and lower probability level around a mean
 Point Estimate +/- Precision amount = Confidence Level

Error

Precision - measurement of variability

Precision amount ½ the confidence level

Precision % = Precision amount/Point estimate

Variance - distance between each set of data points and their mean

Standard deviation – square root of the variance

Error rate – number of claims in error or dollars in error

Error rate – can measure number of dollars in error or dollars in error

Prior history of error requirement to determine sample size

Sampling error – getting a poor sample yielding poor estimate The difference the sample and the frame being extrapolated over
Non sampling error – claims not in error labeled as in error, mistakes in coding, reporting, stratification, inaccurate documentation.