Strafford

Presenting a live 90-minute webinar with interactive Q&A

Medicare and Medicaid Audit Overpayments: Challenging Statistical Sampling and Extrapolation

TUESDAY, SEPTEMBER 22, 2020

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, NY

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 AND MEDICAID AUDIT OVERPAYMENTS: CHALLENGING STATISTICAL SAMPLING AND EXTRAPOLATION" Webinar September 22,2020

This is a list of terms and definitions that help explain some of the statistical concepts in this webinar. They are working definitions to aid in understanding the statistics used in Medicare and Medicaid audits. This is not an attempt to provide the most statistically authoritative or mathematically precise definition.

MPIM CMS Medicare Program Integrity Manual Chapter 8 is 19 pages of verbal guidance for using statistical sampling and extrapolation. It was revised in 2019. https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/pim83c08.pdf

RAT-STATS is a free DHHS statistical software package used to calculate sample size, random number tables, and extrapolations. https://oig.hhs.gov/compliance/rat-stats/index.asp

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.

Inferential 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 is 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.

Probability space is the mathematical structure of a probability model (frame) that includes all possible samples (given size from given frame size) from a family of events (claims). A probability can be assigned to each event (claim).

Parameter refers to a DISTRIBUTIONAL characteristic of a frame.

- **Parametric Statistics** are probability estimates based on the parameters of a normal distribution.
- **Non Parametric Statistics** test make few or no assumptions about the underlying distribution of the population.

Universe – claims paid to a Provider in a specific timeframe.

Research Design Associates
721 E Ponce de Leon
Decatur, GA 30030
www.researchdesignassociates.com
(404) 373-4637

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

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 the sampling frame to be audited for overpayments.

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."

Seed is a number (or vector) used to initialize a pseudorandom number generator**.** A pseudorandom number generator's number sequence is completely determined by the seed. Thus, if a pseudorandom number generator is reinitialized with the same seed, it will produce the same sequence of numbers.

Representative sample is unbiased subset that 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.

Statistically Valid Random Sample (**SVRS)** is a 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 if (and only if): it meets the requirement of the chosen 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 addresses the impact of non-sampling errors.

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

Null Hypothesis a statistical hypothesis that is tested and 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 a 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

Research Design Associates www.researchdesignassociates.com
721 E Ponce de Leon (404) 373-4637
Decatur, GA 30030

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).

Central Limit Theorem (CLT), in probability theory establishes that in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (informally a bell curve) even if the original variables themselves are not normally distributed. The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions. For an excellent explanation, please see the Khan Academy explanation. https://www.khanacademy.org/math/ap-statistics/sampling-distribution-ap/sampling-distribution-mean/v/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 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 that universe parameter.

Confidence level – upper and lower probability level around a mean.
Point Estimate +/- Precision amount = Confidence Level.
MPIM says the lower 90% one-sided confidence level should be adequate.

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 –The difference between the sample and the frame upon which the sample is extrapolated over
Non sampling error – claims not in error labeled as in error, mistakes in coding, reporting, stratification, inaccurate documentation.

Research Design Associates
721 E Ponce de Leon
Decatur, GA 30030
www.researchdesignassociates.com
(404) 373-4637