# Statistical Sampling in Healthcare Audits and Investigations

Michael Holper
SVP Compliance and Audit Services
Trinity Health

Stefan Boedeker
Managing Director
Berkley Research Group LLC



## Topics for Today's Discussion

- Use of statistical sampling and extrapolation of overpayment liabilities in health care audits and investigations
- Medicare Program Integrity Manual (PIM): CMS guidance on use of statistical sampling in audits
- **Breaking News:** Recent CMS update to Medicare PIM
- Fundamentals of proper statistical sampling and common errors identified in applying to health care audits and investigations
- Experiences in defending health care providers against use of flawed statistical sampling and extrapolation models

## Use of Statistical Sampling is wide-spread in health care audits and investigations

- OIG Audits and Investigations (State and Federal)
- DOJ Investigations
- False Claims Act Cases – Whistleblower/Relator or Government Intervenes
- ZPIC Audits
- Corporate Integrity Agreements
- Private Pay Disputes
- CMS Contractors



## OIG Audits and Investigations (State and Federal)

- OIG statistical sampling documentation typically gives detailed instructions to users
- Use of a statistician in the sampling process from design to extrapolation is recommended, but not mandated
- Result: sampling instructions may be applied incorrectly by users
- Biggest drawback
  - OIG typically does not share “workpapers” with auditees
  - Intermediate reports with extrapolated liabilities circulate early
  - “Final” reports can take extended time

## DOJ Investigations: False Claims Act Cases Whistleblower/Relator or Government Intervenes

- DOJ inquiry initiated based on internal information
- Invitation to present information against allegations without specificity
- First response may or may not end DOJ inquiry
- Next stage, specific documents are requested
- Request based on a statistical sample designed by an outside expert at direction and specification of DOJ
- Often leads to lengthy and costly defense costs and potential litigation



## ZPIC Audits

- ZPIC Auditors autonomy to conduct investigations into alleged overpayments
- Frequently driven by data mining, deviation from threshold “norms”, outlier providers targeted for audit
- Aggressive coding reviews conducted (>90% “error rates” are not uncommon)
- ZPICs are required to provide documentation of statistical sampling and extrapolation with repayment demand letters
  - May not always provide information without insistence by provider
- ZPIC auditors may have varying levels of statistical knowledge; less experienced auditors may mechanically attempt to follow Medicare Program Integrity Manual (“PIM”)
- ZPIC Auditors sending statisticians to administrative appeal hearings more frequently

## Corporate Integrity Agreements

- Independent Review Organizations (IRO) typically follow CMS guidelines for statistical sampling as defined in Medicare PIM
- In most cases, a two stage sampling process is required:
  - Discovery, probe or pilot sample to determine the sample size to achieve a pre-determined confidence level and margin of error
  - Full sample to determine actual overpayments based on extrapolations
- If the financial error rate in the probe sample is below a certain threshold, then a full sample may not be required
- Only overpayment amounts in the probe sample will require repayment



## Disputes with Private Payers

- Providers and private payers increasingingly arbitrate their disputes
- Arbitrators order the parties to hire statistical experts to:
  - Design sampling plan
  - Determine sampling methodology, confidence level, margin of error, and sample size
  - Select claims for Arbitrator review
  - Extrapolate findings from audit based on Arbitrator's rulings per claim

## Medicare Program Integrity Manual (PIM)

- Issued by Centers for Medicare & Medicaid Services ("CMS")
  - *Chapter 8: Administrative Actions and Statistical Sampling for Overpayment Estimates", available at http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/pim83c08.pdf*
- The PIM is frequently cited as reference in questions of statistical sampling by CMS, HHS OIG and DOJ
- PIM is not a book about statistical sampling written by or compiled by experts in statistics
- PIM is a general set of guidelines that, in theory, should enable auditors to conduct valid statistical sampling audits



## CMS Updates to PIM: Guidance on Use of Statistical Sampling

- Issued 9/28/18, effective 1/02/19 (Transmittal 828)
- Guidance applies to Medicare contractors:
  - Medicare Administrative Contractors (MACs)
  - Recovery Audit Contractors (RACs)
  - Supplemental Medical Review Contractors (SMRC)
  - Unified Program Integrity Contractors (UPICs)
- Lack of prior specific guidance led to:
  - Successful challenges to sampling and extrapolations during administrative appeals
  - Frustration by providers about inconsistencies in use of sampling and extrapolation by Medicare contractors

## Sec. 8.4.1: Determining When a Statistical Sampling May Be Used

- When contractor determines a *sustained or high level of payment error* exists
- After documented educational intervention has failed to correct the payment error
- What is a "sustained or high level of payment error"? Includes:
  - >/= 50% error rate in contractor or other medical reviews
  - Prior history of non-compliance for same or similar billing issues, or historical pattern of non-compliant billing practices
  - CMS approval provided in connection with a payment suspension
  - Information from law enforcement investigations
  - Audits or evaluations conducted by HHS OIG



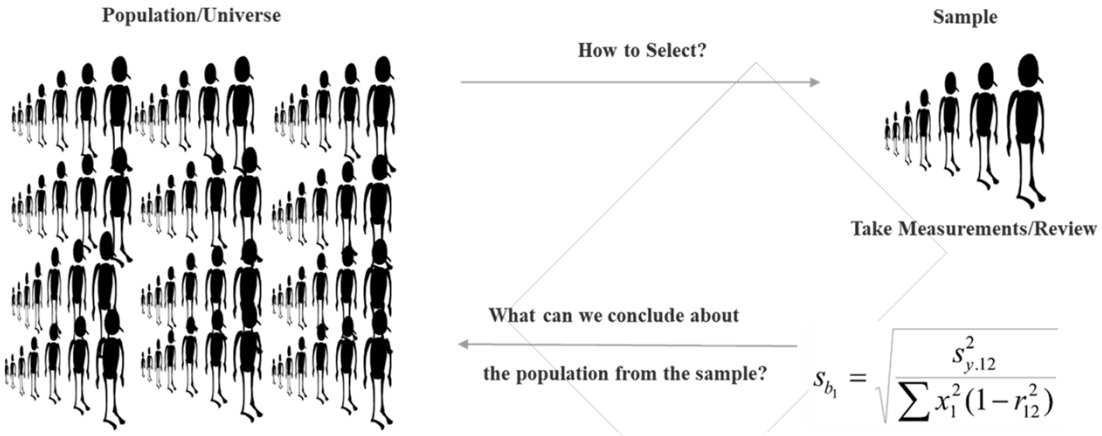
## Other Updates to PIM Chapter 8

- 8.4.1.1 – General Purpose
  - Failure by contractor to follow requirements does not necessarily affect validity of statistical sampling performed or projected overpayment
  - Appeals challenging validity of sampling methodology must demonstrate actual error in methodology affecting overpayment amount
- 8.4.1.5 – Consultation with a Statistical Expert
  - Requirements for contractors to consult with a Statistical Expert prior to releasing findings or overpayment demand letters
  - Minimum educational requirements for Statistical Expert defined
- 8.4.3.2 – Defining the Universe, Sampling Unit and Sampling Frame
  - Additional requirements for contractors to document and retain key information necessary to recreate the sample
- 8.4.7.1 – Recovery from Provider or Supplier
  - Requires contractor to obtain approval from CMS prior to issuing findings letter to provider for overpayments >$500,000 or 25% of provider's Medicare revenue

Trinity Health

BRG Berkeley Research Group

## Fundamentals of Statistical Sampling






Trinity Health

BRG Berkeley Research Group

## Non-Probability Sampling Techniques

- The core characteristic of non-probability sampling techniques is that samples are selected based on the subjective judgment of the researcher, rather than random selection
- Methods: Quota Sampling, Convenience Sampling, Self Selection Sampling
- Results from non-probability samples cannot be extrapolated to the underlying universe with a confidence level and margin of error

***Therefore, the quality and reliability of results from non-probability samples cannot be evaluated***

Trinity Health

BRG Berkeley Research Group

## Probability Sampling Techniques

- The core characteristic of probability sampling techniques is that samples are selected based on random selection utilizing selection probabilities for each element in the universe
- Methods: Simple Random Sampling, Stratified Random Sampling, Cluster Random Sampling, Probability Proportional to Size Sampling, etc.
- Allow us to draw valid conclusions about characteristics in the population and quantify confidence level and margin of error which measure the reliability of the sample results

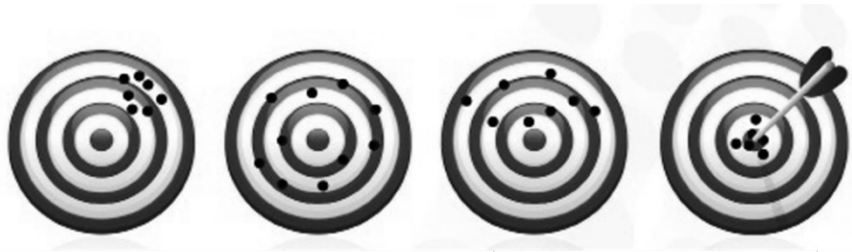


Trinity Health

BRG Berkeley Research Group

## Fundamentals of Proper Statistical Sampling – Key Terminology

- **Validity** – a statistical sample is valid when the mechanics of statistical sampling were applied properly
  - In that sense, a sample size of 2 to estimate the average weight in a population can be statistically valid
  - Obviously, a sample that small is neither accurate nor precise and thus unreliable
- **Precision** – a statistical sample is precise when the point estimate has a small margin of error
  - A scale that works properly and measures the weight within one ounce but is 20 pounds off is very precise, but obviously neither accurate nor reliable
- **Accuracy** - a statistical sample is accurate when it measures what it is supposed to measure
  - A nurse wrote down the weight of patients but the doctor had asked to see the patients' blood pressure
  - The measurements of the weight may be very precise but they are inaccurate to draw conclusions about the blood pressure, and therefore, they are unreliable
- **Reliability** – ***a statistical sample has to be valid, accurate, and precise to be reliable.***



| Inaccurate | Somewhat Accurate | Inaccurate | Accurate |
|---|---|---|---|
| Precise | Imprecise | Somewhat Precise | Precise |
| Unreliable | Unreliable | Unreliable | Reliable |



## Random Vs. Representative

- Randomness and representativeness are two different aspects of statistical random sampling
- Randomness in sampling refers to the fact that each element in the underlying universe must have a known positive selection probability
  - In its simplest version, in the so-called Simple Random Sample approach, each element in the universe has the same selection probability
- Representativeness refers to the property of the sample to represent the underlying universe well enough so that reliable inference can be drawn
  - Unfortunately, there is no precise statistical definition for “representativeness” of a sample.
  - However, the most frequently mentioned attributes of a representative sample are:
    - Adequate representation of the universe with respect to relevant characteristics of the universe elements (e.g., healthcare claims); and
    - Representative with respect to what is being estimated (e.g., alleged overpayments)

Trinity Health

BRG Berkeley Research Group

## Random Vs. Representative

- It is a false conception that a simple random sample is necessarily representative, as easily seen by an example of lottery numbers:
  - A random draw could yield the numbers 1, 2, 3, 4, 5, and 6
  - This particular random draw has the exact same probability of occurring as any other combination of six numbers out of a total pool of 49 numbers (e.g., 9, 16, 21, 32, 39, and 44)
  - However, the first six numbers in sequence are obviously not representative for the numbers from 1 to 49
  - Similarly, the random draw of even numbers 4, 12, 18, 22, 36, 44 has the exact same probability of occurring as any other six number combination
  - However, this sample – even though drawn randomly – would lead to the wrong conclusion that there are only even numbers in the universe because it is not representative of the odd and even numbers from 1 to 49
- Conversely, a sample that was not selected randomly but is representative of the population, cannot be used to extrapolate with a confidence level and a margin of error



Trinity Health

BRG Berkeley Research Group

## Random vs Representative

| | Representative | Non-Representative |
|---|---|---|
| Random | • Valid<br>• Accurate<br>• Precise<br>• Reliability<br>• No systematic bias<br>• Statistical significance<br>• Predictions about population are possible | • Valid<br>• Measured accuracy can be misleading<br>• Measured precision can be misleading<br>• Unreliable<br>• Can lead to systematic bias<br>• Statistical significance can be calculated but may be misleading<br>• Predictions about population can be misleading |
| Non-Random | • Invalid<br>• Impossible to assess accuracy<br>• Impossible to quantify precision<br>• Impossible to assess reliability<br>• Impossible to assess if systematic bias exists<br>• No statistical significance<br>• Impossible to make predictions about population | • Invalid<br>• Inaccurate<br>• Imprecise<br>• Unreliable<br>• Systematic bias<br>• No statistical significance<br>• Impossible to make predictions about population |

## Fundamental Steps of Proper Statistical Sampling

- Proper Universe Definition
- Proper Selection of Sampling Unit
- Construct Sampling frame
- Calculate Sample Size
- Determine Sampling Methodology
- Apply Random Selection Mechanism
- Proper Review Procedures/Objective Measurements
- Proper Extrapolation to the Population
- Address Issues of Non-Sampling Error

***Sampling demands attention to all phases of the process: poor work in one phase may ruin a sample in which everything else is done well***

*William Cochran*

*(Author of the book "Sampling Techniques" – an authoritative treatise that statisticians and most Government agencies cite)*



## Proper Universe Definition

- Every statistical analysis must start with:
  - The precise definition of the underlying population
  - The parameter of interest in that population
  - The hypotheses to be tested
- This step is critical because:
  - It defines the pool from which the sample items will be selected
  - It also defines the pool to which the sample results can be extrapolated

## Proper Universe Definition – Common Errors and Successful Appeal Challenges

Example 1:

- The universe definition included only on a subset of claims recognized as “high risk”
- Other claims with the same codes in the same years were ignored

Example 2:

- Universe used for review did not match the target universe of review
- A number of claims had been part of previous audits

Example 3:

- In a review spanning several codes in several years across multiple facilities, a number of claims appeared multiple times



## Proper Selection of Sampling Unit

- The sampling unit is the unit of analysis that is randomly selected from the sampling frame representing the universe
- If a line item on a claim is the sampling unit, then the sampling frame has to be comprised of all line items in the universe; similarly if the sampling unit is a claim or a beneficiary
- Based on measures derived from the sampling unit in the sample, the number of sampling units in the sampling universe is used in the extrapolation process

## Improper Choice of Sampling Units – Common Errors and Successful Appeal Challenges

Example 1:

- Individual days were selected as a sampling unit
- Days for the same patients were highly correlated and could not be used as independent sampling units
- Extrapolation based on days was disallowed

Example 2:

- Patients were selected as a sampling unit
- Patients were clusters of a largely varying number of claims and the review had cherry picked the claims for the randomly selected patients

Example 3:

- So-called “penny-sampling” randomly selects a penny of an invoice and then attempts to evaluate the penny selected
- Method was rejected because a paid amount is comprised of statistically dependent sampling units of pennies, and therefore, the extrapolation based on errors found in pennies was incorrect



## Construct Sampling Frame

- The sampling frame is a physical list of all the sampling units in the population
- It is of utmost importance to ensure that the sampling frame utilized is an adequate representation of the population
- Errors of inclusion (items that don’t belong into the universe are on the sampling frame)
- Errors of exclusion (Items that belong in the universe are missing on the sampling frame)
- Combination of both
  - Result in invalid extrapolations

## Improper Sampling Frame – Common Errors and Successful Appeal Challenges

Example 1:

- Sampling frame was incomplete and focused on a subset of claims that was recognized as "high risk"
- Not every claim in the universe had a chance of being select, thus biasing the extrapolation upward by including only the high risk claims

Example 2:

- Sampling frame was over-inclusive and thus the average overpayment was multiplied by a number of claims that was too large

Example 3:

- The sampling frame included claims that had been part of a previous investigation/review and the extrapolation was disallowed



## Calculate Sample Size

- Sample size has to be determined such that the sample results can be reported with high confidence and low margin of errors
- Once the confidence and margin of error have been determined, the actual calculation of the sample size is based on an analysis of the underlying population data
- Use of probe or pilot sample:
  - Estimate variation in the universe
  - Calculate sample size necessary to achieve the desired confidence and precision

## Calculate Sample Size – Common Errors and Successful Appeal Challenges

Example 1:

- Sample size chosen was too small to justify the approximation of a normal curve
- Therefore, the confidence limit used as the overpayment determination was incorrect.

Example 2:

- Review used a “one-size fits all” sample size of 30 and the random selection of 30 claims did not even include claims for each of the codes and years that were the scope of the review

Example 3:

- While the PIM argues that a small sample size is not an argument against the validity of a sample, small samples are often neither accurate nor precise, and often lead to mis-representation of important aspects of the universe



## Determine Sampling Methodology

- The choice of the appropriate sampling methodology must be based on:
  - The intended use of the sample results
  - The accessibility of the universe
  - The availability of sampling units for testing; and
  - The required confidence level and margin of error to deem the sample results reliable

## Determine Sample Methodology – Common Errors and Successful Appeal Challenges

Example 1:

- Simple random sampling was applied and the small sample of 30 did not cover important aspects of the universe

Example 2:

- Cluster sampling was used but the approach did not take into account the different size of the clusters and the extrapolation was disallowed

Example 3:

- For convenience reasons, the review was based on a block sampling approach which is non-statistical in nature, and therefore, could not be used for extrapolation purposes



## Application of Proper Random Selection Mechanism

- Results from non-random samples cannot be extrapolated to the entire population with a confidence level and margin of error
- Therefore, the quality of non-random samples cannot be evaluated in a statistically valid or reliably quantifiable way

## Application of Proper Random Selection Mechanism – Common Errors and Successful Appeal Challenges

Example 1:

- Reviewer had made a mistake in the application of the random number generator which lead to a large degree of duplicative random numbers across strata

Example 2:

- The same starting seed was used in a multiple facility review leading to a systematic selection across different facilities, which invalidated the random nature of the sample

Example 3:

- For convenience reasons, the review was based on a block sampling approach and the blocks (months of claims) were selected without random numbers



## Summarize Results and Extrapolation to the Population

- The final step in statistical sampling is the summary of the sample results followed by the extrapolation to the entire population from which the sampling units were initially selected
- The extrapolation methodology is determined by the way the sample was selected

## Improper Extrapolation – Common Errors and Successful Appeal Challenges

Example 1:

- The sample was based on a stratified design but the results were extrapolated using a simple random sample design which put an inflated weight on the large claims in the sample which significantly inflated the extrapolations

Example 2:

- In several cases, errors in the calculations invalidated the lower confidence limits which were used to determine overpayments.

Example 3:

- Results from a non-random sample were treated like a random sample and erroneously extrapolated using formulas that are only valid for random samples



## Contact Information

**Michael Holper**
SVP Compliance and Audit Services
Trinity Health
Email: holperm@trinity-health.org
Office: 734.343.1638

**Stefan Boedeker**
*Managing Director*
Berkeley Research Group
Email: SBoedeker@thinkbrg.com
Office: 310.499.4924