

Sampling Methods

Introduction

Sampling is a process of selecting a representative part of a population in order to estimate the organization's performance, without collecting data for its entire population. Using a statistically valid sample, an organization can measure its performance in an effective and efficient manner. Sampling is a particularly useful technique for performance measures that require primary data collection from a source such as the medical record. Sampling should not be used unless the organization has a large number of cases in the measure set population because a fairly large number of sample cases is needed to achieve a representative sample of the population. For the purpose of sampling national quality measures, the terms "population", "sample", "effective sample", and "case" are defined as below:

- A "population" refers to all patients who share a common set of specified, administratively derived data elements. This may include ICD-9-CM diagnosis codes or other population characteristics such as age. For example, the population for the acute myocardial infarction (AMI) measure set includes all patients having a principal diagnosis of AMI from Appendix A, Table 1.1
- The "sample" is the fraction of the population that is selected for further study.
- "Effective sample" refers to that part of the sample that makes it into the denominator of a measure. This is defined as the sample for a measure minus all the exclusions and contraindications for that measure in that sample.
- A "case" refers to a single record (or an episode of care) within the population. For example, a health care organization may have 100 patients who experienced an AMI as the principal diagnosis in the first quarter. The organization's population would include 100 cases or 100 patient records for the AMI measure set in January.

To obtain statistically valid sample data, the sample size should be carefully determined and the sample cases should be randomly selected in such a way that the individual cases in the population have an equal chance of being selected. Only when the sample data truly represent the whole population can the sample-based performance measure data be meaningful and useful. The measurement system is responsible for ensuring that their sampling techniques are applied consistently.

Organizations collecting a sample of cases must meet the following sampling requirements.

1. Sampling Availability

Sampling is done by national quality measure set, except for Pregnancy. For measures that require medical record abstraction, sampling should be done using available databases that contain discharges by quarter, ICD-9-CM diagnosis codes, and other necessary administrative data. After determining sample cases for each measure set, the EOC level data elements are collected from medical records. The specific measure set sampling populations are defined below:

- Acute myocardial infarction (AMI) measure set: Patients with *ICD-9-CM Principal Diagnosis Codes* for AMI as defined in Appendix A, Table 1.1;
- Heart failure (HF) measure set: Patients with *ICD-9-CM Principal Diagnosis Codes* for HF as defined in Appendix A, Table 2.1;
- Pneumonia (PN) measure set: Patients with *ICD-9-CM Principal Diagnosis Codes* for pneumonia (PN) as defined in Appendix A, Table 3.1;
 - OR Patients with *ICD-9-CM Principal Diagnosis Codes* of Septicemia or Respiratory Failure as defined in Appendix A, Table 3.2 and Table 3.3 accompanied by an *ICD-9-CM Other Diagnosis Code* of PN as defined in Appendix A, Table 3.1;
- Pregnancy and Related Conditions (PR) measure set: The PR core measure set is unique in that samples are drawn from two distinct groups, mothers and babies.
 - Measures PR-1 and PR-3, include patients with a *ICD-9-CM Principal or Other Diagnosis Code* on one of the following Appendix A, Tables – 4.01, 4.02, 4.03, or 4.04.
 - Measure PR-2 includes all live born neonates that are less than 28 days old (*Discharge Date – Birth Date < 28 days*).
 - In order to take advantage of data collection efficiencies, one available sampling option for the PR core measure set is to collect data on the entire population for PR-1 and PR-3 (mother records), and select a sample for PR-2 (baby records). This option will allow organizations to reduce their data collection burden for PR-2, which may require some medical record abstraction, while encouraging 100% data collection for measures that may be derived entirely from administrative data sources (PR-1 and PR-3);
- Surgery Infection Prevention (SIP) measure set: Patients with *ICD-9-CM Principal and/or Other Procedure Codes* for SIP as defined in Appendix A, Table 5.01 to 5.08

Note: Health care organizations are NOT required to sample their data. For measure sets that can be derived entirely from administrative data (such as the PR set), it may be simpler to submit all cases. Similarly, if sampling offers minimal benefit (i.e., an organization has 80 cases for the quarter and must select a sample of 76 cases) the organization may choose to use all cases.

2. Sample Size Requirements

Organizations selecting sample cases for AMI, HF, PN and/or PR measure sets should ensure that its measure population(s) and effective sample size(s) meet the following conditions:

- *The effective sample size for a measure set is at least 35 cases per quarter; and*
- *The required sample size is at least 20% of the measure set population for the quarter (see Tables 1 through 4 for measure set specific sample size requirements).*

For organizations selecting sample cases for the Surgery Infection Prevention (SIP) measure set, an alternative sampling procedure is required. Organizations selecting sample cases for this set should ensure that each individual strata population and effective sample size meets the following conditions:

- *Case selection occurs within each of the seven individual strata (e.g. CABG, cardiac surgery, hip arthroplasty, etc.). The effective sample size within a strata is at least 10 cases per quarter; and*
- *The required sample size is at least 10% of the strata population for the quarter (see Table 5 for specific sample size requirements).*

An organization or measurement system may choose to use a larger sample size than is required. Organizations whose measure set population size is less than 60-78 cases per quarter, depending on the measure set, cannot sample for that measure set.

Note: Measurement systems should monitor health care organization samples to ensure that sampling procedures consistently produce statistically valid and useful data. Because the sample for a measurement set will rarely be equal to the effective sample due to exclusions and contraindications, organizations selecting sample cases should over-sample their population to obtain an adequate effective sample size. The over-sample rate will differ for the different measurement sets because the rate of exclusions and contraindications vary between sets. The following sample size tables for each of the measurement sets automatically build in this over-sample rate to obtain the required sample sizes. These over-sample rates are based on a national data base reflecting experience collected over an eighteen month reporting period. The number of cases sampled will be routinely monitored to determine whether it is sufficient.

Table 1: Sample Size Based on Population Size for the Acute Myocardial Infarction (AMI) Measure Set

<i>Health Care Organization's Measure</i>	
Average Quarterly Population Size "N"	Minimum Required Sample Size "n"
≥ 1556	311
387 – 1555	20% of population size
78 – 386	78
< 78	No sampling; 100% population required

Table 2: Sample Size Based on Population Size for the Heart Failure (HF) Measure Set

<i>Health Care Organization's Measure</i>	
Average Quarterly Population Size "N"	Minimum Required Sample Size "n"
≥ 1522	304
379 – 1521	20% of population size
76 – 378	76
< 76	No sampling; 100% population required

Table 3: Sample Size Based on Population Size for the Pneumonia (PN) Measure Set

<i>Health Care Organization's Measure</i>	
Average Quarterly Population Size "N"	Minimum Required Sample Size "n"
≥ 1207	241
300 – 1206	20% of population size
60 – 299	60
< 60	No sampling; 100% population required

Table 4: Sample Size Based on Population Size for the Pregnancy and Related Conditions (PR) Measure Set

<i>Health Care Organization's Measure</i>	
Average Quarterly Population Size “N”	Minimum Required Sample Size “n”
≥ 1250	250
311 – 1249	20% of population size
62 – 310	62
< 62	No sampling; 100% population required

Table 5: Sample Size Based on Population Size for the Surgery Infection Prevention (SIP) Measure Set

<i>Health Care Organization's Measure</i>	
Average Quarterly Strata Population Size “N”	Minimum Required Strata Sample Size “n”
≥ 350	35
121 – 349	10% of population size
12-120	12
< 12	No sampling; 100% population required

Sample Size Examples

- An organization using AMI measures has 100 AMI discharges during the fourth quarter. Using Table 1, the sample size is seen to be a minimum of 78 AMI patients for this quarter.
- An organization’s population size for the PN measure set is 2,400 PN discharges during the first quarter. Twenty percent of 2400 equals 480 pneumonia patients -- which exceeds the maximum sample size condition given in Table 3 (i.e., 241); thus, the required sample size would be at least 241 pneumonia patients for that quarter.
- The HF measure set population size for an organization has been 500 patients per quarter during the past year. The required sample size using Table 2 would be 100 (20% of 500) heart failure patients per quarter -- as this number is smaller than the maximum condition (i.e., 304 cases) and larger than the minimum condition (i.e., 76 cases).
- The SIP measure set population sizes for an organization is 5, 50, 15, 140, 35, 60, and 120 respectively patients per quarter respectively for the seven strata. The required sample sizes using Table 5 would be 5, 12, 12, 14, 12, 12, and 12 respectively for the seven strata.

3. Sampling Approaches

Simple random sampling or systematic random sampling must be used.

- Simple random sampling - selecting a sample size (n) from a population of size (N) in such a way that every possible sample of size n has the same chance of being selected.
- Systematic random sampling - selecting every k^{th} record from a population of size N in such a way that a sample size of n is obtained, where $k \leq N/n$. The first sample record (i.e., the starting point) must be randomly selected before taking every k^{th} record. This is a two-step process:
 - a) Randomly select the starting point by choosing a number between one and k using a table of random numbers or a computer-generated random number; and
 - b) Then select every k^{th} record thereafter until the selection of the sample size is completed.

To ensure that the sampled data represent the health care organization's measure set population, sampling techniques are determined by the measurement system using simple or systematic random sampling methods. The measurement system is responsible for ensuring that the chosen sampling techniques are applied consistently across participating health care organizations.

Sampling Approach Examples

For an organization with a measure set population size of 350 heart failure (HF) discharges per quarter, the sample size would be 78. To select a random sample of 78 HF patients:

- Simple random sampling:
 1. Generate random numbers for individual HF patient records from a random number function using a statistical software package or computer programming language.
 2. Sort data by the random numbers either in an increasing or decreasing order.
 3. Select the first 78 HF patient records as the random sample.
- Systematic random sampling:

1. In this example, the organization's measure set population size= 350 and the sample size = 78. Divide the population size by the sample size and take the quotient (i.e., the integer portion) as the sampling interval k . The sampling interval $k = 350/78 = 4.5$. Thus, every 4th HF patient record will be selected from the measure population until 78 cases are selected.
2. To ensure that each HF patient has an equal chance of being selected, the "starting point" must be randomly determined before selecting every 4th HF patient record. This can be done using a computer random number generator or a random number table to randomly choose a number between 1 and 4 as the starting point.

Note: Measurement systems must transmit two data elements, *Sample* and *ICD Population Size* for measure data transmission. The *Sample* data element indicates if the data being transmitted for a health care organization have been sampled, or whether the entire population was used for the specified time period. The *ICD Population Size* data element indicates the measure population size, prior to the application of data integrity filters, measure exclusions, or the application of the measure set common logic and/or sampling. Please refer to the *ORYX[®] Technical Implementation Guide* for more information about transmission data elements.

Example:

A hospital uses AMI ICD-9-CM Principal Diagnosis Codes (as listed in Appendix A, Table 1.1) to identify 120 AMI cases during the second quarter. From these 120 cases, the hospital randomly selects a sample of 78 cases. Data for these 78 cases are collected and sent to the measurement system and are then used to calculate the hospital's rate for each AMI measure. During data transmission to the Joint Commission, the measurement system would transmit "Y" in the *Sample* field since the data being transmitted represent a sample of the entire population and "120" in the *ICD Population Size* field since this represents the total number of cases that were eligible to be included in the measure population.