An Explanation of the “8 and 30” File Sampling Procedure Used by NCQA During Accreditation Survey Visits

MAY 1, 2001

Prepared by Ernest Valente, Ph.D.

Overview

The National Committee for Quality Assurance (NCQA) conducts on-site surveys as part of the health care organization[1] (HCO) accreditation process. Survey teams assess the performance of health care organizations on a variety of standards during these visits, in order to review the quality and efficiency of fundamental management processes. Performance on a number of accreditation standards is assessed by examining record files that contain information about how HCOs handle certain activities. For example, NCQA’s utilization management standards require that appropriate professionals be involved in utilization management reviews. During accreditation surveys, NCQA surveyors examine utilization review records for evidence that appropriate professionals were indeed involved in the utilization review process. HCOs are assigned scores based on the proportion of files that show evidence of compliance with the reviewed standard, and plans that meet this standard in a high proportion of the reviewed files are assigned a higher score on this element than plans with a lower proportion of compliant records. Table 1 shows the NCQA accreditation standards that include record file reviews.

This document explains the record file review procedure used by NCQA during the HCO accreditation survey process. The procedure is commonly referred to as the “8 and 30” file sampling procedure, because the procedure involves reviewing an initial sample of eight files, then reviewing an additional sample of 22 files if any of the original eight fails the review (a total of 30 records). The principles in this document apply to the universe of health care organizations that the NCQA accredits, including Managed Care Organizations (MCO), Managed Behavioral Health Care Organizations (MBHO), Preferred Provider Organizations (PPO), and POC and Organization Certification. This document includes four sections: a brief overview of the purpose of the “8 and 30” file review procedure; a review of the appropriateness of the current procedure in a changing managed care environment; an assessment of the statistical appropriateness of the current procedure; and a description of improvements to the procedure. sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinction.

Section 1. Purpose of the “8 and 30 File Sampling” Procedure

Part of the NCQA accreditation process involves the examination of a sample of record files for verification that the management process documented in the record complies with NCQA standards. For example, records of utilization management denials are examined to determine whether UM decisions are being made in an appropriate timeframe and by qualified professionals.

NCQA uses a quality-control sampling procedure to inform the record review process and to classify HCOs into one of three performance categories: HIGH (90% - 100% of the records are in compliance with the review criteria); MEDIUM (60% - 89% in compliance) or LOW (less than 60% in compliance). Thirty files are obtained for review from each HCO, along with 10 alternate records to replace records that, when reviewed more closely, do not meet criteria for inclusion in the sample (a sample of 30+10). Eight of the 30+10 files are reviewed for compliance with NCQA criteria. If all eight are determined to be in compliance, then the HCO is determined to be in HIGH compliance. If one or more of the initial eight are out of compliance, then 22 additional records are examined, so that 30 total records are reviewed to assess the compliance status of the HCO. NCQA refers to this sampling procedure as the “8 and 30 file sampling” procedure.

Delegates. HCO services are often provided by “delegate” organizations instead of the HCOs themselves. Delegate organizations are engaged by HCOs to provide defined services, such as emergency care authorizations or physician credentialing services, to some or all of the HCO’s members. NCQA includes records from delegate organizations in the file review procedure. It is important to note that the reviewed services are sometimes part of larger service processes (e.g., emergency room denials are reviewed, but they are only a subset of the larger service of emergency room authorizations).

[1] “Health care organizations” refers to the universe of organizations that the NCQA accredits, such as Managed Care Organizations (MCO), Managed Behavioral Health Care Organizations (MBHO), Preferred Provider Organizations (PPO), and POC and Organization Certification.

The proportion of delegate records in the sample is based on the amount of the reviewed service provided by delegates (e.g., ER denials), not the amount of the larger service (ER authorization). This means that the representation of delegate organizations in the sample may not be proportionate to their provision of the larger service. For example, it is possible that a delegate organization performs emergency room authorizations for 20% of the membership of an MCO, but processes 70% of ER denials for that MCO (the MCO itself handles the other 30% of ER denials based on the other 80% of the membership). In this case the proportion of emergency room denials processed by the delegate organization (70%) is greater than the proportion of ER authorizations processed by that delegate overall (20%). The sample proportions are based on ER denials (not ER authorizations), so 70% of the ER denial records in the sample will come from the delegate (not 20%), and 30% of the records will come from the MCO.

Section 2. Is the "8 and 30" File Sampling Procedure Still Appropriate in the Current HCO Environment?

The 8 and 30 file sampling procedure was developed in 1996. NCQA staff, in consultation with faculty of the Department of Biostatistics at Johns Hopkins University, recently reviewed the procedure to determine whether modifications are necessary because of changes in HCO organization and management since 1996. The three major changes in the HCO environment that might affect the "8 and 30" procedure are:

- Member population size differences are becoming greater across HCOs because of consolidation through mergers and acquisitions. Is the 8 and 30 file sampling procedure appropriate for such a wide range of population sizes?
- Carve-outs and subcontracting are increasing the number of distinct delegate organizations that may be performing the same administrative functions on behalf of the same HCO. Is the 8 and 30 procedure accurate when a large number of delegate organizations must be included in the sample?
- The rise of specialty contracting organizations (such as managed behavioral health organizations [MBHOs]) is increasing the demand to compare performance on subclasses of management activity (e.g., MBHOs would like to compare their performance on behavioral health records to subsets of behavioral health records processed by MCOs). Can the 8 and 30 procedure be adapted to answer these sorts of questions?

These circumstances in the HCO environment are causing several concerns about the implementation of the 8 and 30 file sampling procedure. As a result, several questions about the statistical appropriateness of the 8 and 30 file sampling procedure have been raised:

A. **How can NCQA make a scoring decision based on such a small initial sample?** The initial sample of eight records seems too small to yield statistically meaningful results, and seems unfairly biased when records from delegate organizations are included.

B. **Why is the sample size the same for HCOs of all sizes?** The same number of records is sampled for health plans with very different numbers of eligible records in the population (e.g., all ER denials). Thirty records are sampled regardless of whether the HCO has 100 eligible records or 10,000 eligible records. Those plans with relatively small populations of eligible records feel that they are being subjected to increased scrutiny since a much higher proportion of their population is being sampled.

C. **Isn't it unfair that all records are sampled in small HCOs?** HCOs with a small population of eligible records (50 or fewer) feel that it is unfair that most or all of their records are sampled (when only a small proportion of the total population is selected from larger plans).

D. **What happens when there are a small number records in the total population**? NCQA sometimes samples additional records of a different type when there are fewer than 30 records of the desired type. For example, in some cases when there are fewer than 30 medical necessity denial records eligible for review, benefit denial records are sampled to make up the difference. Is this appropriate?

There are also questions about the appropriateness of the procedure given the increased prevalence of delegate organizations:

E. **Is the “8 and 30” procedure appropriate when the sample includes too many “small” delegates that do not represent the performance of the HCO overall?** The 30 record sample is composed of records from the HCO being assessed and records from any delegate organizations that perform reviewed services (e.g., emergency room [ER] denials). Records handled by delegate organizations are included in the sample in roughly the proportion that the organizations perform the reviewed service (although there is some deviation from proportionate representation because delegates are selected using a random process). If one of the eight preliminary records fails review, and that failed record comes from a delegate that serves a small proportion of members in the larger service, HCOs sometimes feel that the 8/30 process is unfair. The perception of unfairness comes from a feeling that “small-proportion” delegate records should not be 4 included in the first review of eight. (“If you are going to assess using such a small sample how can you include records from a delegate that serves so few of our members?”).

F. **Should the sampling methodology yield a proportionate sample when delegates also perform the activity under review?** Records from delegate organizations are randomly selected from a proportionate sampling frame, but this procedure does not necessarily yield a proportionate sample, especially when there are large numbers of delegates that process relatively small numbers of cases. Is the procedure where the sample is proportionately composed of delegated records still appropriate given the increased prevalence of delegation, or should the procedure be modified to fully sample the records of some delegates?

The 8 and 30 file sample sometimes includes subclasses of records, typically defined either by the entity that processed the record (such as a specialty contracting organization) or by the type of activity captured in the record (e.g., behavioral health cases). Currently, neither entity nor activity subclasses are analyzed.

G. **Can the “8 and 30” procedure be adapted to inform decisions about subclass comparisons?** Is it appropriate to “force” subclass files into the sample, so that there are a sufficient number of subclass records to make it possible to compare the performance of plans on particular subclasses of activities (such as behavioral health records processing)?

Section 3. Statistical Appropriateness of the “8 and 30” File Sampling Procedure

The 8 and 30 file sampling procedure maximizes the information about reviewed services derived from a small amount of data. The procedure allows HCOs to be classified into three performance categories (LOW/MEDIUM/HIGH) with a minimal amount of file review work from the perspective of both the HCO and NCQA. This efficiency is dependent on an extremely sensitive combination of sampling formality and strict decision rules:

I. The procedure is designed to provide information only about the performance of a single HCO, in the aggregate, on a conceptually uniform class or subclass of activities (e.g., denial notices or ER denials). It is not appropriate to draw inferences about multiple activities or subclasses from the same sample of 30 records. A separate sample is required for each distinct subclass in order to draw valid inferences about subclasses. For example, ER denials and medical necessity denials are subclasses of denials that NCQA reviews with separate samples of 30 files each, because of their importance to consumers.

II. Population inference based on the 8 and 30 sample is accurate if the sample is fully representative of the population of records in the HCO for the service being reviewed (the sample must be a microcosm of the population of all HCO records). Classification decisions based on the sample would be the same if based on the entire population, so the procedure is fair when applied to all of the records in a small population. The 8 and 30 sample should have these characteristics:

a) The sample should be selected through a random process.

b) The sample should be fully representative of the population of reviewed services (ER denials), not larger services (ER authorizations). For example, NCQA assesses quality performance during the

processing of ER denials (not the number of ER authorizations that result in denials), so the sample should look like all ER denials in the HCO (not ER authorizations). This sampling strategy is appropriate because NCQA wishes to answer the question “Was the typical ER denial processed appropriately?” not “What was the likelihood that a typical member experienced an appropriately handled ER denial?” See Section 1, *Delegates* for a discussion of this issue.
c) The proportion of delegated records in the sample should be equal to the proportion of all cases delegated by the HCO, and the proportion of sampled records that are handled by any one delegate should be equal to the overall proportion of HCO records handled by that delegate.

III. The statistical test underlying the classification decision is based on the binomial distribution, the characteristics of which are very well known. This fact allows the classification decision to be based on a very small sample of 30 files, regardless of the size of the population of eligible records. The use of the binomial distribution is possible because the decision based on the 8 and 30 file sampling methodology is BINARY. That is, the decision based on the file review falls into one of two possible categories ("in compliance"/"out of compliance").

IV. The decision to categorize an HCO as "HIGH performing" in the initial file review of eight records is statistically reliable if based on a “LOW performing” percentage of 60%.[2] Currently, HCOs that are found to have fewer than 60% of reviewed records to be "in compliance" are classified as "LOW performing" organizationally. The statistical soundness of the decision to classify an HCO as “HIGH performing” if there is a perfect score in the initial sample of eight records hinges on this 60% criterion. The estimate based on eight initial records is not very precise, but it is just precise enough to determine that an HCO is performing better than “LOW compliance” when all eight initial records are in compliance.[3]

When properly conducted, the 8 and 30 file sampling procedure is statistically appropriate and efficient. Our review of the statistical appropriateness of the procedure addressed some of the issues described in Section 2. The random selection of files and the reliance on the binomial distribution makes the initial review of eight files valid (see Section 2, item A). In addition, reviewing the same number of files is valid regardless of the size of the HCO being surveyed (see Section 2, item B). Reviewing all of the eligible records in a HCO with a small number of eligible records is not unfair, since inferences based on a random sample lead to the same conclusions as inference based on the entire population (see Section 2, item C). It is likely that any modifications of the above conditions (e.g., so that plans performance on behavioral health records can be 2 This is based on the exact confidence intervals for the binomial distribution where N=8. The additional precision necessary to distinguish between the “HIGH” and “MEDIUM” levels of compliance is achieved when N=30. 3 HCOs that would be classified as “MEDIUM compliance” if 30 records were reviewed could possibly be categorized as “HIGH compliance” on the basis of the initial review of eight records. 6 compared) would result in a more demanding and costly procedure for NCQA and the HCOs involved in the accreditation process.[4]

Section 4. Improvements to the “8 and 30” File Sampling Procedure

While NCQA has concluded that the 8 and 30 sampling procedure is fundamentally sound, we have implemented several improvements of the current procedure. The improvements are designed to preserve the integrity of the procedure in a

[2] This is based on the exact confidence intervals for the binomial distribution where N=8. The additional precision necessary to distinguish between the “HIGH” and “MEDIUM” levels of compliance is achieved when N=30.
[3] HCOs that would be classified as “MEDIUM compliance” if 30 records were reviewed could possibly be categorized as “HIGH compliance” on the basis of the initial review of eight records.
[4] It should be noted that the same record file sometimes informs several standards, so performance on each separate standard may not be independent of performance on other standards informed by the same record. For example, if denial records that aren’t processed by appropriate professionals were less likely to be processed quickly, then a sampled record that fails UM 3 would also be more likely to fail UM 4. There is currently no empirical information about the independence of separate standards elements that are sampled from the same file. However, this does not pose a problem for statistical inference even if the separate standards are not independent. Performance on different standards is not statistically compared, so clustering is not an issue. Plans are not strongly advantaged or disadvantaged by non-independence, because the picture of overall performance will be balanced, yielding an accurate estimate of plan performance (e.g., both good and poor performance would be consistent across standards).

cost-effective manner. The six changes listed below are being made for the following accreditation and certification programs, effective July 1 ,2001: MCO, MBHO, PPO, and POC and Organization Certification.

1. Currently, individual files are chosen for the sample by first randomly selecting a date, then sampling the first alphabetical file on that date. If no eligible files appear on that date, a file is chosen from the next day. Different sets of 40 dates are used for separate areas of activity (e.g., one set of 40 random dates is chosen for medical necessity denials, one set for ER denials, and one set for appeals). Dates are eligible to be chosen only once for each activity (*sampling without replacement*). This file sampling procedure will be changed to so that it is more randomized.

   NCQA will continue to select different sets of random dates for each activity, but the same date will be eligible for selection more than once within an activity (*sampling with replacement*). NCQA will provide randomly selected primary dates and several alternate dates for each activity area. The alternate dates will be used if no eligible files appear on the primary date (instead of using the next day). A random process will be used to sample the individual records selected for review within dates (rather than selecting the first file in alphabetical order). These improvements will increase the representativeness of the 8 and 30 sample (see Section 3, item II-a).

2. Currently, records are randomly selected from delegates with probability proportional to their size. For example, if 25% of an MCO’s ER denials can be attributed to delegate A, 50% to delegate B, and 25% to the MCO itself, then there is a 25% chance that a record from delegate A is selected for the sample, a 50% chance that a record from delegate B is selected and a 25% chance that a record from the MCO is selected. This is likely to yield a sample that is composed of 10 (25%) records from delegate A, 20 (50%) records from delegate B, and 10 (25%) records from the MCO. However, the actual numbers may differ because of the variability inherent in chance processes. To ensure that the file review process adequately captures information about delegate organizations, the sample of 40 records will be based on the actual proportion of reviewed services performed by each delegate organization. In the example above, we would randomly select 10 files from delegate A, 20 files from delegate B, and 10 files from the MCO instead of leaving these numbers to chance. Each record in the sample will represent 2.5% of reviewed service (1/40 = .025). Due to rounding, it may be necessary to add a record or two to fully represent the delegate activity, so the final sample may have more than 40 records. This change will insure that delegates are accurately represented in the 8 and 30 sample (see Section 2, item F and Section 3, item II-c).

3. Delegate organizations that handle less than 2.5% of the reviewed records for an HCO will be pooled and a subset will be randomly selected for the sample, so that the group is represented in the sample in its total proportion. For example, if 15 delegate organizations provide five percent of a reviewed service when their percentages are added together, then two slots in the sample of 40 will be filled by two delegate organizations selected at random from the group of 15 (*with replacement*). See Section 2, item F.

4. Files will be randomly selected for inclusion in the initial sample of eight, with a selection probability proportionate to their size. In this way, the population of all reviewed files will be better represented in the initial review of eight records, as records from larger delegates are more likely to be selected. See Section 2, item E.

5. NCQA will continue to sample all eligible records in HCOs that have small populations of eligible records. See Section 2, item C and Section 3, item II.

6. NCQA will not supplement a small number of files with files of another type (e.g., benefit denial records for medical necessity denial records). See Section 3, item I.

7. NCQA will not "cap" or restrict any one type of file or activities in the sample of 30 records (e.g., the number of pharmacy denials will not be limited in the sample when evaluating medical necessity denials). See Section 3, item II.

8. NCQA will not adapt the "8 and 30" file review for activities where we wish to draw separate inferences about performance across plans from the same sample of 30 records. That is, we will not “force” the sample to include subgroups of records to address questions such as "Does MBHO A do a better job with behavioral health records than MCO B?" See Section 3, item I.

Conclusion

Our review of the “8 and 30” file sampling procedure confirmed that it is a cost effective and statistically appropriate method of gathering data about the overall performance of an HCO. As a result of this review and changes in the HCO environment, especially the 8 delegation of defined activities to separate organizations, NCQA has made changes to strengthen the fundamental procedure. The changes explained in this document are meant to improve the implementation of the 8 and 30 file review procedure, and to introduce new features that allow the application of the procedure to circumstances that did not exist at the time that the procedure was established.

Table 1. Summary of Accreditation Standards Reviews that Include Reviews of Sampled Records Files for 2001 Standards

2001 MCO Standards Category	Accreditation Standard	Files Reviewed 8/30?	Beh. Health?
Utilization Management	UM3. Appropriate Professionals	Denials[1]	YES
	UM4. Timeliness of UM Decisions	Denials	YES
	UM5. Clinical Information	Denials	YES
	UM. 6 Denial Notices	Denials	YES
	UM 8. Appropriate Handling of Appeals	Appeals	YES
	UM 11. Emergency Services	Emergency Denials	YES
Credentialing and Recredentialing	CR 3. Initial Primary Source Verification	Credentialing	YES
	CR 4. Application and Attestation	Credentialing	YES
	CR 5. Initial Sanction Information	Credentialing	YES
	CR 6. Initial Credentialing Site Visits	Credentialing	YES
	CR 7. Recredentialing Primary Source Verification	Recredentialing	YES
	CR 8. Recredentialing Sanction Information	Recredentialing	YES
	CR 9. Performance Monitoring	Recredentialing	YES

[1]Or other UM files if there are fewer than 30 denials available

NOTE: Cells that are the same color indicate that the same sample of files is used to assess compliance with multiple measures.