CHAPTER 2

Keeping the Health Care Sampling Gains Going

Catherine Sreckovich, Alan Peterson and B. Bo Martin

I. Introduction

This chapter explores how government regulators, managers in provider and payor organizations, consultants, and other individuals have advanced statistical sampling in the United States health care industry.[1] The major initial cause for this statistical sampling development was the government's need to reasonably oversee its payments for health care services.[2] These government payments were largely direct and indirect individual payments for a massive number of health care claims from health care providers and health plans. Since these government claims and later commercial payors' claims could not all be reviewed in detail, questions related to whether the payments themselves were reasonable, and whether the claims were fair, could be instead answered in part through statistical sampling. Moreover, changes in health care funding, utilization review policies, laws, court cases, government regulations and field practices led to changes in the characteristics of what is sampled and the methods of sampling.

For government regulators who use statistical sampling in health care, this chapter tells them 1) that their statistical sampling applications to date have been reasonable as well as helpful and successful, 2) that the constructive regulatory government style adopted to date has generally been sound, and 3) that both the statistical sampling applications and the constructive style should be retained and improved.

For individuals in provider, non-government payor and other health care entities, this chapter tells them that 1) the use of statistical sampling is an important component of testing compliance as well as quality; 2) those who are not presently using such statistical sampling should consider its use, and 3) anywhere statistical sampling skills are needed, these skills should be acquired or enhanced timely. It is noteworthy that many individua ls involved in health care compliance expect major increases in "whistleblower" activity beginning in 2007. This expectation arises from requirements of the Deficit Reduction Act of 2005, among other similar developments, to educate provider staff on whistle-blowing and on fraud and abuse in state Medicaid programs, as well as an increasing attention of the federal government on quality of care issues and major increases in state government compliance.[3]

Moreover, it is also noteworthy that a slogan of "transformational results" has more recently emerged in the federal government communication with various audiences—most particularly public audiences on health care compliance.[4] This newer slogan is put forth in many variations, all of which focus on major

behavior changes in health care compliance as well as on quality. Common references include assertions that testing (sampling) of transactions and that reviewing of trends in transactions are having and are expected to continue to have significant and helpful transformational effects. This slogan appears to apply well beyond federal payments for Medicare transactions, again because of the Deficit Reduction Act among other reasons. An underlying belief implicit in this transformational slogan is that problems considered systemic in health care can be sufficiently identified for warning, notice and reasonable cures.

One message arising from the transformational results notions is that many entities involved in health care should spend the time to upgrade or strengthen their statistical sampling methodology appropriately. Such action can help further improve the tasks of regulatory oversight and further improve the health care compliance culture as well as enhance results. Some areas with potential for improvements are summarized herein.

This chapter builds upon the well done predecessor work by Perling and Barreau in 2002, among others.[5] It takes a longer perspective over several decades on the developments of statistical sampling in health care compliance, and it includes some recent developments since their 2002 work. This chapter has the following sections: Section II has a review of statistical sampling in theory and application; Section III delves into the widening use of statistical sampling in health care in the United States; and Section IV focuses further on how the government regulatory uses of statistical sampling made the general uses of statistical sampling more effective and more fully transparent. Section V addresses some areas of statistical sampling that currently merit more attention for all; Section VI applies these current statistical areas to current challenges and remedies, especially for the states; and Section VII attempts to look forward for the uses of statistical sampling to further increase health care compliance strength and broaden its topical coverage. Section VIII concludes the chapter.

II. A Brief Review of Statistical Sampling in Theory and Application

In General

When regulators, practitioners or other individuals who are conducting an investigation or research about a population want to know answers to certain types of questions involving data, there are, most often, three options: the first is to study a population in its entirety; the second is to study a sample and to extrapolate the results of the sample to the population. There are many situations where studying a sample is preferred over studying the entire population. The third option would be a combination of the two based on the needs and the constraints of the investigation or research. If sampling were selected, there are two further choices: statistical sampling and judgmental sampling. Again, these two approaches suit different needs and constraints when answering certain questions.[6]

Statistical sampling is mostly about the practice of making statements regarding the characteristics of a large population based on a reasonably thorough review of a representative sample. This practice has many relatively formal rules, largely based on scientific knowledge. These rules concern how a population is defined, how a "sampling frame" of the population is created, how a random sample is drawn, how a thorough review is accomplished and how an extrapolation about the population is determined. Randomness of a sample helps enormously to assure its representativeness. In fact, randomness is essential in order to assert that the sample is an unbiased estimator of the population mean.[7] Stratification of a sample

can help the precision of extrapolations, including appropriate reliability based on given constraints for a sample size.[8]

Typically, statistical sampling provides answers or estimated answers about quantities in a population: for example, the average payment on a claim (a quantity) times the number of claims equals the estimated total payments for a population of claims. However, statistical sampling can also provide answers about qualities of a population: the chance of a claim having an error of "x" in its payment (a quality) times the number of claims equals an estimated total payment error for a population of claims. Statistical sampling can provide combined answers: the average payment on a claim times the chance of a claim having an error times the number of claims equals the estimated total amount of dollars for claims with payment errors for a population. All of these answers remain estimates; but their precision has the virtue of a reasonably known degree of confidence, again subject to rigor; e.g., the incidence of error.[9]

In health care and other industries, statistical sampling also has many other formal rules based on government regulations, court decisions and practical principles developed during field use. For example, the necessity of having sufficient documentation in order to help reach agreement on the legitimacy of a claim submitted for payment has been or is being largely settled. The degree of effort to search for each item drawn in a random sample is becoming more consistent as the definitions of common records; e.g., billing claims and electronic health records, are gradually becoming more uniform and more useful among health care government providers and commercial payors.[10]

A Fundamental Area of Statistics

Sampling with its assumptions of random occurrences and its formulas to predict distributions of occurrences were among the first areas in the broad discipline of statistics to be formally developed.[11] This development happened in large part because statistical sampling is, in its most fundamental respects, a branch of mathematics. Terms such as sets and subsets, permutations and combinations, discrete values and continuous approximations, and others still taught in introductory statistics classes today, were developed for studies of the highest importance for abstract situations and real relationships.[12]

To take an example of sampling from the nineteenth century, a researcher collected data and then applied statistical formulas to test an assumption of whether the incidence of a certain type of accident had a common probability of occurring among similarly-sized groups in a large organization from year to year.

An analogy to the health care industry requires only a small intellectual leap, and it should often be taken in practice to improve the rigor of analyzing and ultimately reducing errors. A regulator or a researcher or manager may want to apply statistical formulas and data about an event of interest, defined perhaps as a sentinel event in patient care, to predict the total number of such sentinel events in a hospital's many wards from year to year. Considering the above, it is a fitting and essential step to apply, as appropriate, such laws of statistical sampling for analyzing the "root cause," or the combined effects of many health care service processes, to help in assessing whether there is a common trait concerning sentinel events for all episodes of patient care in a hospital's wards rather than distinct causes in, say, certain wards or by certain practitioners. Nonetheless, both the common trait and the distinct causes often should be reviewed as will be later shown herein.[13]

An idiosyncrasy of the U.S. health care industry is that the major applications of statistical sampling occurred initially in the area of analyzing and resolving billing errors. Applying analysis and determination rigorously to analyzing and improving quality of care is also accomplishable. This will be a benefit to addressing the challenges faced by management and leaders to try to reduce the rate of health care errors and accidents.

Statistical Sampling's Positive Effect

Regulators, managers and other individuals have applied statistical sampling for a long time in many industries and in a variety of circumstances. There is a straightforward reason: statistical sampling can add value and has added value in quality, usefulness, improved efficiency and reduced risk in its applications.[14] All have, over time, improved older statistical sampling methods and developed new methods which have been tested and applied in different settings. And as statistical sampling methods have improved, their acceptability and their applications have also been enhanced.

1. **From Statistical Sampling to Quality Improvement.** A major example of how managers applied statistical sampling with positive effect is the evolution of William E. Deming's thoughts and practices from theories of statistical sampling to the principles and techniques of quality management. In 1950, Deming published *Some Theory Of Sampling*. This text is dense with mathematical formulas on statistical distributions; it also has a chapter with the interesting title, "The Sample As a Basis For Action."[15] This chapter foreshadowed Deming's later work on quality as well as on statistical sampling, which is more well-known.

   Statistical sampling theory and applications improved throughout the twentieth century. After his major contribution to statistical sampling theory, Deming spent many years in Japan guiding manufacturers there on how to improve the quality of their processes by means of statistical sampling applications. In 1982, Deming published *Out Of The Crisis*. This Deming book has fourteen points for management, including emphasizing long term planning, and building quality into the product and process.[16]

   The evolution of statistical sampling over many years from abstract theory to practical instruction, as well as practice, and as developed and exemplified by Deming plus a few others in the last century, has ultimately constructively affected even the health care industry. For example, in 2007, Virginia Mason Medical Center in Seattle sent its management team to study the Toyota production system in Japan.[17] If history is an indicator of the present trends, the authors believe this Seattle to Japan team will find themselves increasing their use of statistical sampling among other lessons learned.

2. **Some Overall Comments.** As statistical sampling has added value, and particularly as statistical sampling has reduced risks in general, expanded applications have made it easier to progress statistical sampling applications from one field to another and from one area to another. Leaders in government, business and academe have continuously studied previous statistical sampling as well as new concepts to try to make them better. Accordingly, as new concepts and new techniques were invented, they have been tested, tried and increasingly used.

There have also been some occurrences of "failed" or less than successful efforts to use statistical analyses; for example 1) government contract cost allowability estimates, etc., or 2) some poor short-run decisions not to make sampling-indicated improvements timely due to hesitations over the short-term costs involved, or 3) users who did not understand either the subject matter or the application of statistical sampling.[18]

In health care, the misapplication of statistical analyses has yielded at times misleading results. One "bell curve" approach for the statistical analysis of health care evaluation and management coding is an example. A regulator may expect to see a bell curve of codes, with 1s and 2s on the left side and 4s and 5s on the right based on statistical sampling of many providers. However, it is not always appropriate to

apply that rule individually to each physician practice, because a single physician could, in fact, have a higher incidence of 4 and 5 codes. In this illustration, results from the group may not always be applied to the individual. Other challenged examples include extrapolating the results of statistical sampling to time periods, geographic areas or operational areas that were not included in the sampling frame from which a random sample was drawn—both part of a broader problem of taking a sample from only part of the population and then trying to extrapolate to the entire population. Nonetheless, the authors urge "trying" and embracing statistical sampling where appropriate.

III. Early and Widening Use of U.S. Health Care Statistical Sampling

Summary Chronology: Government Postures on Health Care Since the 1960s

The history over the past forty years of the government's expenditures and oversight for health care services in the United States is roughly marked by three periods:

Periods of Government Expenditures and Regulatory Oversight in Health Care

Period	Years	Brief Description
One	Mid-1960s – Mid-1980s	Provider Cost Regulations; Allowable Costs Emphasis
Two	Mid-1980s – Mid-1990s	Emphasis On Payments For Diagnosis Related Groups; Anti-Fraud Legislation; Increased Activity Of The Department of Justice And The "Plaintiff's Bar"
Three	Mid-1990s – 2007	Continued "Constructive Posture" Of Regulation; e.g., Operation Restore Trust

Although all three above periods have involved a government constructive posture of regulation, some believe that a fourth period is upon the United States—a period that could be described as one with a hoped for "transformational result." This was, as previously noted, a time that is to combine government expenditures and quality of care attention in part through improved use of analysis and data, including statistical sampling, as well as some new incentives and to do as successfully.

The Relationship Between Government Expenditures and Oversight

The character of government expenditures—federal as well as state and local—has changed somewhat over the years since the 1960s. However, in no period have these government expenditures for health care services fully facilitated the development of a well-capitalized industry.[19] This thin capitalization has placed increased financial importance, and consequent regulatory stress, on government health care expenditures for services when providers must rely on these expenditures to fund their increasing capital and operating costs.

Somewhat slowly but surely—now viewed years later—new major health care government laws, regulations, rules and court interpretations expanded along with these growing government expenditures.[20] In the first phase of government health care regulation, such regulation focused to a great extent on provider allowable costs and the related structure and amount of government reimbursement. Federal reimbursement through both Medicare and Medicaid continued to draw significant regulatory incurred cost attention and then shifted to place a greater emphasis on provider billings and claims related to payment for services as the periods from the 1960s progressed.

Close Tolerance Review by Government not Feasible

The breadth and variety of government health care support as well as government outlays continued to increase into the new century:

Personal Health Care Expenditures 1970 and 2005, In Billions[1]

Year	Total	Private[2]	Public[3]
1970	$ 63	$ 41	$ 22
2005	$1,661	$746	$746
Increase	$1,598	$874	$724

1 Source: Table 5, "Personal Health Expenditures" Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group.

2 Includes out-of-pocket, private health insurance and other private funds.

3 Includes Medicare, Medicaid, state and local plus federal.

The sheer volume of all of the health care support transactions was and is immense. And trained government resources to supervise, carefully check and review the reimbursement details were simply not available.

The time required to perform careful, detailed and complete reviews of either the millions or billions of transactions in health care would have required significant government expenditures over and above reimbursement expenditures for services themselves. Such efforts would have involved significant subcontracts, internal hiring and training dollars, years and years for reasonable government efficiency and years of "start-up" political criticism. Accordingly, the careful, detailed and complete checking of reimbursements was seen not to be cost-effective.[21]

However, when government reviewers conducted compliance reviews on a very limited basis, they found both carelessness as well as some "gaming" and enough of the latter to cause serious concern. Moreover, some of the problems indicated a lack of internal controls or a lack of billing control together with some disdain for appropriate compliance.[22]

Increased Federal Regulation Led to Increased Statistical Sampling Challenges

Government regulation and oversight grew along with health care coverage and expenditures. Moreover, the added laws and regulations were "following the money." Not surprisingly, new laws sought to reduce the risks of fraud, waste and abuse. Regulations followed regulations; more rules followed as well.[23]

In the later chronological periods, the federal health care Inspector General Offices were "beefed-up" or were created in the first place, and some state agencies followed suit. Virtual armies of audit and review eyes were increasingly focusing on government outlays for health care.

As indicated above, and to "fast-forward" a bit, compliance recoveries reported for 2006 were said to be some $2.2 billion, a rough indication of important health care challenges in the past as well as improvements yet to be accomplished.[24] In that dollar recovery area, it is fair to point out that settlements in any year relating to billing or quality of care often involve health care claims that had been submitted for payment over the course of several years.

The results of certain of the government reviews and audits involved some generally large and well-publicized requests for repayment:

Some Recent Large Health Care False Claims, Fraud and Billing Settlements[25]

Company	Amount Recovered *(in Millions)*	Year Settled
HCA	$631	2003
HealthSouth Corp.	$325	2004
Schering-Plough Corp.	$293	2004
Tenet Healthcare Corp.	$900	2006

Among others, frequently used descriptions for allegations relating to these recoveries include false claims, fraud and systematic fraud, billing violations, kickbacks, etc. As expected, providers and commercial payors challenged the government's preliminary findings and conclusions; and many of the providers' as well as commercial payors' challenges were reasonably founded as well as ultimately constructive as to the statistical sampling methodologies. Some of the challenges related to the statistical sampling methodologies themselves.

IV. Greater Effectiveness and Transparency in Statistical Sampling by Government

Government Uses of Statistical Sampling in Health Care

For many users to better understand the improvement and progress in health care statistical sampling, it is helpful to gain some insight involving both the past as well as today and the future.

The past health care statistical sampling progress is valuable for expansions of compliance monitoring; e.g., the stepped up emphasis on quality of care compliance where new learning curves will happen. Knowing the progression of law and regulations is useful to health care. Also important is avoiding regulatory challenges and mistakes, which could slow down health care statistical sampling and quality improvement progress.

Understanding the present is important so that health care statistical sampling progress becomes even more widely established—e.g., in newer Inspector General programs and state areas. The Centers

for Medicare and Medicaid Services seeks to make recent statistical sampling developments quite relevant in a greater number of health care services in which their funds are involved.

For the future, much of newer health care coverages programs can and will be sampled for compliance. Also, a greater education about fraud, waste, abuse and whistle-blowing, is to occur; statistical sampling will be in heavy demand. And not the least, states' compliance, their partners' compliance, quality of care and more will drive up applications and uses of statistical sampling.

The challenges of more statistical sampling include considerable learning curves—the improvements, the expansions, the training on health care statistical sampling will be summarized herein. The constructive regulatory postures are reflected in betterments indicated.

Government Reviews and the RAT-STATS Statistical Sampling Methodology

It is common in the United States health care compliance to hear the phrase, RAT-STATS, as is explained herein. Statistical sampling methodology and execution involve a broad spectrum of effort from early considerations and planning throughout to reporting as well as use. As indicated and with hindsight, regulatory emphasis on and greater regulatory use of statistical sampling were inevitable—either to test billing practices with a reasonable cadre of auditors as well as review personnel or for other reasons. And, to try to help estimate the overbillings as well as to estimate errors, statistical sampling was also widely used. Of course, there came a clamor from organizations being regulated for more "transparent" (e.g., published) government audit, review and estimating practices, as well as more uniformity in them. This led to a need for more generally accepted use of some sampling disciplines, which, in turn, led to a need for development of more detailed methods for health care statistical sampling users.

Health care compliance testing by statistical sampling developed in more than a few review areas, including Medicare are Medicaid billing, labor law, the Health Insurance Portability and Accountability Act, Stark, Anti-Kickback, Deficit Reduction Act, False Claims Act, etc. Moreover, estimates of damages themselves often have and will have some reliance on or interrelationships with sampling and statistics.

1. **A Word of Caution on RAT-STATS.** The federal government's RAT-STATS program was an important part of this development of statistical sampling methods. Certain discussion matters on RAT-STATS should be clarified in this respect:
   a. Often in health care communication, the phrase RAT-STATS, is used to refer generally to the government's statistical rules. For example, these government rules include specifications on sample size, as will be discussed later in this chapter.
   b. The software program with algorithms that determines sample sizes, draws random samples, extrapolates sample results and performs other calculations under the government's statistical guidance also bears the name RAT-STATS. This statistical software program was developed by the Department of Health and Human Services, Office of the Inspector General, Office of Audit Services. The RAT- portion of this acronym stands for the Regional Advanced Techniques (RAT) staff of the Office of Audit Service. The Office of Audit Service's RAT staff of San Francisco developed a set of statistical modules to assist them in their Health and Human Services reviews and audits. The ongoing development of this software has broadened and continued through today.
   c. At other times, the acronym RAT-STATS is used to refer to the government's health care statistical rules, and simultaneously, to refer to the broad spectrum of sampling methodology and effort the government used or uses.

The authors choose not to change the above rather "loose" communications and instead try to be reasonably clear in their own usage herein.

2. **Flaws in Initial Practices and RAT-STATS Software.** It is also noteworthy that, in these earlier years of use of statistical sampling in health care regulatory compliance, initial resistance in some health care provider organizations as well as payor organizations suggested that statistical sampling and quality added a layer of unneeded costs.[26] This particular discernment on the use of statistical sampling should have been an issue for only a few years until training had sufficiently spread among leadership and other persons in these organizations.[27]

   Government health care agencies as well as provider and commercial payor organizations had caused many of their own problems with sampling and made some of the same ones a major challenge.[28]

   The government's so-called RAT-STATS practices—broadly defined and especially the ones early made transparent—were under-designed for their eventual use as can be seen with hindsight. One example was a shortness on planning and execution coverage guidance, which caused sampling work to often be hastily done or not well cross-checked. Another example was stratification, a widely used statistical sampling technique, which was not available in the module for drawing random samples by the RAT-STATS software.[29]

   Provider and commercial payor organizations trying to use statistical sampling did the same things as the government did, sometimes believing that using the government's method might be more acceptable to government auditors and reviewers. Both found themselves in unnecessary disputes or in long negotiations.

Generally speaking, there was a lack of rigor, including reviews and cross checking, in the earlier health care statistical sampling process. Some specific issues were: 1) conclusions drawn, as well as projections of overbillings, from early statistical sampling work were initially not well-communicated to audiences or readers; neither were there requirements that errors be reasonably well-defined; 2) legal judgments were made in carrying out the statistical sampling process and were not reasonably disclosed; and 3) the populations being sampled were not as well understood as they might have been, nor were the procedures and practices underlying the population itself sufficiently well understood.

RAT-STATS Improvements

Responses to criticism and then proposed improvements of the use of RAT-STATS and RAT-STATS software contributed to its improvement from the 1970s to its current version. Examples of these proposed improvements included the drawing of random samples by strata and the extrapolation of sample results based on proportions of errors rather than average amounts of errors. Of these two, the former has been addressed reasonably well in the current version, while the latter remains and will likely be proposed on again for change in forthcoming versions.

As with many tools for solving technical use challenges, each version of RAT-STATS has needed to find a balance in the trade-off between ease of use for the application of common steps and flexibility for the application of less common, or "safety-valve" steps. Moreover, no software can replace the benefits of education and experience in organizing and monitoring the quality of all steps in the use of statistics in health care.

In 2002, the statistical modules were transferred to the Microsoft Windows-based environment.[30] This transfer of computer environments reflected the demand for a more common set of statistical modules for sampling and extrapolation purposes that could be relied upon by both government auditors as well as providers and commercial payors conducting their own reviews. The current version of RAT-STATS, a further improvement, was released in early 2007.

Statistical Sampling Advancements Through Rulings and Court Decisions

Some government rulings as well as court decisions aided use of statistical sampling. The need for statistical sampling in health care billing had become steadily more significant as the flow of money from the government increased and the complexity of the claims submitted by the providers simultaneously increased. The Health Care Financing Administration[31] justified reliance on statistical sampling for post-payment reviews of claims for essentially pragmatic reasons. For example, in 1986, shortly after the introduction of the more complex Prospective Payment System to be used for billing claims for inpatient hospital services, HCFA issued Ruling 86-1 which, in part, said, "a case-by-case review [of claims] could require a significant diversion of staff from ongoing claims processing, and the cost of determining the cost of overpayment would be prohibitively high unless a sampling method were used." Among other things, this Ruling allowed and, indeed encouraged fiscal intermediaries, contractors to the federal government, to sample adjudicated or paid claims according to some sampling guidelines.

These earlier post-payment reviews and their guidelines were argued about in disputes and in courts by health care providers. Yet, despite assertions of a lack of rigor or related issues, the courts did not totally endorse and did not rule out the use of statistical sampling as a means of investigating practices and determining damages. Courts ruled clearly that the use of statistical sampling did not violate constitutional principles such as the right to a trial by jury. Courts did maintain, considerably later, and with less precise instruction, that any use of statistical sampling was required to generally conform with other constitutional principles such as due process.[32] More recently, in a provider-payor decision, where a contention that a statistically insignificant sampling did not state a claim for fraud was denied, the court held that the language of the parties agreement called for a "survey" of HMO member satisfaction.[33] Such court rulings, even when statistical results were barred from a particular case, contributed to the continued improvement of sampling techniques in health care claims. Reconciliation of reasonable statistical sampling techniques with procedural principles of United States legal processes helpfully occurred over time.

Courts repeatedly supported the notion that the government had an obligation to monitor the expenditure of public money and that such government monitoring should be done in a reasonably cost-effective manner.[34] The courts then further articulated some guidance on their view of process with respect to statistical sampling for the estimation of overpayments in a population of claims based on statistical sampling to be acceptable when resolving a dispute. This court guidance on process included:

a. The health care statistical sample must be reasonably representative. Random sampling is a method to create a representative sample. Representativeness is important. As an example, improper stratification could lead to the health care sample no longer being considered representative, absent a sound justification for why such stratification were established as appropriate. Other unsound practices can be equally weak.

b. A health care provider should have an opportunity to challenge and rebut the results of government-estimated improper payments in the statistical sample. A successful rebuttal could generally be made of the sampled claims that were considered representative of the others in the population

by the government. This opportunity pretty much appeared to remove some part of a complaining provider's ability to argue a generalized notion that each and every claim considered improper for payment had to be identified in the government's population; e.g., for damages.[35] The court's posture meant that the government did not have to prove each and every dollar of error detail-by-detail in the entire population. Nevertheless, the government or another plantiff must prove considerable in its statistical sampling process.

c. The government should produce thorough documentation of its entire sampling process in the case. For example, if a health care statistical sample were maintained to be representative of, say, 10,000 claims, then a summary listing of all 10,000 claims in the population had to be produced along with the reasonably detailed listing of the representative claims in the sample. This summary listing of the population is termed the sampling frame. Having access to such basic elements as a sampling frame for review and challenge by the health care provider—so that the sample results were replicable—was considered fundamental in relationship to the legal principles of fairness as well as due process.

d. The government has the burden to demonstrate that its health care statistical sampling method was reasonably reliable. This burden in part established the need for sufficient expertise and training in the government and its experts.

e. The government has to follow its own health care rules and regulations on how to conduct statistical sampling. For example, where the government had rules that set criteria for a minimum sample size, then a provider could challenge the extrapolation of overpayments from a sample whose size was below the minimum.[36]

This guidance process was established in court rulings as well as in administrative rulings and agency rules. Over time, health care counsel, auditors and reviewers could incorporate the guidance into their practices to try to assure better and more timely resolution of disputes based on statistical sampling.

The use of statistical sampling was thus advancing in practices among health care providers and commercial payors while the court guidance on process for statistical sampling was being formalized in case after case. Much later, courts ruled that the use of statistical sampling could be required for the monitoring or enforcement of a settlement; and thus parties often had to become "ready" by improving and testing their own policies as well as practices and their use of statistical sampling. These developments may call for a few comments regarding government and provider field practices involving statistical sampling. To summarize, the law and related court decisions—relevant and important as they are—cover some small percentage, estimated at less than 1 percent, of all instances of health care administration. The greater percentage is accepted billings and quality of care issues, including negotiated settlements. Thus, it should not be expected or assumed that there will be detailed court rules on the many possibilities of disputed areas. Using the sampling and related projections as an example, it is fair to say that, in health care negotiations and settlements, both sides may very well have different views and positions. The government may believe the "law" and regulations call for, or totally allow and support, sampling as opposed to requiring more reviews of all disputed dollars item by item. The provider may not agree or accept the government positions on the law and regulations. Both, most often, can and do discount their own and the opposing position in negotiations.

Also, often the result is an agreed-on "bottom line" settlement on gross dollars as well as general, but not all, practices. Neither agree to all of the details. And hence, since so much settlement occurs, the field practices indirectly support the use of statistical sampling for the parties' benefit on an overall basis. Statistical sampling in health care thus goes on in its application in major ways.

Next, Corporate Integrity Agreements with the government[37] formalized what were then emerging field practices among the health care industry. The review of health care claims under a Corporate Integrity Agreement involved another significant improvement in the use of statistical sampling of claims. These agreements frequently mandated ongoing current statistical review of a provider's claims rather than only post-payment reviews. Many Corporate Integrity Agreements required that statistical sampling should be used to review relatively more claims being submitted for payment contemporaneously and required relatively fewer claims submitted in earlier periods to be reviewed than past practices.

In other words, statistical sampling thus had become established, acceptable and agreed to be part of quality improvement in health care as well as part of court processes and dispute resolutions, at least in the area of claims for payments and increasingly for quality of care issues.[38]

Evolution in Health Care Statistical Sampling—Challenges and Progress

The challenges of applying statistical sampling in health care have led to an evolution of some statistical sampling approaches and practices worthy of review.

1. **Approaches for Determining Sample Sizes for Reviews.** It is common to find, among health care providers and commercial payors and other parties who rely on statistical sampling, misunderstandings regarding appropriate sample sizes for purposes such as billing reviews, quality evaluations or other compliance testing.[39] These sample size misunderstandings have largely arisen among practices in the field, such as when management must rely on sample results for monitoring health care business practices and deciding to change business practices. These misunderstandings have also arisen in the courts, when one party challenges the validity of a sample to make extrapolations about a health care population based on the sample size in addition to challenging other parts of the methodology of the statistical sampling.[40]

   In the authors' experience, health care individuals appear to be more likely to question the validity of conclusions made from small sample sizes than from larger sample sizes. It should be noted, however, that the standard deviation of a small sample will be greater than that of a larger sample—other factors being equal. Some fairly common misunderstandings as to sample size in health care have long existed. In reality, statistical sampling must be carried out well and used effectively—and when that occurs, small sample sizes can be just as meaningful as large sample sizes.[41] The authors agree. However, some individuals also hold that CMS benefits in important ways from its "sponsorship" of small sample sizes in statistical sampling.

   A review of major trends on health care sample size may be helpful. Some methods for determining sample sizes, when estimating overpayments in a health care claims population, illustrate the progression of statistical expertise in the health care industry. This progression began with either an intended or accidental fairly rigid specificity of sample sizes in order to "get going." The progression had an intermediate phase of reliance on variance-influenced formulas, and it has now advanced to better deliberative planning and interpretation on sample sizes. This statistical sampling progression was logical and most of the related current developments are sound.

   The appearance of rigid specificity, with respect to sample sizes, had been codified in the Medicare Carriers Manual, Part 3, Sampling Guidance Appendix (initially in the 1980s).[42] This Appendix indicated that a sample size was required to increase as the anticipated magnitude of overpayments in a claims population increased. The following table is developed from information in the Sampling Guidance Appendix:

Overpayment Range	Sample
Less than $5,000	80
$5,000 to $14,999	125
$15,000 to $24,999	200
In excess of $25,000	320

If the anticipated overpayment were less than $5,000, then a sample size was required to be at least 80 claims. If the anticipated overpayment were between $5,000 and $15,000, then the sample size was required to be at least 125 claims, and so on until an anticipated overpayment greater than $25,000 required a sample size of at least 320 claims.

This form of specifying sample sizes was useful while it was in effect, because the guidance was a way to initially establish what constituted a sufficient sample size by government, and possibly providers and other payors. As said earlier herein, in cases where a government agency selected and analyzed a claims sample, a provider could and often did challenge the extrapolated overpayment for the claims population when the government's sample size was less than the specified size.

However, these specified sample sizes, in that early state, did not seem to bear a relationship to the statistical principles of confidence levels and precision for approximately interpreting a sample's representativeness of a health care claims population. Consider two different populations of such claims: 1) A first health care claims population with a small number of health care claims having large and varying overpayment errors, or 2) a second health care claims population with a large number of consistent and small overpayment errors. Suppose that these two populations, which vary in the characteristics of their overpayment errors from claim to claim, actually have the same estimated amount of overpayments among all of their claims in total. Yet, the first health care claims population will require a larger sample size than the second health care population to achieve the same precision in estimating its total overpayment. The larger the variation in errors, the larger the sample size must be to achieve a defined precision. A sample size should be influenced by the variation of the errors in a health care population when a reasonably consistent precision is a goal of the statistical sampling.

That earlier specification of sample sizes, as set forth in the Medicare Sampling Guidance Appendix, did not take the consequences of variation among claims into account, even though such variation was fundamental to the motivation for random sampling for determining billing overpayments in the first place.

Upon the publication by the Office of Inspector General's Self-Disclosure Protocol in 1998,[43] determining sample sizes by means of a variance-influenced formula had the effect of overtaking or replacing the earlier specified sample sizes. The Self-Disclosure Protocol laid out the major characteristics of conducting a statistically valid sample, just as the Medicare Sampling Guidance Appendix had tried to do. There were two major changes.[44]

First, the Inspector General Self-Disclosure Protocol constructively provided guidance on the use of a "probe" sample to help users understand and estimate the degree of variation for an error, such as an overpayment amount, in a claims population. Essentially, the probe procedure involves taking a preliminary sample to estimate the population variance and using that estimated variance to determine the sample size needed to achieve a pre-determined level of precision. The benefit at

issue for health care users involved the potential for segmentation of the claims population with respect to errors. For example, there might be lower frequency of overpayment error for claims with large payment amounts. The understanding from a probe sample could also be developed in a stratified sample in which strata (i.e., non-overlapping portions) of a health care claims population would be sampled in different proportions depending on different error rates.

Second, the estimation of the degree of variation for an error became a variable in a formula to calculate a sample size, or sample sizes if stratification were to be adopted. The Office of Inspector General's approach calibrated its desired confidence and precision to the anticipated variation based on the probe sample. Because of this change to a variance-influenced formula, the sample size required for estimating overpayments for health care claims could be easily challenged not by whether its size met some rule based on the magnitude of the anticipated overpayment, but rather by whether its size were sufficient to achieve the government's standard for confidence and precision. This was hoped to lead to a more fair result. The Office of Inspector General could reject a provider's estimate for total overpayments if that estimate's desired confidence level and precision had not been fulfilled in the determination of the sample size.

After a gradual transition in the last few years, a more deliberative planning and interpretation of sample sizes now is occurring—e.g., when the parties involved in settling an overpayment issue or putting together a sound sampling plan can agree on such things as stratification and sample size. This agreement among the parties should occur after taking into account many factors from the available information and requires an advanced understanding of the statistical issues by the parties. Such is the current state of self-disclosures that are being developed by providers and commercial payors as well as with advanced understandings with the Office of the Inspector General and also for some internal provider and commercial payor testing.

Following initial recognition of a lack of compliance in billing that resulted in the overpayment of claims for some period of time, a provider may develop a sampling plan—e.g., one including, stratification, sample size, random sampling method and extrapolation formulas. The provider may seek agreements with the Office of the Inspector General (or other relevant official) for this sampling plan before further steps are undertaken.[45] A probe sample remains part of this development. The provider may present multiple factors, such as the anticipated magnitude of the total overpayment and the cost and staffing for conducting the sample review, which affect their proposed stratification and sample size. This deliberative planning and interpretation requires an advanced and timely recognition of what a sample can and cannot achieve, given its size as well as the parties' desired confidence and precision.

Importantly, the Office of Inspector General recognizes that the validity of a sampling plan is contingent upon a provider having a reasonably sound compliance program in place for the longer term.[46]

Even in the current phase of deliberative planning and interpretation, there remain areas where the setting of sample sizes has its challenges. The Medicare Internet-Only Program Manuals now have a section that indicates the validity of their extrapolation of a sample's results, including total overpayments, cannot be easily challenged by a provider solely because of a small sample size. This section's statements merit citing at length:[47]

> The size of the sample (i.e., the number of sampling units) will have a direct bearing on the precision of the estimated overpayment, but it is not the only factor that influences precision. The standard error of the estimator also depends on (1) the underlying variation in the target

> population, (2) the particular sampling method that is employed (such as simple random, stratified, or cluster sampling), and (3) the particular form of the estimator that is used (e.g., simple expansion of the sample total by dividing by the selection rate, or more complicated methods such as ratio estimation). It is neither possible nor desirable to specify a minimum sample size that applies to all situations. A determination of sample size may take into account many things, including the method of sample selection, the estimator of overpayment, and prior knowledge (based on experience) of the variability of the possible overpayments that may be contained in the total population of sampling units.
>
> In addition to the above considerations, real-world economic constraints shall be taken into account. As stated earlier, sampling is used when it is not administratively feasible to review every sampling unit in the target population. In determining the sample size to be used, the PSC or Medicare contractor BI or MR unit shall also consider their available resources. That does not mean, however, that the resulting estimate of overpayment is not valid, so long as proper procedures for the execution of probability sampling have been followed. A challenge to the validity of the sample that is sometimes made is that the particular sample size is too small to yield meaningful results. Such a challenge is without merit as it fails to take into account all of the other factors that are involved in the sample design. *(emphasis added)*

The final two sentences of this excerpt on the merit of challenging a sample size must be interpreted in the context of the entire excerpt from the Manual. The authors believe the assertions of "…without merit" to be a serious exaggeration; fortunately an exaggeration mooted by the document itself. The earlier and later sentences in the very same paragraph have important conditions and considerations on how a sample size is to be determined. Otherwise, an indication by the government of approval relating to sample sizes that are quite small could create a risk of the "luck of the draw" for those providers and payors whose claims have high variation in errors on a claim-by-claim basis. Without the above context, there may be an unclearness risk created for providers, some of whom now may believe they have neither the easy ability to reference the government's own specification of sample sizes developed over time nor the easy ability to place a sample size in the context of a reasonable confidence *and* precision.

The appropriate current means for determining sample sizes requires continued understanding that all parties must have some ability to influence the process and methods, especially at the early stages of the sampling process. Such a "safety valve" ability for reasonableness exists presently from understanding that relevant parties can negotiate sound approaches for the circumstances; and that appeals are possible.[48] Some issues for clearness or agreement might include 1) planning including, sample size concepts, outcomes expected, reporting, measurement objectives, etc; 2) sampling methods including random, judgmental or combinations, sample sizes and validity, probe use, significance issues, sample records and work done in the sampling; 3) overall work papers and records; 4) adjustments to planning; 5) new issues (plus changes needed); 6) projections and extrapolations; and 7) reporting communications, etc.

2. **Probe Samples, Strata and Outlier Identification, Alternative Estimators.** A probe sample is a relatively small sample that is reviewed before a full statistical sample is developed. The purpose of a probe sample is to learn more about the characteristics of the population, and to learn more about how the review can best be undertaken—the two learnings are strongly inter-related. The use of a probe sample was part of the Self-Disclosure Guidance by the Office of the Inspector General. The authors have observed that probe samples are now regularly considered and often undertaken, and the documentation of their results are reported for many types of health care reviews.

   The evolution of statistical sampling concepts in addressing strata, outliers, and other population characteristics in health care has been "coming along," but has not been as effective. Strata occur in populations when an attribute of relevance or possible relevance, say an error rate for billing, tends to vary based on other, observable characteristic(s) of the items. For example, claims for large dollar amounts may have a different error rate, because they are processed by the provider or the payor in a different manner than claims for small dollar amounts. Outliers can be thought of as an extreme type of strata; e.g., claims for very large dollar amounts tend to receive the greatest scrutiny when being processed. For both strata and outliers, analysis prior to the review should be undertaken to confirm their presence in the population and the reasoning for incorporating these characteristics in the review should be established before the review begins. There has been less guidance from regulation and court rulings on outliers and strata, compared to guidance on sample sizes and probe samples. The authors have noted that reviews relying on stratification and outliers methods tend to be accepted when done reasonably and disputed when not.

   Alternative estimators, such as ratio estimators, are often theoretically sound, but they have not yet been fully embraced in the evolution of statistical sampling in health care.[49] A ratio estimator uses the percentage of errors in a sample, rather than the average amount of errors in a sample, to estimate the total amount of errors in a population. A ratio estimator has some advantages—for example, when reviewing errors for payments on claims, a ratio estimator will not predict that the total amount of errors exceeds the total amount of payments in the population, when some other methods will make that mistake. And a ratio estimator can be more efficient; i.e., have more precision for the same sample size and confidence level, than an average estimator for some populations. However, the average estimator appears more commonly in the guidance in health care regulations that have been recently published.[50]

3. **In Process—Documentation of Statistical Sampling.** The requirements for increased rigor in sampling mechanics, such as documentation and other areas, were alluded to in the section on government rulings and court decisions. Documentation has improved with respect to records of the sampled items reviewed and records of the complete statistical sampling process.

   One element of documentation improvement involves the records of the characteristics of items being reviewed. Such records should be defined reasonably consistently for all of the items. When feasible, this consistency should be planned for before the review of the sampled items is begun. This improvement can be assisted by what is learned from a probe sample (discussed below). Further, the absence of contemporaneous documentation for an item being reviewed in a sample should be addressed early. It is common, perhaps to an excessive degree, to follow unclear assertions about medical chart documentation—"If it is not documented, then it did not happen." Effort can be necessary to establish alternative reasoning or facts that establish the characteristic being reviewed. In any case, such agreements for documentation should now be regularly established before the items in the sample are reviewed.

The second element of improvement for documentation involves the records of the complete statistical sampling process. These records should enable the reviewer to present information about: the testing objectives, the definition of the population, the definition of the sampling unit, the definition of the strata (when applicable), the sampling frame, the confidence level and desired precision selected, results of the probe sample (when applicable), the sample size, the technique for random selection, the random seed number(s) that can be used to recreate an identical random draw from the sampling frame, the results of the review for every item reviewed in the sample, the extrapolation methods, and the conclusions drawn about the population. Moreover, it should be possible for independent reviewers to understand the work, to test it appropriately, to recreate the sampling itself, etc. It is now common for such documentation to be presented for self-disclosures, for settlements and for some court rulings. Statistical experts may often be relied upon for helping present and explain this statistical sampling documentation.

4. **In Process—Stepping Up Rigor.** Health Care work with statistical sampling still requires considerable improvement in rigor. Constancy of purpose and continual improvement must occur. Long-term planning must replace short term reaction.[51]

5. **Training in Statistical Sampling for Health Care Compliance.** Based on the experience of the authors, a number of constituencies are well-along in training on statistical sampling. Examples include: 1) presentations on challenging the use of statistical procedures at a recent conference of the American Health Lawyers Association, 2) presentations on the many advantages and some cautions of providing self-disclosure assessments, 3)the distribution of training audit manuals that include statistical sampling sections by Medicare and some Medicaid agencies and 4) training sessions at consulting firms that provide investigative and compliance services to providers and commercial payors in health care. Constituencies who are well along in statistical sampling training should keep up the good work. Constituencies getting started along those training lines should step up their statistical sampling work training. And those not started should "get going."

Some Developments in Use of Statistical Sampling in Health Care

1. **In General.** The correlations of anti-fraud, waste and abuse legislation with other increases in government health care support, as well as with major refund recoveries by plaintiffs, was not accidental. Unfortunately, the health care industry had earlier welcomed the increased government funding somewhat naïvely, without stepping up its own control and compliance efforts.

2. **Two Significant Changes—Quality of Care and Preventive Care.** It was not possible for early health care system reviewers—government and corporate—to miss the fact that the quality of care delivered to people varied greatly from place to place as well as from one health care provider or commercial payor to another.[52] Moreover, in some cases, it has been alleged and sometimes found that the care delivered or reimbursed was or appeared to be inadequate or possibly not delivered at all—at the least, this became a possible cause for concern. Rules, reviews and audit approaches have expanded their regulatory reach toward quality of care—also not an accidental development.

   Similarly, it was generally noted by reviewers that considerable compliance effort was devoted to past health care billing problems and for past quality of care problems, which conceivably might have been avoided. A corollary concern thus arose that too little compliance effort was being directed toward avoiding health care problems before they happened or before such problems

became large. Savings were believed possible. Compliance requirements and efforts thus began to shift toward greater attention on, as well as greater investment in, quality of care and in steps to try to prevent or reduce all types of health care problems. Most health care constituencies joined in both as goals, although some were a tad tardy in their "getting going."

This "newer look" of compliance did not set aside important health care compliance issues related to earlier regulatory emphases, such as billing. But, attention moved toward more balance on a broader range of regulatory compliance and with related newer compliance objectives training as well as training in general.[53]Quality of care is not being fully reviewed yet for compliance, but statistical analyses, including statistical sampling, is happening and increasing in quality compliance. As an example, quality initiatives presently underway include the pilot Medicare studies for physicians and hospital outcome measures; and some reviews of these initiatives are now being published. Yet, harking back to the billings paradigm, what is being reviewed currently often has a focus on inappropriate care or omitted care.[54] To the extent that providers have been paid for this care, the services may even be considered part of "false claims."

V. All Constituencies Have More to Do

Government Initiative in Developing and Improving Statistical Sampling

The federal government has had a reasonable group of strong regulatory leaders who exercised good judgment as a huge wave of new laws and stepped-up regulations swept over the United States health care industry. Many of these federal individuals deserve considerable credit for constructively dealing with the health care regulatory issues described herein. Major regulatory errors could have had enormous negative consequences to United States health care.

Perhaps the best news is that internal government regulatory "panic"[55] did not happen. The second best news is that government health care regulators not only listened to criticism of their new and changing rules and then often moved to revise major areas of those early rules, but they also were reasonably timely. Regulators also strengthened training—a step often either omitted or done tardily in the usual regulatory world. The major importance of this part of government regulatory improvement in health care is this: the recognition of the need for change by identifying the improvements and curing, or trying to cure, the problems happened in a timely way with minimal defensiveness.[56]

The government health care regulatory team has had and still has a daunting task as their work relates to health care generally, as well as in its use of statistical sampling. Since statistical sampling was the best way to proceed in many circumstances and in the history now available, it is fair to say now that statistical sampling played a major role in the government's health care progress in those earlier times.

Assistance in Education and Training by Private Organizations

Professional groups, such as the Health Care Compliance Association, the American Health Lawyers Association, the American Bar Association's Section on Health Care and others,[57] responded to the needs of providers in helping them better understand health care compliance needs.

These groups also educated audiences and other parties on the early uses of statistical sampling in health care, through educational seminars, journals and newsletters. They contributed professional time

to the identification of and thoughtful analysis of statistical sampling issues, particularly as those issues were raised in compliance activities as well as for related recommended improvements.

For example, individuals constructively challenged weaknesses in the initial RAT-STATS concepts. Their urging of improvement in RAT-STATS helped in the identification of problems and, as developed above, some important improvements happened.

Reliance on Statistical Sampling by Commercial Payors and Providers

Since the government has tried and has reasonably succeeded in the use of statistical sampling, many health care providers and commercial payors seem to be doing more with statistical sampling than ever before.[58]

CMS and its predecessor agencies, as well as what are called Medicare Intermediaries and Carriers, have long relied on statistical sampling to help assess billing accuracy in their utilization review and audit efforts. However, providers early on primarily relied on statistics to challenge the results of some of the regulatory assessments in reviews as well as audits. It has only been within the last decade or so that providers and commercial payors have developed more effective formal compliance programs and have begun to broadly test their organization's performance in accordance with their own testing programs, many times through statistical sampling.[59] In other words, with the development and implementation of formal compliance programs, health care providers and commercial payors have more and more developed their own processes to test compliance and have themselves developed sampling programs and processes. Instead of primarily critiquing others' methodologies, providers and commercial payors have become more adept at sampling and the application of results they obtain through sampling for better control.

A Bit More Dispute or Case Information

The authors have reviewed a judgmental sample of recent litigation cases that involved the use of statistical sampling and found that better basic and established statistical sampling concepts reasonably used have not often been successfully challenged; although, as expected, any incorrect application continues to be disputed.

A few health care administrative law and court cases involving statistical sampling have been briefly mentioned in this chapter, mostly in general. References to sources for those and more are provided. The good news is that the health care administrative law and court decisions show a relatively positive trend overall. It appears that such administrative determinations and the law are "settling in" and making the best of the regulatory as well as the legal health care situation constituencies find for themselves. It may be noteworthy that compliance officials in providers and payors as well as state personnel need to pay careful attention to transparency in their statistical sampling efforts. The private organizations need to understand that their government counterparts should follow state and federal requirements for rule-making in the statistical sampling area of audits as well as reviews. This is an area where room to improve is important and where counsel's assistance may be helpful because formulas are quite vulnerable to inadvertant abuse.[60]

Our health care systems cannot rest on their favorable laurels for a variety of reasons, and the same is true for health care use of statistical sampling in control as well as compliance. First, the states' work on Medicaid compliance is significantly increasing. Second, more whistle-blowing state activity is forecast for the near future.

VI. Users of Statistical Sampling for Health Care Compliance—Still a "Way To Go"

A large number of health care payors and providers beefed up their compliance policies, practices, organizations and personnel. As mentioned earlier, training helped as well.[61] Such health care firms both contributed to compliance training through their own leaders and also relied upon some helpful consultants, advisors and professional groups. Much has also been done in strengthening compliance competence, improving quality control, determining and refunding overpayments voluntarily and generally improving their organizations.

Increasing States' Use of Statistical Sampling in State Reviews and Audits

Medicaid expenditures for health care—like Medicare outlays—have risen dramatically over the years since the 1960s. However, states' compliance efforts have lagged significantly—as pointed out by the Comptroller General Accountability Office reviews, by the Centers for Medicare and Medicaid Services, by the Health and Human Services' Inspector General and others.[62] By all relevant measures, it is known and asserted and was predictable that Medicaid and other state agencies would also be increasing their health care compliance efforts.

There is an unwritten—and not often talked about—policy and practice in government health care regulation as well as in its administration. Such an approach has benefited United States health care importantly. It has allowed constructive and flexible a) resolution of disputes, b) helpful implementation of greatly increased regulatory rules to abide by and c) considerable efficiency, among many advantages for the appropriate regulatory circumstances.

The concept permitted regulatory discussions, negotiations and settlements with great health care savings, including the varied uses and improvements to date of statistical sampling. Such a regulatory approach requires judicial-like fairness in results and a good deal of transparency. Many states and some of their partners have used and continue to use statistical sampling more and more in their utilization reviews, often for quality benefits as well as compliance.

State Medicaid agencies' occasional tardiness in their compliance efforts is often attributed to the two sorts of "partnership attitude" these agencies have: 1) with federal Medicaid, as well as 2) with providers and health plans.

Unfortunately, a shortfall in that partnership posture is a lack of or quite inconsistent health care compliance—something that can be a serious problem downstream for these state agencies unless remedied reasonably soon.[63] The authors recognize that Medicaid agencies are often challenged by providers because of payment levels, and that Medicaid agencies view their existing provider network as important to reasonable state care.

State agencies have also been challenged in their efforts at monitoring managed care programs. They often lack valid encounter data that are necessary for some of their reviews.

As discussed earlier in this chapter, the Deficit Reduction Act and other requirements are likely to cause state Medicaid agencies to increase compliance activities related to Medicaid payments. CMS, for example, is now planning the implementation of the Medicaid Integrity Program. CMS will contract with eligible entities to help carry out certain specified activities including reviews, audits, education and identification as well as recovery of possible overpayments. CMS plans to further increase its oversight of

state program integrity efforts as well as provide training and other practices guidance to state program integrity units.[64]

State Medicaid and other related human services agencies should not, however, wait until regulations are in place to try to comply with federal intent and to test their own compliance with other federal regulations that govern their programs. There are numerous federal regulations and rules affecting many aspects of Medicaid payment that states should consider, ranging from provider certification to policy and eligibility. States should also test to assure that they are in compliance with their own state plans, are transparent in their practices and are properly, not marginally, compliant with the intent of federal laws.[65] Accordingly, the authors suggest that the states' lagging compliance development leads to a corollary prediction of a huge use for statistical sampling in the near future in both states' and providers' compliance work relating to Medicaid. It is also likely that other state-purchased or state-directed health care services will be similarly evaluated for compliance with policies and regulations.[66]

In addition, statistical sampling could be used by CMS and internally by state agencies in a variety of increasingly relevant areas, including education and training of their staff as well as contractor staff, 1) to assess whether managed care entities have compliance programs and are testing them; 2) to determine whether data sharing (where permissible) of a useful nature occurs, as well as; 3) to assess whether state payment accuracy study results are being followed-up with the studying of encounter data and more.[67]

One Challenge of Statistical Sampling for Underpayments and Overpayments

It may be appropriate to point out one more health care area where statistical sampling has and has not been administrated as fairly as it might have been—the area of counting error effects. It is not pleasant to hear of a provider being wrongly overpaid; e.g., payments for services not actually provided but billed, or services incorrectly billed. The reverse is not fair either; it is also not pleasant to hear of a provider being wrongly underpaid; e.g., no payments for services actually provided and billed reasonably.

How "errors" are counted as a result of statistical sampling depends on the intended use of the reporting. If an overpayment amount is to be calculated to assess payment accuracy, for example, for refunds or disclosure purposes, it is generally best that overpayment in the sample as well as in any related estimation be offset against any underpayments claimed and vice-versa. This view has—somewhat gradually—increasingly become the administrative health care practice.[68]

If the purpose of statistical sampling is to assess the presence of systemic errors, the outcome of the sampling is a report of the error rate in terms of number of claims billed or paid incorrectly. This is simply for quality control of one sort or another. In this latter case, the error rate should be expressed as the total number of errors (whether they would result in potential over or underpayments). Finally, it is sometimes appropriate that both error rates; i.e., dollars over and underpaid and claims accuracy, be reported.

Sound sampling communication calls for appropriate disclosure of both error rates along with reasonable explanations, if available.[69] Among reasons for all constituencies' increasing interest in such communication is the possibility of what are called "systemic" problems, which can only be discerned from a review of the number of errors, the character of errors, the overpayment or underpayment amounts, etc.

Other Statistical Sampling Challenges

In testing compliance with policies and procedures in health care, a decision must be made as to the most reasonable approach; i.e., a review of the universe item by item, sampling, or some of both. Statistical sampling can be used far more broadly than presently. However, those who are doing the sampling and subsequent testing should address challenges that the authors have noted from personal experience, in compliance reviews, litigation and other compliance and quality of care activities, as demonstrated in the authors' "Top 10" list below:

1	Sample character randomness
2	Sample size accuracy
3	Sample stratification rationale and method
4	Unusual representativeness circumstances (i.e., lack of representativeness)
5	Mistakes and weak "judgment calls" in classifying discrepancies as errors
6	Poor documentation of work performed or conclusions reached
7	Misinterpretation of results
8	Absence of alternative procedures for "look-ups" (i.e., alternative documentation of care having been provided), outliers, etc.
9	Inadequate quality control of sampling work
10	Absence of "hedges" in communication about sampling results

Such health care challenges can be largely cured.

Some Effective Conceptual Remedies

Sampling can provide meaningful and useful information, as discussed throughout this chapter. However, the authors have noted that organizations which adopt the following practices seem to address the above and other challenges successfully:

Requirements and Practices to Remedy Some Sampling Challenges

A Few Requirements	Improved Practices
• Use of qualified personnel,[70] overall and adequate resources, including individuals with health care industry experience	• Individuals entrusted with the responsibility to conduct reviews or audits through sampling have been educated about the uses and methodologies of statistical sampling, are experienced and have access to other individuals informed in these matters if they do not.
• Development of proper and sound sampling objectives and adequate planning, including a well thought-out work approach that defines what is to be tested and the best approach for such testing	• The objectives for a sampling project are sound as well as known and the approach has been designed accordingly. The parameters of the review meet reasonably specific and appropriate objectives. Reviewers do not delve into "new" and "interesting" areas in the course of the review, which could compromise the validity of the statistical sample.

• Rigorous and intelligent adherence to work plans	• Reviewers properly do their work in the sampling. Reviewers do not stop carelessly if the "desired" results appear to have been achieved early (e.g., if no errors have been found in the first 50 observations, discontinue the review); the sampling program has been carried through as designed.
• Clear communication regarding any conclusions[71] and limitations in conclusions; i.e., necessary "hedges," etc. happened[72]	• Communication is judicial (even-handed) to the extent possible. Thorough analysis of findings and reasonably accurate representation in written summaries or reports as appropriate. Although the Office of Inspector General of the Department of Health and Human Services does not appear to give much mechanical leeway to providers in developing their sampling programs in the *Provider Self-Disclosure Protocol*, reviewers should take great care and communicate how they make decisions about compliance or non-compliance with regulations, policies and procedures. In coding reviews, for example, reviewers may disagree. Any conclusions should state these and other issues that affect the classification of an observation as "correct" or "incorrect." Reviewers or auditors should express findings within a range and identify discrepancies as "potential errors" as opposed to asserting them to be "errors."[73]
• Proper documentation	• Planning and execution of the work as well as conclusions call for appropriate documentation as well as appropriate retention of records—the latter being a subject not sufficiently considered most of the time. This issue has been and is important; it applies to government, providers, commercial payors and others. Moreover, there are some sound practices within the recent changes in the Federal Rules Of Evidence relating to retention of electronic records, effective December, 2006. Where litigation is present, threatened or likely, great and careful attention is appropriate for information retention. A genuine and significant interest in documentation by means of electronic health records to help health care improve quality and lower and better control costs has arrived.[74]
• Proper (before conclusions) challenges to reviewer or auditor results	• Internal and at times external questioning reviewers' or auditors' findings should happen; reviewers or auditors can make mistakes, may not have the depth and breadth of experience necessary, or may not have the level of specialization required to test a particular aspect of compliance.
• Involvement of competent legal counsel, as necessary	• Engaging outside counsel where appropriate, so that the review may be conducted under attorney-client privilege or may get outside counsel's experience. Careful counsel consideration of findings and next steps in addressing any problems discovered.
• Use of experts, when appropriate	• When statistical sampling is to be significantly relied on going forward, (as in an overbilling situation or as in a quality of care liability issue) the individuals designing the sampling plan have been well-educated and experienced regarding the use of statistical sampling in testing health care compliance in general and on the issues. Occasionally, for example, when an investigation is underway, or when litigation is pending, an outside expert has been relied on.
• Sound task "wrap-up"	• Timely wind-down of task work, including completion of administration responsibilities, contractual responsibilities, file completions, sign-off, etc.

And, as always when addressing compliance, it is necessary to have "support from the top"—including leadership, support and supervision throughout the statistical sampling evaluation. For the best improvement, the sampling organization and the sampled organization must be "getting it"—to use the authors' phrase—throughout.

VII. Looking Ahead—Crucial Importance of Compliance Strength and Statistical Sampling

The Growing Trend Toward Avoiding Systemic Challenges

There is a modest but growing body of evidence that the government and some organizations providing health care have quite wisely chosen to focus considerable of their current compliance efforts on removing or curing what are called "systemic" challenges. This attention serves as an alternative to continuing to dispute individual issues, prosecute individual health care providers or react to the risk after the problem emerges later.[75] The benefits of this "looking ahead" policy and practice are significant, including helping health care constituencies put scarce resources toward important improvements, before, rather than after a problem becomes serious. The benefits also include the addition of a more experienced cultural element of continuous quality improvement that provides more skill and experience, with the goal of "getting it," in the sense of need for reasonable compliance, etc.

Quality of Care Compliance

As the federal government is moving toward more systemic issues to review and audit on quality of care, then it has been and is up to providers and others to step-up or begin "testing" their compliance relating to quality health care.[76]

One part of finding the problem and curing any quality fall-down is to learn about the following failures or challenges from individuals in health care and from reviews.[77]

1. Systemic fall-downs
   a. From system design challenges
   b. From system installation and integration problems
   c. From "debugging" the systems
2. From testing for effective operation
3. From monitoring analysis indications
4. From execution failures indications

Statistical sampling can both help identify and can later help cure the quality problem.[78]

Quality of care disputes appear to be getting more attention on the criminal side of the legal landscape. It may also be noteworthy that some states have such quality of care cases under their felony laws, all of which means the risk should be very carefully monitored. Moreover, in health care, the opportunities of improvements in quality are immense in their scope.

Not encouraging continuous quality improvement is expensive from most points of view.[79] Compliance strength must be organizational, complete and have strong follow-through. Improvements in quality and compliance should be an important objective of health care organizations as well as complete for all departments and functions. As said earlier, there is no substitute for a strong compliance "tone at the top." Moreover, there must be an effective follow-through in the health care organization. Otherwise, systemic challenges are likely to exist (at a minimum) and to continue.

The authors recognize a lingering tension in the health care industry between attributing errors to systemic processes versus attributing errors to individuals. This tension is appropriate, because either attribution may in fact be correct, one may in fact be incorrect, or both may be correct for a particular matter to be investigated. The tendency in health care too often has been to identify errors by individuals and not to sample or analyze errors in the systemic processes in a timely manner.

There are limits to what can be accomplished by any constituency, but reasonable analysis and information are needed. Statistical sampling can aid in such analysis and decision making areas in health care.

Making Continuous Improvement a Standard in Health Care and in Statistical Sampling

As a general proposition, continuous improvement in general and in statistical sampling is important in achieving improvement in the U.S. health care system across the board. Even so, strengthening of statistical sampling at the level of those persons providing health care services remains necessary as well.

As discussed above, the new RAT-STATS of 2007 was a major improvement over earlier versions. There is still room for improvement, however. For example, the proposal that sampling results should be extrapolated based on proportions of errors rather than average amounts of errors, is a well-established view in statistical theory, but has not yet been fully accepted in published guidelines for extrapolation.

Given that the government has put providers on notice about the "transformational results" sought, it seems incumbent on all parties to note what are called the early "red flags" and move aggressively to respond to them. Statistical sampling can help identify "red flag" items before such items become a larger problem. Early testing and timely responses are still a health care challenge; the tendency in too many instances is to wait until a problem is identified. One way to find red flags early is testing health care policies, practices and systems, including using statistical sampling.[80] An important issue emerging is whether a provider or commercial payor in health care had a system, or a reasonable system, in place to try to achieve quality and related compliance. Simply stated, such quality and compliance systems should exist and organizations should be working hard to achieve quality and compliance levels of excellence. Statistical sampling can be effectively used to help health care organizations determine whether they are moving toward quality and excellence and where they can improve.[81]

For government, other payors and providers, improving compliance in a sound fashion may well involve further consideration of new controls, new organizational structures, new systems and processes, new and improved training and, likely, some new or improved culture.

The result has been shown to lead to and should continue to help achieve better health care quality. One important support process to the improvement in health care practice has been and is statistical sampling. Health care providers and commercial payors should spend the needed time on continuous quality in general.

VIII. Conclusion

Statistical sampling applications have played an important role in health care compliance and quality testing and evaluation. As commercial payors and providers become more skilled in the use of statistical testing and as boards of directors and consumers look to relevant persons to continue to demonstrate the quality of services within health care entities, it is likely that their performance will be measured and reported based, at least some degree, on statistical sampling.

As senior management and boards of directors of health care providers assess where their organizations stand on billing, quality of care and other compliance requirements, statistical sampling can help greatly improve these tasks—if well done. Then, in reviewing whether the improvements are working efficiently, accomplishing their objectives and improving control as well as compliance, statistical sampling can be used effectively in many ways by the same senior management and their chief compliance officers and their directors. Doing so, thus, can help transform the health care culture as well as enhance overall results.

All constituencies who have a stake in health care compliance will benefit from continued assessment of statistical sampling methodologies and related tools available to develop these methodologies. Each must determine how these processes might be improved for each organization. At the same time, use of such methodologies and tools can help enhance performance and improve overall quality. Importantly, the strength of health care improvement must aid those persons who directly deliver health care in the United States. Our people desire improvement.

The spirit of health care improvement, including by means of statistical sampling, must live on.

Endnotes

CHAPTER 1
Developing an Effective Compliance Team

1. Although this chapter focuses its attention on health care providers, the principles and recommendations summarized here are believed to have broader application.
2. This is a common regulatory development. The authors refer interested parties to the United States public utility, defense, telecommunications and railroad industries. What is often overlooked is a key reason for the stepped-up self-governance: The industry fears greater government regulation will result in significantly greater inefficiency, a weaker health care system, poor service and higher costs. In other words, private participation – or a joint public-private solution – is better.
3. As to the often-asked question—"Can an industry, such as health care, avoid greater regulation at a time of increased federal support?"—the short answer is "No." The reason is what is known as the Golden Rule: "Whoever has the gold makes the rules." With greater reliance on government reimbursement comes more government regulation. In the long run, regulation can be eased but rarely is it totally eliminated.
4. The government often requires this direct reporting relationship to the board and senior management.
5. These are generalizations designed to focus organizations on key requirements important to the success of the compliance team. The titles used here may not be the same as those used in some organizations.
6. For a similar grouping, see *Report on Medicare Compliance*, "When a Compliance Program Is a Sham: One Chief Compliance Officer's Nightmare and How to Avoid It," Atlantic Information Services, Inc., November 16, 2000, p. 3.
7. The board of directors should be acutely aware of compliance activities and should ask questions of the compliance team. In addition, the board should be informed about what the chief compliance officer is doing and how the chief executive officer is responding to the findings of the compliance plan. The board should make sure the organization's senior leaders see to it that resources are in place for activities such as corrective actions and training. The board's active involvement will help the organization maintain an effective compliance program.
8. While the chief compliance officer position does not necessarily require the technical or clinical skills such as those possessed by an MD or RN, such skills can be valuable. Moreover, it would be wrong to assume that a technically trained person should not be a chief compliance officer, or to require the chief compliance officer to be, for example, a financial person.
9. Many of the views summarized in this document were shared by the authors with the health care industry on two separate occasions: the first in a presentation to the Association of American Medical Colleges, September 24 – 26, 1997; the second in a presentation to The Minnesota Hospital and Health Care Partnership, November 3, 2000, at the University of Minnesota.
10. The Internal Audit department does have leaders, but generally speaking, has few personnel to spare.
11. Here "operations skills" means pricing, providing health care services, negotiating, anticipating compliance risks, etc.
12. Current organizational thought believes compliance to be an area requiring specialized training and, more significantly, a totally independent function within the organization. If the compliance function is considered to be too "close" to Internal Audit, it could jeopardize the perception of its total independence.
13. It is a fact that the turnover among chief compliance officers is quite high.
14. It is fair to say that demonstrated leadership in compliance and experience in health care would "tax the available pool of leaders too much in today's health care facts and circumstances." Health care compliance has only been a high priority focus in the last few years. Practically speaking, people with experience in both health care and compliance are difficult to find. It may be necessary to get the best person and train that person up by way of competent outsider help and inside organizational assistance.

15. Although a somewhat redundant assertion here, a strong chief compliance officer will have an understanding of, and appreciation for, dealing with complex and sometimes ambiguous regulations.
16. For a brief peek into the United States General Accounting Office's (GAO) views of how the Centers for Medicare and Medicaid Services must not only improve its regulatory efforts, but also develop better regulatory oversight, see "Medicare Improper Payments: While Enhancements Hold Promise for Measuring Potential Fraud and Abuse, Challenges Remain," United States General Accounting Office Report to Congressional Requestors, September, 2000. This report discusses the direction the GAO recommends the Health Care Financing Administration should proceed regarding federal payments, what health care chief compliance officers may be facing in the future, and some interesting background and more current developments.
17. If the chief compliance officer does not possess this statistical knowledge, it is important that he or she seek outside support in this area.
18. Here "the market" means the customers, funding or reimbursement sources, the securities markets for publicly held securities, etc.
19. Not only is this "clear objectives" requirement important for the goal of achieving compliance, it is even more significant for the efficient delivery of health care services.
20. Training in health care compliance need not be unduly expensive. Indeed, such health care training can be Spartan. Health care personnel appreciate training and respond to that training very effectively.
21. For two versions of excellent text on internal controls and auditing, see Richard L. Ratliff, Wanda Wallace, Glenn Sumners, et al., *Internal Auditing Principles and Techniques* (2nd ed.), The Institute of Internal Auditors: Altamonte Springs, FL, 1996.

CHAPTER 2
Keeping the Health Care Sampling Gains Going

1. This chapter does not explore the advancement of statistical sampling in the health sciences itself, which is a different history.
2. In this chapter, "government payments for health care services" refers to payments made by fiscal intermediaries for the services that they provide, as well as the payments these intermediaries make to health care providers, health plans and other parties. Payments are for Medicare, Medicaid and other federally-sponsored health care services. In other sections of this chapter, reference is made to commercial payors; i.e., non-government payors and payments for health care services.
3. The Deficit Reduction Act of 2005, Public Law No. 109-171, offers states with certain "effective" false claims laws a 10 percent incentive reduction in required Medicaid monies to be paid to the federal government. See John Boese, "The Civil False Claims Act and Health Care Compliance: Recent Developments at the Federal and State Level," (Lecture, Health Care Compliance Association 11th Annual Compliance Institute, Chicago, IL, April, 2007). See also John Gibeaut, "Seeking the CURE: With Health Care Fraud Rampant, States Are Urged to Pass Their Own False Claims Acts, But Foes Warn of Windfalls for Plaintiffs Lawyers," *ABA Journal*, Oct. 2006.

   For a compendium of a few relevant statutes and other authorities on this subject see Cheryl Wagonhurst & James Sheehan, "Emerging Government Enforcement of Quality of Care and Compliance Solutions," (Lecture Abstract, Health Care Compliance Association 11th Annual Compliance Institute, Chicago, IL, April, 2007), 15.

   Also, there are proposed rules from the Department of Defense, General Services Administration, and National Aeronautics and Space Administration which contain provisions related to educating and training employees about fraud and abuse as well as compliance issues. See Gabriel Imperato, "New Proposed Rule for Federal Contractors Highlights Fraud Aimed at Training Employees About Fraud and Abuse," *Compliance Today*, May 2007.
4. See: Daniel Levinson, "General Session," (Lecture, Health Care Compliance Association 11th Annual Compliance Institute, Chicago, IL, April, 2007) and Department of Health and Human Services, Office of the Inspector General, "Semiannual Report for Fiscal Year 2006."

5. Lester J. Perling and Annette G. Barreau, "Using Statistical Sampling To Challenge Medicare Overpayments And Assess Compliance Efforts," in *Monitoring & Auditing Practices for Effective Compliance*, ed. John E. Steiner (Minneapolis: Health Care Compliance Association, 2002), 11-24.
6. In simplest terms one form of sampling, at times, is judgmental, and it is often referred to as that: judgmental sampling. Another form of sampling relies significantly on mathematics or using statistical analysis and mathematics; this form is referred to as statistical sampling, and it uses considerable rigor in mathematical principles.
7. Random samples come in two "flavors"—simple random samples and stratified random samples. A sample randomly selected is a number of observations from the population as a whole. A stratified random sample subdivides the population into subcategories, or strata, and randomly selects a number of observations from each strata. For example, a population may be stratified by age strata.
8. The authors assume a reasonable degree of understanding about statistical sampling practice among readers, and encourage all to study or refresh their understanding if and when necessary.
9. Estimates can often be refined. In this example, further analysis could examine whether claims with payment errors have payments above, or below the average for all claims.
10. As a recent example, the requirements of the Health Information Privacy and Accountability Act for providers and commercial payors have helped statistical sampling with respect to common formats for patient records by standardizing claims information that aided sampling with the creation of more consistent structures for data sets that can more easily be tested.
11. This formal development was part of the so-called Enlightenment Era in Western society. For example, in 1809 Carl Gauss developed the formula for a particular probability distribution that is now so widespread and accepted in applications that it is called simply the "normal" curve. Heinz Kohler, *Statistics for Business and Economics* (Glenview, IL: Scott, Foresman and Company, 1985), 284-285.
12. Like much of mathematics that could, say, forecast the rotation of heavenly bodies by a system of equations, it was often said that the predicted values of statistical sampling developed from the abstractions of mathematics had such strikingly equivalent values in the people world we inhabit that there could almost seem to be a "ghost in the machine"—a ghost which knows the answer.
13. This is where concerns about systemic problems meet concerns about individual practitioners.
14. This is a view from 65,000 feet. Some attention to more than short run costs is, of course, necessary.
15. William Edwards Deming, *Some Theory of Sampling* (New York: Dover Publications, Inc., 1966).
16. William Deming, *Out of the Crisis,* (London: MIT Press, 2001), 23-24. Deming's 14 points, originally published in 1982, provided a basis for the transformation of American industry and can be summarized as follows: 1. Create constancy of purpose toward improvement of product and service. 2. Management must awaken to the challenge, must learn their responsibilities, and take on leadership. 3. Cease dependence on inspection to achieve quality. Eliminate the need for inspection on a mass basis by building quality into the product in the first place. 4. End the practice of awarding business on the basis of price tag. Instead, minimize total cost. . 5. Improve constantly and forever the system of production and service, to improve quality and productivity, and thus constantly decrease costs. 6. Institute training on the job. 7. Institute leadership. The aim of supervision should be to help people and machines and gadgets to do a better job. 8. Drive out fear, so that everyone may work effectively for the company. 9. Break down barriers between departments. People in departments across the company must work as a team, to foresee problems of production and in use that may be encountered with the product or service. 10. Eliminate slogans, exhortations and targets for the work force asking for zero defects and new levels of productivity. Such exhortations only create adversarial relationships, as the bulk of the causes of low quality and low productivity belong to the system and thus lie beyond the power of the work force. 11. Eliminate work standards (quotas) and management by objective and substitute leadership. 12. Remove barriers that rob employees of their right to pride of workmanship, including abolishment of the annual or merit rating and of management by objective. 13. Institute a vigorous program of education and self-improvement. 14. Put everybody in the company to work to accomplish the transformation. The transformation is everybody's job.

17. Charles S. Lauer, "What You Can Do," *Modern Healthcare*, May 28, 2007, 25.
18. The related concerns involve the broader discipline of statistics, plus management and health care services and not simply a narrow mechanical definition of statistical sampling.
19. A federal initiative of capital consequence on bed capacity for health care providers was the Hill-Burton Act of the 1970s. Since then, some public capital, accompanied by "roll-ups," mergers, etc., has contributed importantly to a still undercapitalized health care industry in the United States considering capital to include federal, state and local funding, public company debt as well as equity, etc.
20. Notably earlier, for example: Fraud, abuse and waste legislation; D. McCartney Thornton, "The Anti-Kickback and Stark Laws," (Lecture, Health Care Compliance Association 9th Annual Compliance Institute, New Orleans, LA, April, 2005); *Health Insurance Portability and Accountability Act of 1996*, Public Law 104-191, 110 Stat. 1936, codified in sections of 18,26,29, and 42 U.S.C.); "Standards for the Privacy of Individually Identifiable Health Information; Final Rule," 45 CFR Part 160, Federal Register 65, no. 250, December 28, 2000; United States Sentencing Commission, *2006 Federal Sentencing Guidelines Manual* (effective November 1, 2006).
21. Somewhat parallel is the fact that the U.S. Internal Revenue Service does not and cannot check all of the details of every tax return.
22. The government—largely through the U.S. Department of Justice—aggressively pursued such emerging health care problems. The "Plaintiffs Bar" struck as well for their plaintiffs. The related stories of providers' abuse were damaging to many in the health care community.
23. This parallels other regulated industries—utilities, space, defense in the U.S.
24. James Sheehan, "Government Session," (Lecture, Health Care Compliance Association 11th Annual Compliance Institute, Chicago, IL, April, 2007).
25. "By The Numbers," *Modern Healthcare*, June 18, 2007, 32.
26. In reality, sampling costs were not the challenge. Studies have demonstrated that improvements to quality reduce health care costs. Some early decision makers, however, chose not to invest in long-term improvement, because of short-term costs. In this regard, such a problem still exists in health care. However, time has helped more and more health care leaders in all constituencies "get it"—thinking on compliance and quality gets better and has gotten better each year.
27. For example, individuals considering statistical sampling in a small population; e.g., 300 or fewer for a population of claims, should know that it made more sense to re-check an entire universe instead of trying to conduct statistical sampling.
28. The authors believe that the challenges occurred from relatively reasonable judgments made in a "start-up" sense.
29. On stratification and related sample size criticisms for earlier versions, see Jorge Sirgo, Jessica Pollner, and Laura Schweitzer, "Evaluating RAT-STATS Statistical Sampling Software," *Today's Corporate Compliance*, September, 2000, 8-9.
30. There was sponsored research at the Mathematics Department, United States Naval Academy. "RAT-STATS II." Researcher: Professor Thomas J. Sanders. Sponsor: Department of Health and Human Services, OIG, Office of Audit Services. (Reference accessed at *http://www.usna.edu/AcResearch/2003SummaryPDFs/2003_DivMathSci_MathDept.pdf*, on June 25, 2007.).
31. Health Care Financing Administration frequently called HCFA, was the predecessor to the Centers for Medicare and Medicaid Services.
32. *Blue Cross & Blue Shield of New Jersey, Inc. v. Philip Morris, Inc.*, 113 F. Supp.2d 345 (E.D. N.Y. 2000). And importantly, *Chaves County Home Health Services Inc., v. Sullivan* 289 U.S. App. E.C. 276931 F. 2d 914, D.C. Cir. 1991.
33. *Crider, etc. v. Keystone Health, etc.* U.S. District Court, E. Pennsylvania, 2001-CV-0561, September, 2003.

34. This list on court guidance is summarized from Lester J. Perling, "Challenging the Use of Statistical Sampling in Overpayment and False Claim Cases," (Lecture Notes, *American Health Lawyers Association*, 2001).

35. There is an old health care "answer" to a proponent of having to prove each and every dollar of all claim items. It goes like this, "Tell that proponent of all dollars, all claims, that the next time he or she goes for a blood test to request that the nurse take all of his or her blood and not just a sample."

36. See the later section on the evolution of sample size as well as later reference on rule making involving formulas.

37. Indeed, the government challenges of health care payments together with later provider and commercial payor payments led to many, many settlements which used statistical sampling in presenting issues among the parties, in the fact-finding, in the negotiation, etc.

38. In areas, such as the definition of basic billing errors, related billing issues or quality problems, the authors are aware that technical concept requirements involving "intent" are important by law, regulation and court interpretation.

39. Similar misunderstandings occurred in other industries' use of statistical sampling, but a good time ago.

40. *Daytona Beach General Hospital, Inc. v. Weinberger*, 435 F. Supp. 891 (M.D. Fla. 1977).

41. And there arise circumstances in health care where larger samples are relevant.

42. *Medicare Carrier's Manual. Part B, 3: Claims process*, (Washington, D.C.: U.S. Dept. of Health and Human Services, Health Care Financing Administration: Supt. of Docs., U.S. G.P.O., distributor, 1986).

43. There may have been some earlier, similar publications not consistently well known.

44. Another method for determining sample sizes is based on the likelihood of observing zero errors among all the elements in a sample under the assumption of an error rate to be tested. The sample size is increased until it is sufficiently unlikely that the error rate to be tested would result in zero errors being observed. This method is used in "discovery sampling" by some auditors in other industries where there is guidance on a directly asserted tolerable error rate, but it has not been adopted in health care guidance where the tolerable error rate tends to be, and properly should remain, unspecified.

45. A similar approach may apply to Medicare carriers or fiscal intermediaries, who are themselves audited by the government.

46. Of course, both have to start somewhere.

47. U.S. Department of Health and Human Service, *Internet-Only Manuals (IOMs)*. Section 3.10.4.3—Determining Sample Size. (Reference accessed at *http://www.cms.hhs.gov/Manuals/IOM/list.asp*, on June 25, 2007.)

48. Health care regulators have a good record to date on such flexibility as a practice, especially in administrative determinations; that view should be "codified;" e.g., taking forward the "guidelines" concept to regulations or possibly to law and retaining the guideline "blessing."

49. An estimator is the mathematical formula that transforms a sample's many results for all items drawn and reviewed into a statistical result. For example, the average estimator is the sum of the overpayments divided by the number of claims that were drawn and reviewed in a sample; the ratio estimator, which can be an alternative estimator, is the sum of the overpayments divided by the sum of all the payments for claims that were drawn and reviewed in a sample. Both an average estimator and a ratio estimator may be valid to use for some samples.

50. See the Medicare Internet-Only Manuals.

    It should be noted, however, that the variance of a ratio is computed using a different formula than the variance of a standard variable. The formula for the variance of a ratio must account for the variance in the numerator, the variance in the denominator and the covariance of the two.

51. Indeed, where false claims are asserted and where statistical sampling is used, the government or other plaintiff must prove the existence of false claims legal requirements claim issue by claim issue to establish violations. Only if the existence of false claims is proved directly, can a sample be used to estimate damages.

52. Such a finding should have been expected.

53. Important developments, beyond sampling applications, included organizational improvements, the emergence of a class of chief compliance officers, better controls, more training, etc.

54. Seth W. Glickman, Fang-Shu Ou, Elizabeth R. DeLong, Matthew T. Roe, Barbara L. Lytle, Jyotsna Mulgund, John S. Rumsfeld, W. Brian Gibler, E. Magnus Ohman, Kevin A. Schulman, and Eric D. Peterson, "Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarction," *Journal of the American Medical Association*, 297 (2007): 2373-2380.

55. It is commonly held that, for whatever reason, governments tend to consistently overreact to embarrassing developments and that overregulation occurs, often creating inefficiency.

56. For a summary of five years (1997 – 2001) of joint reports on detecting, prevention and prosecution by Health and Human Services and the Department of Justice, see "Health Care Fraud and Abuse Control Program for Fiscal Years 2000 and 2001," GAO-02-731, June 2002.

57. Including a handful of relevant health care executives, consultants and the Society of Corporate Compliance and Ethics, which latter group was not in existence until recent years, but which has effectively focused on and developed helpful guidance as well as various recommendations.
    One major contribution of these "external" training efforts was "cover" to help "bridge" until organizational internal training got better.

58. For the uninitiated readers, there is a learning curve involved in the effective use of statistical sampling. The cost is worth the investment. The authors believe that, even without government improvements or without a constructive regulatory posture, providers and commercial payors would have increased their statistical sampling use and would have brought much progress. Nonetheless, the government aided that process.

59. An early, major compliance program was developed as part of the Columbia HCA compliance in the mid-to-late 1990s. More recently, also see a provider-payor decision where a contention that a statistically insignificant sampling did not state a claim for fraud was denied, because the language of the parties agreement called for a "survey" of HMO member satisfaction. Crider, etc. Keystone Health, etc. U.S. District Court, E. Pennsylvania, 2001-CV-0561, September, 2003.
    Larger providers and payors would be wise to expand their present use of statistical sampling; a single hospital provider at a minimum should be using random number selections for some of its internal audit tests. Of course, some simple judgmental sampling will help as well. And judgment samples for all institutions will continue to be valuable when analyzing many identified problems.

60. Florida provides an example of a decision about a state's obligation in health care statistical sampling to follow that state's own laws on rule making involving use of a formula. The Florida statistical sampling formula at issue had only been used for some 18 years without formal rule making. Compliance officials need to use care and to consider going to competent counsel, where appropriate, on both sides of rule making regulations as well as reimbursement.

61. The improvement was gradual, but it was from reasonable decisions also. The health care compliance improvement had gained its own momentum, including statistical sampling, and had its own "roll," started well before Sarbannes Oxley.

62. Some references are:
    - United States Government Accountability Office, Report to the Committee on Finance, U.S. Senate, *Medicaid Financing: Federal oversight initiative…*, March 2007, GAO-07-214;
    - United States Government Accountability Office, Report to the Chairman, Committee on Finance, U.S. Senate, *Medicaid Financial Management: Steps taken to improve federal oversight…* June 2006, GAO-06-705.; and
    - United States General Accounting Office, Report to the Subcommittee on Government Efficiency, Financial Management and Intergovermental Relations, Committee on Government Reform, House of Representatives, *Medicaid Financial Management: Better oversight of state claims…* Feb. 2002, GAO-02-300.

63. These partnerships are somewhat logical, but they have unpleasant risks of possible compliance abuse or omission.

64. Dennis Smith, "Bolstering the Safety Net: Eliminating Medicaid Fraud," (testimony before the Senate Subcommittee on Federal Financial Management, Government Information and International Security of the Committee on Homeland Security and Governmental Affairs, March 28, 2006; reference accessed at: *http://www.cms.hhs.gov/apps/media/press/testimony.asp?Counter+1822*, on July 13, 2007.)

65. For more information on Medicaid Financing issues, the reader is directed to the March 2007 GAO Report, *Federal Oversight Initiative Is Consistent with Medicaid Payment Principles but Needs Greater Transparency*. This report, among others, comments on how CMS has responded to what GAO refers to as "inappropriate financing arrangements to collect federal matching funds when payments were not retained by or returned to providers in state Medicaid." This issue is believed to have arisen as a result of prior audit tests by GAO and CMS using judgmental sampling. GAO had recommended that CMS provide guidance to clarify allowable financing arrangements and to explain its determinations in writing; i.e., implement clarifying policies transparently. CMS has since (May 25, 2007) issued new rules related to the financing mechanisms used by certain states to be effective July 30, 2007. However, there is a one-year moratorium on implementation beginning effectively May 25, 2007.

    In Fall of 2006, the state of Minnesota had appealed an earlier related federal administrative decision for Medicare and Medicaid Services, involving what are called inter-governmental transfers (the "financing arrangements at issue") and federal rulemaking, to the Eight Circuit Court of Appeal.#06-3263. Argument occurred in May 2007.

    These financing mechanisms have distorted the calculations of Medicaid expenditures in many cases in certain states. For one thing, use of such transfers (as happened), have created unsustainable levels of expenditures, putting providers, beneficiaries and states at risk. For another thing, the inappropriate financing arrangements had the appearance of "gaming" the health care system for colorable advantage and to the risk of considerable credibility impairment. The relevant governmental organizations should not have so tried or approved the practices at issue in the first place.

66. Following a whistleblower trial in Fall 2006, there were motions before the U.S. District Court by Amerigroup Illinois and the United States as well as the state of Illinois, the relator and others involving a) Medicaid, b) judgment, c) a new trial, and d) remittor for the defendant and penalties for the plaintiff. Statistical sampling was involved. Earlier jury damages, after trebling, were some $140 million. Court-imposed penalties were about $190 million. Case no. 02CG074; decided March 2007. Appeals are pending. The matter illustrates states' more current regulatory focus on Medicaid.

67. The so called Medi/Medi projects, implemented in several states, involve comparing data from both Medicare and Medicaid programs to try to reveal fraudulent patterns previously not easily visible to either program, independent of the other. Such data exchange was prohibited prior to 2001, when this program was first implemented in California. See Dennis Smith, "Bolstering the Safety Net: Eliminating Medicaid Fraud," (testimony before the Senate Subcommittee on Federal Financial Management, Government Information and International Security of the Committee on Homeland Security and Governmental Affairs, March 28, 2006; reference accessed at *http://www.cms.hhs.gov/apps/media/press/testimony.asp?Counter+1822*, on July 13, 2007.)

68. However, the same is not as common in allegations of false claims where a defendant is argued to be or is found to have knowingly violated the laws and regulations, has been similarly reckless, etc. Practically speaking, there can be some "mechanical" offsets in statistical sampling on one set of dispute facts or another. See *CMS Recovery Audit Contract Initiative Guidelines*, 2005.

69. This is an area where health care government leadership wasted less time seeking no-offset claim refunds.

70. Qualified personnel include those with strengths in statistical sampling who are involved throughout the compliance review project.

71. Neither understated nor overstated is fair.

72. Including relevant sampling work not done; (e.g., "look-ups" for missing proof).

73. With regard for billing compliance testing, for example, missing documentation may be a "potential error;" if documentation becomes available, the "potential error" may have been misstated.

74. For senior compliance personnel, see the case decisions as well as writings on *Trigon Insurance Co. v. USA*, Civil No. 3:00cv365 (E.D. V.A. 2002)., see *Laura Zubalake v. UBS Warburg LLC*, 02 Civ. 1243 (SAS) (S.D. N.Y., 2005)., and *Morgan Stanley & Co. Inc., v. Coleman Holdings Inc.*, No. 4D05-2606 (Court of Appeal of Florida, 4th District, 2007). All are required knowledge.

    On electronic health records, see Kelly Green and Chad Priest, "Emerging Health Care Techology; Ensuring Compliance and Quality in the 21st Century," (Lecture, Health Care Compliance Association 11th Annual Compliance Institute, Chicago, IL, April, 2007), Graph 7.

75. While this could be said to just be more of the regulators "following the money," the authors believe a far more virtuous motivation exists—a desire or need for market credibility through quality. Again, readers should "think Intel and Toyota" in manufacturing and marketing, for example. Or, think the General Accountability Office in government on quality and avoiding systemic problems.

76. "Management is about human beings. Its task is to make people capable of joint performance, to make their strengths effective and their weaknesses irrelevant." Drucker interview with author in 2004 in Elizabeth H. Edersheim, *The Definitive Drucker*, (New York: McGraw-Hill, 2006).

77. It is often said that, to use a hypothetical, a weak or under-performing physician is known to his or her peers—that is other doctors, nurses, technicians, etc. Something similar is also often true with regard to systems. The health care "trade" often knows of system problems. The quality reviewer only has to ask for their information.

78. For example, it would be helpful to a health care provider to use sampling in checking several departments capture and as well as reporting of adverse events which, in time, could lead to compliance improvements.

79. Investment in health care statistical sampling for quality improvement reasons is like manufacturers investment in computer-aided design and computer-controlled manufacturing processes.

80. Again, adverse events reporting needs to be well done. The effectiveness of adverse events reporting can be tested by statistical sampling.

81. Testing, both sampling and analyzing, the facts relating to providing the right care; doing so at the right time; doing so at the appropriate place with the right persons and equipment. Organizations like Intel and Toyota did it. Health care can do it as well.

CHAPTER 3
Retrospective Versus Contemporaneous Reviews

l. 63 FR. 8987 (February 23, 1998).

2. The seven key elements recommended by the OIG of the Department of Health and Human Services effective compliance program are: (i) written code of conduct; (ii) designation of a chief compliance officer and other appropriate bodies such as a corporate compliance committee; (iii) effective education and training programs; (iv) hotlines and similar processes; (v) the use of audits, reviews, and other evaluation techniques to monitor compliance; (vi) enforcement of compliance standards through well published disciplinary standards; and (vii) the investigation and remediation of identified systemic problems and policies addressing the nonemployment or retention of sanctioned individuals.

3. 63 F.R. at 8996.

4. In its model Guidance published for the long-term care industry, the OIG stated that it "believes that an effective program should incorporate thorough monitoring of its implementation and ongoing evaluation process." See *Fed. Reg.* 14289, 14302 (March 16, 2000).

5. 63 F.R. at 8996.

6. It goes beyond the scope of this discussion to detail all aspects of the controversy surrounding this provision of Medicare law. However, certain constitutional law issues have been raised regarding this provision and there has been some debate as to whether it applies only to individuals and not to corporate entities such as hospitals and other health care providers.