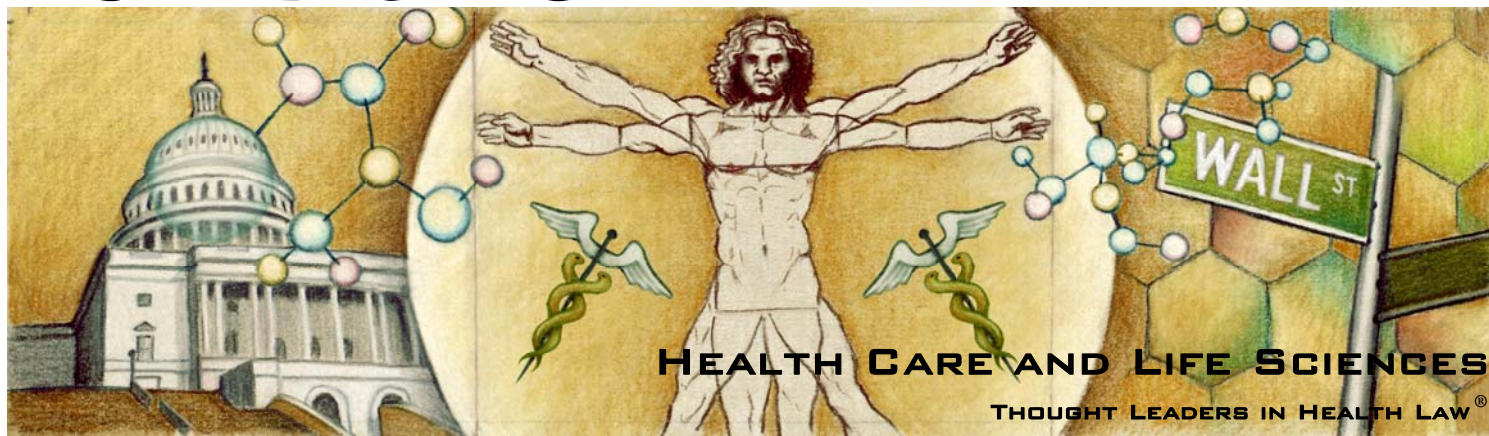


# EPSTEINBECKERGREEN



## Statistical Sampling in Health Care Fraud Litigation

BCBSA National Internal Audit & Anti-Fraud Conference

May 2008

**Steven E. Skwara**  
Epstein, Becker & Green, P.C.  
Washington, DC  
[sskwara@ebglaw.com](mailto:sskwara@ebglaw.com)

# STATISTICAL SAMPLING IN HEALTH CARE FRAUD LITIGATION

May 2008

Steven E. Skwara

This paper is intended to help health care fraud investigators and the attorneys who litigate related cases understand the legal background and basic practical considerations of statistical sampling in health care fraud-related civil litigation. First, this paper discusses the general acceptance of statistical sampling methods by courts in health care fraud cases as well as one recent, anomalous exception. Second, this paper discusses the practical considerations that investigators and counsel should keep in mind when anticipating or participating in health care fraud-related litigation involving statistical sampling of claims.

## **I. Judicial Review of Sampling Evidence in Health Care Fraud Cases.**

Statistical sampling allows one to derive conclusions about a larger population (e.g., insurance claims, people) by measuring a meaningful sample or subset of the individuals within that population. Obvious examples of statistical sampling in everyday life include political and other opinion polls. Litigation-wise, courts have accepted statistical sampling in a variety of circumstances, including political district apportionment cases, pornography prosecutions, drug prosecutions, labor and employment litigation, and others. See Jones & Hagtvedt, *Sample Data as Evidence: Meeting the Requirements of Daubert and the recent amended Federal Rules of Evidence*, 18 Ga. St. U. L. Rev. 721, 723 to 725 (Spring 2002) (collected citations).

To those (like auditors or health care fraud investigators) steeped in the regular use of statistical sampling and extrapolation, the use of such evidence to prove damages in a health care fraud case involving large datasets is obvious. In cases that can include hundreds, or thousands, of potentially fraudulent claims, a claim-by-claim presentation at trial would be unbearably tedious

and costly, if not logistically impossible. See, e.g., Illinois Physicians Union v. Miller, 675 F.2d 151, 158 (7<sup>th</sup> Cir. 1982) (“[I]n view of the enormous logistical problem of Medicaid enforcement, statistical sampling is the only feasible method available.”); Mile High Therapy Centers, Inc. v. Bowen, 735 F.Supp. 984 (D. Colo. 1988) (case-by-case review not administratively feasible). Neither a court nor a jury would have the patience to endure a claim-by-claim presentation in such a case, and counsel would be well-advised not to propose same. Instead, statistical sampling provides an alternative to claim-by-claim proof by providing a means of estimating with reasonable certainty the aggregation of large quantities of data without having to measure each individual data point. Accordingly, most courts accept, indeed encourage, the use of statistical sampling in civil health care fraud cases.

**A. Statistical Sampling to Prove Actual Loss or Single Damages**

In health care fraud and abuse cases, Federal courts have repeatedly held that the government may use statistical sampling methods when calculating claims overpayments or actual damages. In holding that the U.S. Department of Health and Human Services (“HHS”) enjoyed this authority, the Ninth Circuit Court of Appeals found extrapolation based on statistical sampling to be reasonable where the defendant provider had the opportunity to challenge the statistical sampling methodology. Chaves County Home Health Service, Inc. v. Sullivan, 931 F.2d 914 (D.C. App. 1991). The other federal circuit courts considering the government’s use of statistical sampling and extrapolation in recovery cases have held similarly. See Ratanasen v. California, 11 F.3d 1467, 1471 (9<sup>th</sup> Cir. 1993) (“We now join the other circuits in approving the use of sampling and extrapolation as part of audits in connection with Medicare and other similar programs, provided the aggrieved party has an opportunity to rebut such evidence.”); Illinois Physicians Union, 675 F.2d

at 156: Michigan Dept. of Educ. v. U.S. Dept. of Educ., 875 F.2d 1196 (6<sup>th</sup> Cir. 1989).<sup>1</sup>

Essentially, these cases hold that a statistical sampling-based estimate of overpayment creates a rebuttable presumption that defendant providers can challenge through introduction of countervailing or claim-by-claim evidence if the provider chooses.<sup>2</sup>

The use of statistical sampling in health care claims overpayment cases is not entirely settled in all jurisdictions, however. The concept of statistical sampling befuddled the Rhode Island Supreme Court in Garden City Treatment Center, Inc. v. Coordinate Health Partners, Inc., et al., 852 A.2d 535 (R.I. 2004). In Garden City, plaintiff Garden City Treatment Center, an emergency care facility, filed suit against Blue Cross of Rhode Island (“BCRI”) as a result of various claims audits performed by BCRI. In those audits, BCRI sampled a subset of plaintiff’s claims and found overpayment rates of 29% and higher (i.e., 29% of the claims in one sample were found to have included excess payments). BCRI applied those overpayment rates to the entire universe of claims from which each sample was drawn and extrapolated an overpayment amount. Acknowledging that BCRI enjoyed the contractual right to audit its claims, Garden City claimed that BCRI’s extrapolation nonetheless violated the parties’ contract since the contract did not expressly grant BCRI the ability or right to use statistical sampling or extrapolation.

The Garden City trial court determined that the entire case turned on the definition of “audit” in the contract. After an exhaustive research effort — she consulted Black’s Law Dictionary — the trial judge held that BCRI could not extrapolate because she could not find a

---

<sup>1</sup> Private payers should be able confidently to rely upon these precedents too insofar as those cases find sampling generally to be a valid means of estimating overpayment or actual damages.

<sup>2</sup> The opportunity to present claim-by-claim evidence thwarts a provider’s argument that statistical sampling violates the provider’s constitutional due process rights in government overpayment or fraud cases.

dictionary definition of “audit” that referred to extrapolation or statistical methods. BCRI appealed. The RI Supreme Court confirmed the trial court’s decision, holding that BCRI’s contractual right to “audit” Garden City allowed only for a record-by-record review. *Id.* at 543 (“a case-by-case or record-by-record examination more closely fits the ordinary understanding of audit.”) Thus, BCRI could not seek to collect the overpayment amounts reached via extrapolation but still enjoyed the ability to “audit,” for example, all 22,518 claims submitted by Garden City to BCRI in 2000.

Garden City will likely not develop a following. Indeed, Justice Flanders concurring opinion in Garden City correctly takes the majority to task. He criticizes the majority’s focus on whether the dictionary definition of “audit” refers to statistical sampling and instead argues that the Garden City’s motion for summary judgment should have hinged on whether BCRI could establish that it used proper statistical sampling methods. Justice Flanders explained that the use of statistical sampling was a question of evidence, not a contractual issue: “absent some valid contractual provision or law preventing them from doing so, and subject to the rules of evidence governing the admission of expert testimony, I would hold that parties may freely use statistical sampling as a method of proving alleged overcharges in situations such as the one presented by the case at bar.” *Id.* at 547.

While most courts generally accept the evidentiary concept of statistical sampling, it has been rejected in specific cases due to methodological infirmities. For example, the court in United States v. Skodnek, 933 F. Supp. 1108 (D. Mass. 1996), rejected the government’s extrapolation of fraudulent claims for purposes of calculating a “loss” figure for sentencing in a criminal case. In Skodnek, the government had proven at trial that defendant Skodnek had submitted \$157,460 in fraudulent claims to Medicare and private payers; based on those proven fraudulent claims, the

government extrapolated a “loss” attributable to Skodnek of \$1,218,454 during the sentencing phase of the proceeding.

The Skodnek court observed that “[i]n order to accept the picture offered by the government with respect to the entire universe of fraudulent billings, I have to accept that Skodnek almost never honestly billed a single patient, never forgot to bill for time he had actually spent with them . . . , and that his illegal activities never ebbed and flowed over a five or six year period.” Skodnek, 933 F. Supp. at 1117. The court found that this “picture” was undermined by evidence in the case. And commenting on the government’s extrapolation sample, the court disparaged it as a “convenience sample, garnered by a unit whose purpose is to investigate fraud.” *Id.* “Clearly, the interviewers were searching out ‘horror’ stories, the stuff of which criminal prosecutions can be made or sentencing increased [and] there were instances in which [patient] reports apparently inconsistent with the overall conclusions were ignored.” *Id.* at 1118. The cherry-picked sample of fraudulent claims in Skodnek was thus rejected as a basis for extrapolation of the fraud loss for sentencing purposes. (The jury had otherwise found Skodnek guilty on many counts of health care fraud against Medicare and private payers.)

Another case rejecting sampling to prove FCA liability is United States v. Medco Physicians Unlimited, 2000 U.S. Dist. LEXIS 5843 (N.D. Ill. 2000). In that case, the court denied the Government’s motion for summary judgment in part because the Government sought to rely on expert analysis of sixteen patient files and to extrapolate from that analysis to all claims submitted to Medicare by the defendant. The experts determined that the sixteen patient files revealed that the defendant billed Medicare for non-medically necessary services. The Government then argued that all claims submitted by the defendant in the year in question should be deemed false. The court summarily rejected the proposed extrapolation, stating that there was

“no case law or other authority to support such a request.” Id. at 23. Moreover, held the court, the Government had not even shown that Medco billed Medicare or was reimbursed for those sixteen patients.

The recounting of a Social Security Administration (“SSA”) case similarly highlights a court’s rejecting statistical sampling in light of specific methodology problems. See Holland & Hart, *Roaming the Random Range*, March 2003 Health Care Law Bulletin (describing In re American Health Services, HICN 103-01-0077 A (2000)). In this overpayment case, a representative of the Office of Inspector General (“OIG”) testified that his office did not review a 400-claim sample, the minimum necessary statistically, due to “lack of audit resources and availability of staff.” The OIG also failed properly to document and preserve its sampling methodology such that it could be replicated. Therefore, the SSA’s Office of Hearings and Appeals overturned a \$1,248,747 overpayment determination. Id.

Also unsuccessful was a plaintiff relator’s attempt to employ a novel application of statistical sampling in United States ex rel. Kusner v. Osteopathic Medical Center, 1997 U.S. Dist. Lexis 16855 (E.D. Pa. Oct. 23, 1997). In Kusner, a civil False Claims Act case, the relator relied upon Ratanasen and other federal cases endorsing statistical sampling when attempting to ameliorate the fact that neither he nor defendants had relevant billing records for much of the relevant time period. Though the relator (Kusner) had no knowledge of whether underlying individual patient billing records existed, and defendants submitted evidence that individual patient records did not exist and that the billing records that did exist contained inconsistencies and data integrity issues, Kusner argued that an expert statistician could make a projection of damages. Id. at 21. The court held otherwise, noting that Ratanasen and like cases were “distinguishable, however, because they all involve, in some form or another, *actual* billing records from which extrapolations were made.” Id. (emphasis in original).

In other words, Kusner proposed to sample non-existent data. See also United States ex rel. Aflatooni v. Kitsap Phys. Svc., 314 F.3d 995 (9<sup>th</sup> Cir. 2002) (cannot imply falsity of claims in FCA case from circumstantial evidence where no actual claims in evidence).

**B. Statistical Sampling to Prove Liability in False Claims Act Cases.**

While use of statistical sampling to prove actual loss, overpayment, or single damages is generally accepted, its use in cases under the False Claims Act (“FCA”), 31 U.S.C. §§ 3729 to 3733, for purposes of calculating multiple or punitive damages seems to be the subject of greater judicial resistance. At least one commentator suggests that this resistance is justified:

The government is increasingly attempting to rely on statistical analysis to prove both liability and damages in FCA cases. These efforts raise obvious concerns about the reliability and permissibility of using statistical extrapolations to prove the fundamental elements of an FCA claim. The excuse tendered by DOJ is that these cases are so large and complex that statistical evidence is necessary for the cases to go forward. The problem with that analysis, of course, is that DOJ always insists, for a variety of technical reasons, that each false claim is a separate violation of the Act. There is little doubt that if DOJ brought an FCA case based on five false claims, it could not prove one of them and then argue the other four are just like the first one. If that is true, why should DOJ be excused from proving each claim is false when it brings an FCA case based on 5,000 or 50,000 false claims?

Boese, *Civil False Claims and Qui Tam Actions*, § 2.03(C)(4). On the other hand, nothing precludes a FCA defendant from introducing evidence that a claim or claims subject to extrapolation were not false.

Along these lines, in an FCA case tried in 1993, the district court held a psychiatrist liable for excessive billings but declined to extrapolate from a random sample of over half of the defendants’ claims for purposes of penalties. In a footnote to its unpublished opinion, the court explained:

While I recognize the validity of the mathematical and statistical projections based on a review of a smaller number of claims I have declined to extrapolate in the manner urged by the government. My declination is based on the existence at trial of discrete claims which were analyzed and discussed and subjected to cross

examination. I was able therefore to review each claim in reaching my conclusions. While I am mindful of the government's efforts to shorten the trial and present its evidence efficiently and clearly, I am reluctant to accept a statistical sampling as the basis for doubling the alleged overpayment without the same scrutiny and support.

United States v. Friedman, Slip Op. at pp. 6-7, fn. 1, C.A. No. 86-0610-MA (D. Mass.). Thus, though there seems to be no basis for the distinction, the fact that the actual overpayment could be doubled gave the court pause.

In United States v. Cabrera-Diaz, 106 F.Supp. 2d 234 (D.P.R. 2000), a FCA case, the court accepted the government's sampling and extrapolation in establishing liability and awarding damages, but would not extrapolate for purposes of imposing punitive damages. The Medicare carrier to which the defendant submitted his claims identified a statistically valid random sample of 461 claims from a two-year period and found that all but six were false. "The results were then extrapolated from the sample to calculate the amount of the Medicare overpayment." *Id.* at 240. The court adopted the sampling results to extrapolate the damages. The court, however, balked at using the extrapolation to calculate per-claim penalties of \$5,000 to \$10,000 because the resulting total penalty — an additional \$2 to \$4 Million — would be "excessive." While the Cabrera-Diaz court's finding of "excessiveness" would seemingly have resulted regardless of whether the underlying liability resulted from statistical sampling or direct, claim-by-claim proof, the element of reluctance in the court's opinion seems to stem in part from the relative ease with which treble damages of approximately \$1.3 Million resulted in a \$2 to \$4 Million penalty.

In sum, as a legal matter, most courts accept statistical sampling evidence for purposes of estimating single damages or actual loss in health care fraud cases. There has, however, been some reluctance to use extrapolation for establishing FCA penalties. Of course, as set forth in the next

section, judicial acceptance of the concept alone is not enough; the proponent of statistical sampling evidence still must offer competent expert and factual support for that evidence.

## **II. Practical Challenges to Sampling in Health Care Fraud Cases.**

The ability successfully to assert or challenge statistical sampling evidence depends on the adequacy of the underlying processes and methodology. As statistical sampling evidence has become more commonplace, the attacks have become more sophisticated.<sup>3</sup> Proponents, in turn, have been and will continue to be pressed on the adequacy of their sampling processes. Ultimately, though, scrupulous attention to documentation and method by a payer and its expert will often render statistical sampling evidence competent, and therefore admissible.

### **A. Expert Testimony Typically Required.**

As a threshold matter, the use of sampling evidence in a legal proceeding will require an expert witness, typically a statistician. Admission of expert testimony requires among other things that a court be satisfied with the sufficiency of the facts relied upon by the expert; here, the “facts” in question are claims data. Because an expert must understand fully the factual basis of his or her expert testimony, in a claims-based health care fraud case, the expert witness must understand

---

<sup>3</sup> They have not always been sophisticated. In United States v. Krizek, defendant providers agreed that the liability case against them could be tried based on the records of seven patients and 200 associated claims. Ill. F.3d 934 (D.C. App. 1997). After the government prevailed in showing liability for the miscoded HCFA 1500 claims associated with the seven patients, the trial court extrapolated the results to the over 8,000 claims submitted by defendants. Despite having agreed to the sampling, defendants objected to the extrapolation on appeal. And lost. In hindsight, defense counsel perhaps should have considered whether to have at least tried to retain the ability to challenge the statistical extrapolation methodology even if liability were established for the seven members. Instead, defendants agreed to the trial court’s approach that “[t]his case will go to trial on this issue of liability using these seven patients as a representative sample. A determination of liability on the issue of improper coding would be equally applicable to all other claims.” Krizek, 111 F.3d at 941.

fully the source and meaning of the payer data and be aware of any inconsistencies in it. Any manipulation or preparation of data must be well-documented (or “transparent” in current jargon).

## **B. Specific Challenges to Statistical Sampling.**

Once a payer has qualified its expert, the battle of the legitimacy of the sampling evidence will typically focus on the following issues: (1) the “randomness” and reproducibility of the sample; (2) adequacy of sample size; (3) the variability within the sample; and (4) the type of sample.

### **1. Reproducibility.**

The “scientific” validity of a statistical sample and extrapolation demands reproducibility. Where a payer’s statistical methods cannot be duplicated or reproduced, a defendant provider can make a credible argument that that method was non-scientific and thus inadmissible.

Of critical importance in this regard is the expert’s documentation of his or her analysis. Among other things, the expert selecting a universe of health care claims data to be sampled must record the time period to which the claims pertain, the provider(s) to whom it pertains, and the source of the data. Decisions about changes to the universe of data should also be documented contemporaneously. Such decisions could include limiting the time frame for review, eliminating zero-paid claims, eliminating claims data based on patient age or sex, etc. Full documentation of the process ensures that another expert statistician could reproduce the analysis.

The method by which a payer or its expert randomized the sample must be reproducible by others. The expert witness must understand the randomness algorithm. Software applications like RAT-STATS and STARS generate random samples using an algorithm, and typically generate a “seed number.” To replicate a random number sequence, the original seed number must be known. Again, since reproducibility is key to demonstrating that the process is scientific, a record

of the seed number and a full understanding of the process by which a random sample was selected should be viewed as mandatory predicates to admissibility. Failure to document the “seed number” could be fatal to the chances of payer intending to introduce statistical sampling evidence.

## **2. Adequacy of Sample Size.**

Providers frequently challenge the adequacy of sample size, claiming that the payer’s sample size was too small to support the payer’s inferences from the sample. The adequacy of the sample size is essentially a mathematical or statistical function, however, and “[t]here is no case law that states how large a percentage of the entire universe must be sample.” Michigan Dept. of Educ., 875 F.2d at 1206. Generally, of course, the larger the sample, the more precise the extrapolation will be. Nonetheless, a competent statistical expert can use statistical error concepts such as confidence intervals and precision levels to explain within what degree of accuracy his or her extrapolation likely meshes with reality so that even a relatively small sample of claims can be the basis of extrapolation.<sup>4</sup>

## **3. Variability.**

The variability within a sample, sometimes stated as the coefficient of variation, can also give rise to legal challenge. Where the material data within a sample vary widely, the reliability and

---

<sup>4</sup> The RAT-STATS “Companion Manual” downloadable at <http://oig.hhs.gov/organization/OAS/ratstat.html> describes precision and confidence levels in this context. (It also provides a lot of squiggly lines and math formulas for the more astute reader.) While the calculation and inner meaning of confidence intervals are beyond the scope of this paper, a couple of general pointers are offered. The higher the confidence interval, the more likely it is that the extrapolated result(s) from a sample resemble the outcome had the entire underlying data population been analyzed. Typically, confidence intervals of 80%, 90%, 95%, and 99% are generally accepted benchmarks for statistical sampling. The confidence interval will be function of sample size and variability. A confidence interval of 90, for example, indicates that there is a 90% probability that the “real” answer falls within the range of extrapolated results. Extrapolated results are typically expressed in a range from a lower limit to an upper limit with a “point estimate” falling somewhere in the middle. (The difference between the point estimate and the upper and lower limits is synonymous with the term “margin of error.”)

precision of the sample suffer. For example, a random sample drawn from the entire universe of a surgeon's health care insurance claims might include low level office visits that paid tens of dollars and involved surgical claims that paid thousands of dollars. Or, for instance, claims for Medicare beneficiaries might vary materially from claims of privately-insured patients due to underlying demographic differences. These payment or demographic differences would create an element of variability within the sample. This variability can be muted, however, by increasing the size of the sample or by thoughtful selection of the universe of data to be sampled. For example, eliminating the high-priced surgeries in the above example might make sense if the case is focused on the provider's use of office visit codes and would create a better starting universe of data that reduces variability within the subsequent sample. Stratification, discussed below, also minimizes the problems associated with high variability.

#### **4. Simple Versus Stratified Samples.**

Providers frequently question the selection of a simple random sample over a stratified random sample. See, e.g., Ratanasen, 11 F.3d at 1471 (rejecting defendant's argument that stratified sample required). A simple random sample may be used where the underlying dataset is relatively homogenous (i.e., where factors among data points do not vary widely.) Where the dataset is more heterogeneous, however, a stratified sample may be required. A stratified sample comprises sub-samples keyed to the factors that make the dataset relatively heterogeneous.

A successful challenge to a state's use of a simple, or non-stratified, random sample was mounted by a provider in HCA Health Services of Kansas, Inc. v. State of Kansas, 21 Kan. App. 2d 141 (1994). In this case, the state alleged that defendant, a medical clinic, overbilled Kansas' Medicaid program by misclassifying certain emergency and non-emergency room claims and sought an overpayment amount estimated using statistical sampling. The defendant medical clinic successfully argued that the state's failure to determine in the first place whether the disputed

Medicaid claims were homogenous or heterogeneous rendered the state's non-stratified sample infirm. According to the state's own statistical sampling policies, it should have made such a determination based on factors including differing procedure codes, places of service, and servicing providers. The defendant medical clinic's expert witness testified that in fact the identities of the servicing providers differed, the provider types differed, and the place of treatment differed, all of which indicated that the state should have used a stratified sample. The court agreed, holding that the "[state's] failure to determine if the population was homogeneous or heterogeneous before deciding which sampling procedure to use was unreasonable, arbitrary, and capricious." Id. at 157.

In Skodnek, for another example, the court implicitly took issue with the absence of stratification within the sample when wondering whether the extrapolation assumed that Skodnek's "illegal activities never ebbed and flowed over a five or six year period." Id. at 27. Had the government had available to it a truly random, stratified (based on calendar years for instance) sample that included all data both favorable and unfavorable to the government, it may have been better able to persuade the court of Skodnek's sentencing "loss."

In health care fraud cases, the claims data to be sampled may tend to be relatively homogenous where one claim resembles the next. For example, a dataset consisting of an individual provider's office visit CPT codes will tend to be homogenous. In contrast, a provider's claims dataset that includes claims for both small dollar office visits and big ticket surgery claims might require stratification when choosing an appropriate claims sample. Likewise, a dataset that includes multiple providers and multiple sites of service would even more likely be heterogeneous.

## **Conclusion**

In conclusion, courts recognize the general validity of statistical sampling. As long as health care fraud investigators, their attorneys, and their expert witnesses scrupulously document their sampling processes in order to ensure reproducibility and otherwise thoughtfully apply statistical sampling methodologies, courts (outside Rhode Island anyway) will be inclined to accept sampling evidence in civil health care fraud-related litigation.