

THE BROKEN PROMISE OF OBRA '87: THE FAILURE TO VALIDATE THE SURVEY PROTOCOL

MALCOLM J. HARKINS III*

I. INTRODUCTION

Nursing facilities are among the most heavily regulated entities in the American economy. Unfortunately, nursing facilities also are among the most inconsistently and ineffectively regulated entities.

It is not a secret that government inspections of nursing facilities do not result in valid, accurate, and consistent assessments of the quality of care. To the contrary, decades of empirical studies and anecdotal evidence make the point largely without contradiction.

Nursing home regulation is the modern-day analog of the fable in which no one dares to mention that the emperor has no clothes. Politicians, regulators, researchers, advocates for both nursing home residents and industry interests, and courts—virtually anyone with an interest in assessment of nursing home care—treat the results of the regulatory system as trustworthy—at least in so far as such results conform to the speaker’s professional or political interests.

Numerous studies dating back to the 1990s collectively make four points about the unreliability of nursing home inspection results. First, the inspection process does not accurately determine whether a nursing facility is in compliance with the substantive care standards established by federal regulations. Second, the inspection process fails to appropriately and accurately assess the seriousness of violations found. Third, inspection results

* Malcolm J. Harkins is a former partner of the Health Care Department of Proskauer Rose LLP. He currently is a Professor of Practice in the Center for Health Law Studies of the St. Louis University School of Law as well as an adjunct in the School of Public Health and Health Services, Department of Health Services Management and Leadership of The George Washington University.

This article is published with gratitude to Robert Burke, PhD for his valuable insights, comments and suggestions. With thanks also to Taylor Jensen, whose exemplary focus, commitment, patient persistence and work ethic were appreciated more than she’ll ever know. Finally, with very special thanks to Cheryl Cooper, for without her dedication and skill this article could not have been produced.

Any errors in this article are, of course, the responsibility of the author.

are inconsistent. Fourth, there is no evidence that the inaccuracies and variations in survey results are related to differences in quality.

To be sure, many investigators, especially in recent years, have asserted that the regulatory system fails to identify instances of deficient nursing facility care. Indeed, the problems of under-citation, as well as under-classification of the seriousness of deficient practices, have been studied extensively. In fact, the problems of under-citation and under-classification have been the focus of studies of the inspection protocol in recent years. Although recognizing that the survey process is systemically flawed, there is an implicit assumption in such studies that the systemic flaws cause only over-citation and under-citation. There is no empirical evidence supporting that assumption, however.

Any viable effort to reform the nursing home regulatory system must start by acknowledging that systemic flaws in the survey inspection methodology and protocols result in both false-negative and false-positive findings. Such an approach is mandated by law as well as common sense.

The Nursing Home Reform Act mandated that the assessment protocol be “developed, tested, and validated by not later than January 1, 1990,” the date on which the federal government was required to implement the new assessment protocols and procedures.¹ Initially, there was no doubt about the meaning of “tested and validated.”

The Department of Health and Human Services’ Centers for Medicare and Medicaid Services (CMS) acknowledged that Omnibus Budget Reconciliation Act of 1987 (OBRA ’87) required that survey results must be both accurate and consistent.² According to CMS, “[a]ccuracy means not only that the [survey] team has appropriately cited all deficiencies that existed, but also that it has not cited a deficiency when no violation of a requirement has occurred.”³ “Consistent,” according to CMS, means that survey “results are consistent across surveyors, and . . . that the enforcement actions precipitated by the survey results are consistently applied.”⁴ “Inadequate survey performance” is defined by regulation as “a pattern of failure to . . . identify deficiencies . . . [or to] [c]ite only valid deficiencies.”⁵

1. 42 U.S.C. §1395i-3(g)(2)(C) (2012). The prior citation identifies the relevant provision of the Medicare Act. The Medicaid Act contains requirements governing the survey and certification process that are essentially identical to those in the Medicare Act. The majority of those provisions are found at 42 U.S.C. 1396r. Because the language of the Medicare and Medicaid provisions is virtually identical, in most instances, this article provides the citation to the Medicare Act provisions only.

2. Medicare and Medicaid Programs; Survey, Certification, and Enforcement of Skilled Nursing Facilities and Nursing Facilities, 59 Fed. Reg. 56,116, 56,140 (Nov. 10, 1994).

3. *Id.* at 56,146-147.

4. *Id.* at 56,141.

5. 42 C.F.R. § 488.318(a)(1) (2014).

OBRA '87's requirements that the survey methodology produce valid, accurate, and consistent results was neither *sui generis*, nor was it a casually considered suggestion. To the contrary, the requirement wrote into law one of the three key recommendations of the landmark study, "Improving the Quality of Care In Nursing Homes," published by the National Academy of Sciences Institute of Medicine (IOM) in 1986.⁶

The IOM study was ordered by Congress and commissioned by the Department of Health and Human Services (HHS). The IOM study was undertaken after numerous efforts to reform the nursing home regulatory system had failed, despite widespread consensus that the existing regulatory system neither measured nor promoted quality of care.⁷ IOM was viewed as a "fair broker" that possessed the expertise and credibility to reconcile the competing interests and recommend a fair and balanced regulatory system.⁸

IOM made specific and pervasive recommendations for legislative changes to the nursing home regulatory process. IOM compared the three components of the regulatory process to the legs of a three-legged stool: "All are equally important."⁹ The three legs of the regulatory stool are: (1) the substantive care standards; (2) the inspection methodology and procedures (or "survey") used to assess compliance with such standards; and, (3) the means of enforcing compliance.¹⁰

IOM emphasized that the regulatory process could not be effective unless the second leg—the methodologies and procedures utilized to assess compliance with quality standards—was a reliable, valid, consistent, and accurate means of assessing quality. The Report specified that the survey protocol should be validated by "the findings of careful empirical studies," and that tests of the survey protocol should be "carefully analyzed for interrater reliability in use of the instrument" because "[s]urvey findings must be valid and reliable as well as consistent—they should be capable of determining the extent to which a facility is in compliance with [Medicare and Medicaid participation requirements]."¹¹ The IOM Report similarly insisted that "[i]t is essential to incorporate statistically defensible sampling procedures to achieve

6. All three of IOM's recommendations were written into law by OBRA'87. The other two recommendations addressed implementation of resident-focused substantive care standards and improving the effectiveness of enforcement actions taken in response to findings of deficient care. *See, e.g.*, COMM. ON NURSING HOME REGULATION, INST. OF MED., IMPROVING THE QUALITY OF CARE IN NURSING HOMES 70, 71 (1986) [hereinafter IOM REPORT].

7. *See, e.g., id.* at 1-2, 12-16, 238-53.

8. Thomas G. Morford, *Nursing Home Regulation: History and Expectations*, 1988 (SUPPL.) HEALTH CARE FINANCING REV. 129, 130 (1988) [hereinafter Morford].

9. IOM REPORT, *supra* note 6, at 69.

10. *Id.* at 69.

11. *Id.* at 76, 128, 131.

valid, consistent, and reliable findings.”¹² Further, according to IOM, failure to base survey findings on review of a “statistically, valid case-mix-stratified sample of residents” means that a nursing facility with very few residents at high risk of problems and complications will incorrectly appear to have better outcomes than a facility with a more debilitated population.¹³ OBRA ’87 adopted wholesale IOM’s recommendations.¹⁴

This article, beginning in the 1970s, reviews decades of studies of the validity and accuracy of nursing home regulation. Using primarily studies and reports prepared by, or at the behest of, the federal government, this article demonstrates that every study of the system used by the federal government to assess nursing home care has shown nursing home quality assessment is inconsistent, inaccurate, and unreliable. Yet, government regulators, industry officials, resident advocates, courts, and public policy researchers persist in relying on survey data—and encourage others to do so as well—to pursue enforcement actions, to draw conclusions about the relative and absolute quality of nursing facilities, to make conclusions about the state of nursing facility care generally, and to recommend public health policy initiatives regarding nursing facility care delivery.

In the usual case, the ability of regulators and others to utilize data of questionable validity would be subject to judicial review. As a general rule, courts would not permit the use of data or opinion testimony without some showing that the data and the opinions based on them are valid and reliable.¹⁵

12. *Id.* at 131.

13. *Id.* at 114, 115-17.

14. 42 U.S.C. § 1395i-3(g)(2)(C) (2012). The meaning of “valid” is also well understood outside of the context of OBRA ’87. There is a body of case law interpreting “valid” to mean capable of producing accurate and reliable results. The Supreme Court, for example, has held that district courts must, and such courts routinely do, determine the scientific validity of expert testimony as a precondition to admitting such testimony into evidence. *See, e.g., Daubert v. Merrell Dow Pharms.*, 509 U.S. 579, 590 n.9 (1993) (“We note that scientists typically distinguish between “validity” (does the principle support what it purports to show?) and “reliability” (does application of the principle produce consistent results?)”); *see also Pride v. BIC Corp.*, 218 F.3d 566, 577 (6th Cir. 2000) (“Proposed testimony must be supported by appropriate validation - i.e., “good grounds,” based on what is known) (*citing Daubert*, 509 U.S. at 590); *see also Berry v. Crown Equip. Corp.*, 108 F. Supp. 2d 743, 749 (E.D. Mich. 2000) (same); *see also Thompson v. Frostproof*, 103 So. 118, 118 (Fla. 1925) (“Validate, validation and validating are all derivatives of valid, which is defined in the Century Dictionary as meaning to test the validity of, to make valid, confirm, good or sufficient in point of law, efficacious, executed with the proper formalities, incapable of being rightfully overthrown or set aside, sustainable and effective in law as distinguished from that which exists or took place [sic] in fact or appearance but has not the requisites to enable it to be recognized and enforced by law.”). By using the term “validate,” Congress used a term that courts are familiar with and routinely apply to determine whether evidence has sufficient indicia of accuracy and that it is reliable enough to be relied on or considered in determining the rights of litigants.

15. *See supra* note 14.

Nursing home regulation is not the usual case, however. A series of jurisdictional, procedural, economic, and practical barriers insulate the survey protocol from meaningful challenge and judicial oversight; OBRA '87's commands to test and validate the survey protocol and to measure and reduce inconsistency are thereby rendered largely meaningless.¹⁶

16. First, despite OBRA'87 requirement that routine annual (known as "standard") surveys must be conducted using the specified protocol, 42 U.S.C. § 1395i-3(g)(2)(C); *see also* 42 U.S.C. § 1395i-3(g)(3)(A) (requiring federal surveyors conducting validation surveys to assess the performance of state surveyors to utilize the same protocol), the surveyors' failure to utilize the prescribed survey protocol is irrelevant when challenging survey decisions. *See* 42 C.F.R. § 488.305(b) (2014) ("The State survey agency's failure to follow the procedures set forth in this section will not invalidate otherwise legitimate determinations that a facility's deficiencies exist."); *see also* 42 C.F.R. § 488.318(a)-(b) (2014) ("Inadequate survey performance," including the failure," [u]se Federal standards, protocols and the forms to, methods and procedures specified by CMS. . ." and the failure to "[c]onduct surveys in accordance with the requirements of this subpart" does not "invalidate adequately documented deficiencies.").

Federal court review is also not realistic. The Medicare Act denies federal courts jurisdiction to review challenges to the survey and enforcement system until after complete exhaustion of administrative remedies. Exhaustion of such remedies is costly and often takes a year or more to complete, effectively rendering any remedy meaningless. *See* 42 U.S.C. § 405(g) (2012), made applicable to the Medicare Act by 42 U.S.C. § 1395ii (2012) and to appeals by nursing facilities by 42 U.S.C. 1395cc(h)(1) (2012). *See also* *Shalala v. Ill. Council on Long Term Care, Inc.*, 529 U.S. 1, 12-13 (2000); *id.* at 48 (Thomas, J., dissenting); *Peak Med. Okla. No. 5 d/b/a Woodland View Care & Rehab. Ctr. v. Sebelius*, 2010 WL 4809319, at *4 (N.D. Okla. Nov. 5, 2010).

Third, a nursing facility may appeal an adverse finding only when a "remedy," i.e., a sanction, is imposed by CMS. If the provider corrects the alleged problem in order to avoid sanction and assure uninterrupted Medicare and Medicaid payments, the right to appeal is lost. Thus, a provider wishing to appeal must maintain its innocence and accept substantial economic risk to mount a challenge the protocol – a challenge that a court will not hear and adjudicate for years. *See e.g.*, *Bryn Mawr Care Inc., v. Sebelius*, 749 F.3d 592, 603 (7th Cir. 2014).

Finally, the procedural rules and rules of decision utilized in administrative appeals effectively "stack the deck" in CMS's favor. *See generally* Joseph L. Bianculli and Kelly A. Priegnitz, Presentation at American Health Lawyers Association, Long Term Care and the Law Conference: 2012 Survey and Enforcement Case Law Update (Feb. 2012), *available at* https://www.healthlawyers.org/Events/Programs/Materials/Documents/LTC12/papers/N_bianculli_priegnitz.pdf. Such rules make the chances of a nursing facility prevailing minimal and the investment in doing so, at least, questionable. *Id.* at 3-4. ("During 2010 and 2011, the Board issued about 50 decisions on the merits in nursing facility appeals (that is, not including decisions on procedural issues such as timeliness and the like). During this two year period, the Board decided exactly *zero* of these appeals in favor of facilities on the merits. In contrast to past years, the Board did not reverse a single ALJ Decision against a facility. During this same time period, ALJs rendered about 100 decisions on the merits. Of these, about five set aside all cited deficiencies, and another half dozen set aside most deficiencies, and/or most of the sanctions CMS imposed. CMS did not appeal every such decision, but in those that CMS did appeal to the Board, the Board reversed (or remanded for reconsideration) *every* such favorable decision. Contrast this track record to past years, where facilities routinely "won" about a quarter of all cases decided on the merits." (footnotes omitted) Moreover, even when a provider successfully

Consequently, effective enforcement of IOM's "prime directive" and OBRA '87's mandate that nursing facility care be assessed using a methodology that produces valid and accurate results is not possible as a practical matter. The inability to challenge the validity and accuracy of the survey protocol also means that government regulators are largely unaccountable for failing to correct survey systems flaws.

There is, therefore, little incentive—and virtually no compulsion—for regulators to assure that surveys produce valid, reliable, consistent, and accurate findings of both compliance and noncompliance. To the contrary, by continually focusing on under-citation, such individuals advance the politically and professionally desirable goals of enhancing their image and that of the enforcement agencies as regulators who are tough on noncompliance and committed to ferreting out problems. Accurate policing has a limited and less compelling constituency. Thus, declines in deficiency rates and severity are dismissed—based on speculation and without empirical evidence—as attributable to unknown factors and not indicative of care improvements.¹⁷

Relying on deeply flawed data—and encouraging others to do so—without acknowledging the flaws is a betrayal of the public trust. Most importantly, it deceives those whose health, safety, and welfare depends upon the credibility and effectiveness of the nursing facility survey process. It also sends a message to the regulated and to the first level regulators that the survey process is "effective," whether or not it is accurate, because it serves the narrow political interests of regulators and others who rely on the data despite knowledge that the survey is not an accurate or consistent performance measurement tool. Finally, after almost thirty years, it has institutionalized a survey system that is arbitrary, inconsistent, ineffective, and does not promote quality. That result, it seems, is more palatable, administratively and politically, than admitting that the survey process has never been tested and validated as OBRA '87 required and, consequently, that the government has no effective means of distinguishing quality care from deficient care.

II. THE NURSING HOME SURVEY SYSTEM: HISTORY AND BACKGROUND

In order to be eligible for payment by the Medicare and Medicaid programs, a nursing facility must be certified as in compliance with federal and

disputes a deficiency citation, the government cannot be compelled to correct its records or precluded from public disclosures alleging that the facility violated care standards. *See, e.g.*, *Bryn Mawr Care Inc.*, 749 F.3d at 603. Consequently, at least one level, victory is pyrrhic.

17. *Compare* Statement of Leslie v. Norwalk, Acting Admin. CMS, Taking Quality Nursing Home Care to the Next Level (Apr. 19, 2007), available at https://www.nhqqualitycampaign.org/files/NHQC0407_LOS_CMSnorwalk.pdf (discussing extensive, voluntary care improvement programs instituted by nursing home industry).

state quality standards.¹⁸ The federal quality standards are embodied in a detailed set of substantive care standards contained in federal regulations.¹⁹

A nursing facility obtains certification that it is in compliance with federal quality standards by submitting to an inspection, called a survey, performed by state inspectors.²⁰ These inspectors, known as “surveyors,” are generally the same individuals who perform inspections to determine a nursing facility’s compliance with state licensure laws.

The work of the state inspectors is subject to periodic oversight and review by federal officials.²¹ Federal review and oversight may take a variety of forms ranging from desk review of the state surveyors’ findings, on-site surveys by federal inspectors shortly after the state inspection, “looking behind” the state results and comparative federal surveys.²² Importantly, whether a survey is conducted by state surveyors or by federal surveyors, the Medicare and Medicaid Acts require that all surveys must be conducted using the same quality assessment protocols and methodologies required pursuant to OBRA '87.²³

In the first twenty-five years after passage of the Medicaid Act, consensus developed among all concerned that the initial quality standards governing nursing facility care delivery and the enforcement of those standards did not result in the quality of care expected.²⁴ A number of reasons were given to explain why the quality desired was not the quality delivered.

First, it was widely believed that the quality standards assessed the wrong things. The initial substantive care standards in the regulations were primarily structural; that is, the regulations primarily focused upon such things as the organization of the nursing facility,²⁵ the resources available,²⁶ and qualifications of the staff.²⁷ The assumption implicit in those initial regulations was that if the right resources were in place, good results and care outcomes would follow because those responsible for care delivery would do the right thing.²⁸

18. 42 U.S.C. § 1395i-3(g)(3); 42 C.F.R. § 488.330 (2014).

19. 42 C.F.R. §§ 483.1-75 (2014).

20. *See* 42 U.S.C. § 1395i-3(g)(1)(A); 42 C.F.R. §§ 488.10(a), 488.11(a), 488.12 (2014)

21. 42 U.S.C. § 1395i-3(g)(3)(A).

22. IOM REPORT, *supra* note 6, at 37-38.

23. 42 U.S.C. § 1395i-3(g)(2)(C); 42 U.S.C. § 1395i-3(g)(3)(A).

24. IOM REPORT, *supra* note 6, at 12-16, 238-248 (discussing the history of nursing home regulation); *see also* Morford, *supra* note 8, at 131; Gary S. Winzelberg, *The Quest for Nursing Home Quality: Learning History’s Lessons*, 163 ARCH. INT. MED. 2552, 2552 (2003); Patricia A. Butler, *Assuring the Quality of Care in Nursing Homes*, 57 N.C. L. REV. 1317, 1336-37 (1979).

25. IOM REPORT, *supra* note 6, at 53.

26. *Id.* at 296.

27. *Id.* at 76.

28. *Id.* at 54.

Second, it also was widely believed that quality problems were the result, at least in part, of weaknesses in the inspection process. For example, every state had its own approach to conducting inspections.²⁹ Some states allowed individual surveyors to utilize ad hoc approaches; other states required a more structured approach. Some states required assessment of a sample of nursing facility residents; others required that the care of every resident be assessed by the inspectors.

The federal government itself required two different surveys. One survey was done for certification purposes to determine whether the facility was in compliance with Medicare and/or Medicaid requirements, and, therefore, eligible for payment for services rendered to beneficiaries of those programs.³⁰ The second survey, known as the inspection of care (IOC) survey, was done to determine whether a nursing facility's Medicaid residents were receiving adequate care.³¹ The IOC survey determined whether the state was entitled to payment of federal matching funds for the care of each Medicaid resident.³²

The two surveys were generally done in conjunction with one another but using different quality assessment methods.³³ The certification survey assessed the care rendered to the facility's residents generally, i.e., in the aggregate. There was no requirement that each resident's care be individually assessed. The IOC survey was conducted independently of the certification survey, but often by the same state surveyors who had conducted the certification survey. IOC surveyors, however, were required to conduct an individualized assessment of the condition, needs, and adequacy of the care provided based on record review and observation of each Medicaid resident.³⁴

In addition to not requiring that the certification survey be conducted using a uniform methodology, the substantive care criteria applied to evaluate a nursing facility were vague. Consequently, the surveyors also had broad discretion to determine the meaning and application of the regulations in individual circumstances. Such discretion, it was widely believed, introduced an unacceptable level of subjectivity and inconsistency into the survey process.³⁵

Third, even when it was generally agreed that the process had identified quality deficiencies, enforcement was inconsistent. Initially, termination of a facility's Medicare or Medicaid participation was the only remedy for quality

29. *Id.* at 35.

30. IOM REPORT, *supra* note 6, at 69.

31. *Id.* at 140.

32. *Id.* at 319.

33. *Id.* at 140.

34. *See, e.g., id.* at 71.

35. IOM REPORT, *supra* note 6, at 108, 140-42 (describing IOC survey).

deficiencies.³⁶ Termination, however, was rarely used because: (1) the nursing facility's residents would bear the brunt of the sanction when they were forcibly relocated; and (2) there were insufficient numbers of nursing facilities to meet the demand for nursing facility services.³⁷ In short, the passage of the Medicare and Medicaid Act had created a benefit and guaranteed payment for delivery of such services, but there were insufficient nursing home beds available to satisfy the demand for services. This meant that regulators were extremely reluctant to aggressively enforce quality standards because closure of a nursing facility might leave its residents with nowhere to go.³⁸ As a result, surveyors were viewed as consultants whose job was to encourage and assist delinquent nursing facilities to come into compliance with federal quality standards.³⁹

The impact of the systemic flaws in the regulatory system was confirmed by, among others, a 1971 General Accounting Office (GAO) Report titled "Problems in Providing Proper Care to Medicare and Medicaid Patients in Skilled Nursing Homes."⁴⁰ The Report found "many nursing homes participating in the Medicaid program—and in some cases, the Medicare program—were not adhering to Federal requirements for participation."⁴¹ According to GAO, "[t]he nonadherence to requirements resulted primarily from weaknesses in State procedures for certifying eligibility of homes and from ineffective State and HEW enforcement of Federal requirements."⁴²

In the mid-1980s, the regulatory environment began to change dramatically. The consultative approach was displaced in favor of a more aggressive, enforcement-orientated approach. A number of factors contributed to this change: (1) litigation challenging the effectiveness of the inspection process and, specifically, whether that process actually assessed the quality of care provided; (2) a series of notorious reports of inadequate and poor quality care in long-term care facilities; (3) shifting in the care needs of the typical long-term care resident from custodial care to more complex, professional, and

36. *See id.* at 13, 154.

37. *Id.* at 13.

38. *Id.* at 240-41. The problem was two-fold. First, sufficient numbers of facilities and beds to meet demand did not exist or were not located appropriately to serve the population in need. In addition those facilities that did exist, relatively few met the certification standards. In the first year, after adoption of the Medicare program, 6,000 facilities applied for certification but only 740 qualified; another 3,000 facilities were certified as in substantial compliance. *Id.*

39. *Id.* at 150.

40. U.S. GOV'T GENERAL ACCOUNTING OFFICE, B-164031(3), PROBLEMS IN PROVIDING PROPER CARE TO MEDICAID AND MEDICARE PATIENTS IN SKILLED NURSING HOMES 2 (May 28, 1971).

41. *Id.* at 2.

42. *Id.* at 9. This GAO report found that more than one-half of the nursing facilities audited in three states were in violation of federal certification standards pertaining to nurse staffing, physician visits and fire safety. *Id.* at 2-3.

post-acute care; (4) sharply escalating federal and state payments for long-term care coupled with uncertainty regarding whether such payments actually supported quality of care; and, (5) the maturing of the nursing facility industry such that sufficient beds generally were available to meet the demand for nursing facility services.

In Colorado, for example, a group of nursing facility residents sued both the federal and state governments claiming that the inspection methodology used to determine compliance with substantive care standards was ineffective.⁴³ The plaintiffs contended that the survey tools inappropriately assessed only a facility's capability of delivering care, not actual patient outcomes or the effectiveness of such care. As a result, the plaintiffs argued, the government had no idea whether Medicaid payments to long-term care facilities were buying quality care.

In 1984, the United States Court of Appeals for the Tenth Circuit agreed. The court held:

The Secretary has a duty to promulgate regulations which will enable her to be informed as to whether the nursing facilities receiving federal Medicaid funds are actually providing high quality medical care [T]he Secretary has failed to discharge her statutory duty altogether.⁴⁴

On remand, the District Court ordered the Secretary of HHS⁴⁵ to develop a survey protocol that was prescriptive, that was embodied in regulations and that assessed quality based on actual resident outcome, rather than based on whether a facility's care process and structure indicated that it was capable of delivering quality of care.⁴⁶

At about the same time, after several failed attempts to revise the survey, certification, and enforcement processes administratively,⁴⁷ Congress ordered HHS to contract with the IOM to study the existing regulatory system and to make recommendations for improvements. In 1986, the IOM returned its landmark report, "Improving the Quality of Care in Nursing Homes."⁴⁸

After discussing the current state of nursing facility regulation, the IOM made a fundamental point:

43. Estate of Smith v. Heckler, 747 F.2d. 583, 585 (10th Cir. 1984).

44. *Id.* at 591.

45. At the time the agency was known as the Department of Health, Education, and Welfare (HEW). Similarly, at the same time, the component of HEW responsible for administration of the survey process was the Health Care Financing Administration. The name changes have no practical significance for present purposes. Therefore, throughout this article the entities' current names are used – the Department of Health and Human Services (HHS) and the Centers for Medicare and Medicaid Services (CMS).

46. *Id.* at 591-92.

47. IOM REPORT, *supra* note 6, at 1.

48. *See generally id.*

Government regulation of nursing homes for quality assurance purposes has three components: (1) the criteria used to determine whether a nursing home is providing care of acceptable quality . . . (2) the procedures used to determine the extent to which nursing homes comply with the criteria, and (3) the procedures used to enforce compliance. The three components are the like the legs of a three-legged stool: All are equally important.⁴⁹

IOM emphasized that the “development and use of valid and reliable instruments to measure quality are critically important to quality assurance and to regulation.”⁵⁰ According to IOM, the survey protocol should be validated by “[t]he findings of careful empirical studies,” and the survey protocol should be “carefully analyzed for interrater reliability in the use of the instrument [.]” Such testing was necessary because “[s]urvey findings must be valid and reliable as well as consistent . . .”⁵¹ “[S]tatistically defensible . . . procedures,” IOM reiterated, were “essential . . . to achieve valid,⁵² consistent and reliable findings.” According to IOM, this meant that “[a]ll survey protocols (instruments and procedures) should be tested so that when used by properly trained surveyors they produce consistent and reliable findings.”⁵³

In 1987, Congress adopted virtually all of IOM’s recommendations, essentially without modification, in OBRA ’87.⁵⁴ Among the many significant and far-reaching changes made by OBRA ’87, three are important here.

First, the statute shifted the focus of the inspection process away from its prior concern with process and structural requirements. Rather than asking whether a facility was organized in a way that made the facility capable of providing quality of care, OBRA ’87 required that quality be determined by focusing primarily on the actual outcomes of resident care.⁵⁵ In short, following OBRA ’87, surveyors were no longer to assess what a long-term care facility could do; surveyors were to determine whether the facility residents actually received care to meet their needs based on whether those residents achieved or maintained appropriate care outcomes.⁵⁶

Second, the statute contemplated that regulators would aggressively enforce compliance with quality standards. To that end, OBRA ’87 provided enforcement tools, including a variety of intermediate sanctions, including Civil Money Penalties, temporary management, and denial of payment in

49. *Id.* at 69.

50. *Id.* at 56.

51. *Id.* at 128.

52. *Id.*

53. IOM REPORT, *supra* note 6, at 128.

54. Omnibus Reconciliation Act of 1987, Pub. L. No. 100-203, 101 Stat. 1330-1339 (Dec. 22, 1987).

55. *See, e.g.,* Morford, *supra* note 8, at 131 (“The basic change overriding all others is . . . emphasis on patient outcomes”).

56. 42 U.S.C. § 1395i-3(b)(2) (2012).

certain circumstances, short of termination of a nursing facility's participation in Medicare and Medicaid programs.⁵⁷

The new enforcement process was intended to limit the time between discovery of a violation of care standards and imposition of a remedy as a means of encouraging compliance. Moreover, the menu of intermediate sanctions, i.e., other than termination of program participation, was intended to punish the facility without directly burdening the residents as had previously been the case when program termination was the only option for noncompliance. As a result, it was expected that remedies would be utilized without hesitation.

Finally, OBRA '87 also included an express requirement that the survey protocol utilized to assess the quality of nursing facility care be tested and validated prior to the date the statute required that the new approach be implemented, January 1, 1990.⁵⁸ Further, the federal Secretary and the states were mandated to implement ongoing programs to assure that survey results were consistent.⁵⁹ The statute provided:

(C) SURVEY PROTOCOL. Standard and extended surveys shall be conducted —

(i) based upon a protocol which the Secretary has developed, tested and validated by not later than January 1, 1990, and . . .

(D) CONSISTENCY OF SURVEYS. Each State and the Secretary shall implement programs to measure and reduce inconsistency in the application of survey results among surveyors.⁶⁰

The statute also provided that the failure to develop, test, and validate the survey protocol did not relieve a state or the Secretary of the responsibility to survey nursing facilities.⁶¹ Although nursing facility surveys would continue to be conducted primarily by state inspectors, the statute required oversight by the Secretary. Among other things, the statute required that federal surveyors conduct follow-up, or validation surveys, shortly after the state inspections in order to determine if the state survey results were accurate and valid.⁶² The survey methodology utilized to assess nursing facility quality was considered sufficiently important that Congress required the federal oversight surveys to be conducted using the same, "tested and validated" methodology that the state surveyors were required to utilize. Specifically, OBRA '87 provided:

57. 42 U.S.C. § 1395i-3(h)(1)-(2).

58. 42 U.S.C. § 1395i-3(g)(2)(C).

59. *See* 42 U.S.C. § 1395i-3(g)(2)(D).

60. 42 U.S.C. § 1395i-3(g)(2)(C)(i)-(D).

61. 42 U.S.C. § 1395i-3(g)(2)(C).

62. 42 U.S.C. § 1395i-3(g)(3)(A).

(3) Validation Surveys.

(A) The Secretary shall conduct onsite surveys of a representative sample of skilled nursing facilities in each State In conducting such surveys, the Secretary shall use the same survey protocols as the state is required to use⁶³

The statute requires that the Secretary conduct such surveys in a minimum of five percent, but not less than five, nursing facilities in each state.⁶⁴

III. THE MANDATE TO “TEST AND VALIDATE:” THIRTY YEARS OF GOVERNMENT STUDIES OF THE ACCURACY, VALIDITY, AND CONSISTENCY OF SURVEY RESULTS

The Secretary failed to adhere to the OBRA '87 mandate with respect to the development and implementation of the new survey protocol almost from the beginning. Although OBRA '87 required that the survey protocol be tested and validated “not later than January 1, 1990,” i.e., before the new rules were effective,⁶⁵ the Secretary did not contract to test the survey protocol until September 1991, nine months after the testing and validation was supposed to have been completed.⁶⁶ Moreover, the contract provided that the testing of the survey protocol would not be completed for two years thereafter or, stated differently, the testing would not be completed for three years after the statutory deadline.⁶⁷

In fact, however, the testing was not completed within the expected two years after the contract was let. Instead, in 1993, only the preliminary results

63. *Id.*

64. 42 U.S.C. § 1395i-3(g)(3)(B).

65. 42 U.S.C. § 1395i-3(g)(2)(C).

66. ABT ASSOCIATES, INC., EVALUATION OF THE LONG TERM CARE SURVEY PROCESS, FINAL REPORT at i (1996) (hereinafter FINAL REPORT).

67. In 2002 the United States District Court for the District of Columbia held that the Secretary had no obligation to validate the survey protocol prior to January 1, 1990. Instead, the court held that the statute “implic[d] an ongoing process that will result in changes and refinements.” To be sure, OBRA'87 mandates an on-going process to assure that survey results are accurate and consistent. *See* 42 U.S.C. § 1395i-3(g)(2)(D). That requirement, however, is distinct from the mandate to “test and validate” the survey protocol prior to the January 1, 1990 implementation date. The Court also ignores that Congress, by requiring validation of the protocol by a specific date and prior to implementation did not contemplate an open-ended never-ending process that continues to this day. IOM's recommendation and OBRA'87 were intended to eliminate the validity, reliability and consistency problems that previously characterized survey findings. OBRA'87 required a result, not a process.

In the interest of full disclosure, the author of this article was counsel to the plaintiff in *Beverly Health and Rehabilitative Services, Inc. v. Thompson*, 223 F.Supp.2d 73 (D.D.C. 2002).

were available.⁶⁸ Those preliminary results found that the protocol did not produce valid results:

Despite implementation of the minimum data system (MDS) [tool for assessing nursing home residents' needs] and attempts to train surveyors in how to evaluate outcomes through the interpretive guidelines, the current survey does not lead to valid assessments of outcome.⁶⁹

In short, the preliminary results found that the new survey protocol failed to satisfy one of the basic requirements of OBRA '87, that the survey protocol produce valid results.

In 1996—six years after the statutory deadline for testing and validating the survey protocol—the Final Report largely came to the same conclusions. Among other things, the Final Report concluded “that the exercise of surveyor discretion can lead to harmful facility practices not being cited, or practices being inappropriately cited, in both quality of care and quality of life domains.”⁷⁰ Further, the Abt Report concluded, “the structure of the survey, the instruments employed, the training of surveyors, and the inclinations of many surveyors, do not readily result in the resident-centered, outcome-oriented process that the IOM envisioned.”⁷¹ According to the Final Report, “[t]o make judgments about outcomes, surveyors must rely heavily on their professional training and experience and on the brief training provided to them about the survey process.”⁷² Abt also found that:

For all the emphasis that surveyors place on ‘nursing knowledge,’ that knowledge in itself does not provide an adequate basis for many nurse-surveyors to properly make decisions about outcomes.⁷³

As a result, Abt found that deficiency citations were often inaccurate both because actual problems were not cited and because citations were issued when not warranted.

More specifically, Abt studied in depth nine key outcomes corresponding to the most prevalent quality issues in nursing facilities. The Final Report found false negative citations (i.e., under-citation) of quality problems in all of the domains from thirty-nine to seventy-eight percent of the time and found false positive citations (i.e., over-citation) in five of the domains from sixteen to twenty-eight percent of the time.⁷⁴ For example, Abt concluded that twenty-

68. See generally ABT ASSOCS., INC., BRIEFING POINTS ON THE PRELIMINARY EVALUATION RESULTS, EVALUATION OF THE NURSING HOME CERTIFICATION PROCESS (1993) [hereinafter BRIEFING POINTS].

69. *Id.* at 3.

70. FINAL REPORT, *supra* note 66, at 1.

71. *Id.* at 1-2.

72. *Id.* at 72.

73. *Id.* at 96.

74. *Id.*

three percent of citations for failure to improve function and twenty-eight percent of citations for pressure sore prevention and treatment were false positives.⁷⁵ On the other hand,⁷⁶ surveyors did not cite care problems related to failures to improve function in seventy-eight percent of the cases and in sixty percent of the cases with respect to failure to maintain acceptable nutrition parameters.⁷⁷ Overall, the percentage of false negative citations ranged from thirty-nine percent to seventy-eight percent (average: sixty-one percent). The percentage of false positive citations ranged from twenty-eight percent to five percent (average: fifteen percent). Thus, in the aggregate, Abt found that, for nine care domains studied, survey findings, on average, were inaccurate seventy-six percent of the time.

The problem, however, was not just that the survey findings were inaccurate. In addition, the Abt Final Report also found that survey findings and enforcement were not consistent. Abt reported “considerable” fluctuation in the rate at which deficiencies are cited and in the level and type of enforcement actions from region to region and state to state.⁷⁸ For example, in CMS’ Region 9, an average of 13.09 deficiencies were cited per facility, whereas in Region 2, an average of 4.76 deficiencies were cited.⁷⁹ Abt found that it was “difficult for individuals to exercise professional judgment consistently” due to the lack of instructions for selecting resident samples and

75. FINAL REPORT, *supra* note 66, at 96.

76. *Id.*

77. *Id.* at 104. These survey findings examined in detail “correspond to many of the most prevalent quality of care issues in nursing homes and to [the areas] frequently emphasized in quality of care research[.]” *Id.* at 104. The false negative and false positive citation rates for each area are as follows.

<i>Area</i>	<i>False Negative</i>	<i>False Positive</i>
Maintenance of Function	56%	16%
Personal Care	50%	9%
Improvement of Function	78%	23%
Incontinent Care	39%	6%
Pressure Sore Prevention and Treatment	64%	28%
Restraint Usage	56%	16%
Maintenance of Appropriate Nutritional Status	60%	20%
Neuroleptic Usage	67%	5%
Benzodiazepine Usage	75%	8%

Id. at 123.

78. *Id.* at 29, 33-34.

79. *Id.* at 31.

making determinations such as what constitutes a deficiency and whether a particular outcome was avoidable.⁸⁰

Over the course of the next several years, the conclusions of the Final Report were corroborated by other government-sponsored studies. For example, in 1998, a congressionally mandated study⁸¹ conducted jointly by CMS⁸² and Abt Associates found:

[T]he new enforcement regulation does not appear to be working as intended . . . [T]here is considerable variation among States in the degree of enforcement, as measured by rates of deficiency or substandard quality of care determination, [but] it is difficult to separate what proportion of the variation is due to true differences in nursing home quality and what proportion is attributable to differences in surveyor behavior.⁸³

It bears emphasis: The evidence supporting the conclusion that the survey protocol was ‘not working’ consisted exclusively of data that fewer, and less serious, deficiencies were being cited and that substantial variation existed in citation rates in different states.

CMS’s method of testing the validity of the survey process marks a significant and unexplained departure from the analysis previously used to determine the validity and reliability of survey results. The original Abt assessment of the survey protocol—consistent with CMS’s regulations—assessed the survey’s validity based on: (1) whether the survey cited those problems which actually existed; and (2) whether no citations were issued when no requirement had been violated.⁸⁴ The 1998 Report, however, does not

80. FINAL REPORT, *supra* note 66, at ii, 45, 56, 67, 72.

81. Omnibus Consolidated Rescissions and Appropriations Act of 1996, Pub. L. 104-134, § 516(d), 110 Stat. 1321-248 (1996).

82. HEALTH CARE FIN. ADMIN. & ABT ASSOCS., INC., REPORT TO CONGRESS: STUDY OF PRIVATE ACCREDITATION (DEEMING) OF NURSING HOMES, REGULATORY INCENTIVES AND NON-REGULATORY INITIATIVES, AND EFFECTIVENESS OF THE SURVEY AND CERTIFICATION SYSTEM 1 (1998) [hereinafter CMS REPORT]. The portion of the Report addressing the effectiveness of the survey process was written by the Health Care Financing Administration, as CMS formerly was known, not by the outside contractor, Abt Associates Inc. As a convenience, the text refers to CMS as the author. This portion of the Report appears to rest primarily on analysis of others’ research, and on data analysis rather than original field work.

83. *Id.* at 40.

84. See FINAL REPORT, *supra* note 66, at 122. CMS regulations specify a similar definition that tested the accuracy of survey results based on whether (1) appropriate deficiencies were cited; and (2) inappropriate citations were not issued. See, e.g., Medicare & Medicaid Programs; Survey, Certification and Enforcement of Skilled Nursing Facilities and Nursing Facilities, 59 Fed. Reg. 56,146-47 (Nov. 10, 1994).

Likewise, the original assessment of the validity of the OBRA survey emphasized that the integrity of the regulatory system would be undermined by both under and over citation:

In an effort to ensure standardization, HCFA devised a series of interpretive guidelines and accompanying probes for surveyors to use in applying Federal standards. The

analyze the results of the survey system from those two perspectives. Rather, CMS asks only whether sufficient citations—as defined by agency expectations—have been issued.

Because the number and severity of citations fell short of CMS's expectations, the survey results were deemed invalid.⁸⁵ There is no

interpretive guidelines are designed to provide the surveyors with a uniform set of indicators against which they can measure facility performance. The intent is to ensure consistent application of the regulatory requirements by surveyors across the country and to minimize, as much as possible, the influence of any individual's subjective judgment. If different surveyor teams were to introduce different criteria or thresholds to the survey process, the reliability of that process would be threatened. If a survey team were to write fewer – or more – negative findings and deficiencies as a result of its divergence from the mandated process, that too would undermine the regulatory system. Hence, assessing the consistency with which surveyors employ the survey process and the interpretive guidelines is an important objective of the current evaluation.

FINAL REPORT, *supra* note 66, at 42.

85. For example, the 1998 Report noted the agency's prediction that definitional changes made in 1995 would result in "a substantial increase in deficiencies." CMS REPORT, *supra* note 81, at 39-41. In fact, however, the mean number of citations per facility had continued to drop steadily. *Id.* at 538. For example, in surveys where at least one deficiency was cited, the mean number of deficiencies cited in 1994 was 8.3, but by 1997 the mean number had dropped to 6.2. *Id.* In addition, the number of surveys in which no deficiencies were cited increased sixty-nine percent between 1994 and 1997. *Id.* at 538.

CMS's 1998 Report identifies allegedly inaccurate citation rates in specific areas. For example, the Report alleges that surveyors do not accurately identify nutrition problems, an area which Abt also found problematic, but with respect to both over-citation (20 percent) and under-citation (60 percent). See FINAL REPORT, *supra* note 66, at 123; CMS REPORT, *supra* note 82, at 46. The Report notes that citations for certain federal requirements regarding nutrition had declined from fifteen percent in 1991 to less than five percent in 1996. CMS REPORT, *supra* note 82, at 565. The Report dismisses the possibility that the declining citation rates reflects improving care, concluding that the survey and certification results are not valid. *Id.* at 46, 565.

CMS's Report also concludes that the survey protocol does not accurately identify abuse of nursing home residents. *Id.* at 556. Once again, the Report identifies a decline in the rate of citations for abuse, but asserts that the survey results are not valid indicators of improving quality, stating that there is no independent evidence that the problem in fact has diminished. *Id.* at 578.

CMS's Report also identified "considerable variation among States," as measured by rates of deficiency citations, by the range (0 to 14 percent) of substandard quality of care determinations, by the pattern of immediate jeopardy findings and by the disparate range with which the states issued citations at levels that might trigger sanctions (from 22 to 100 percent of facilities in a given state). *Id.* at 541-543. The study also pointed out that, in 1998, five states cited no substandard quality care and another ten states cited virtually no substandard quality care. *Id.* at 42, 542. Again asserting that the citation patterns were not accurate. CMS stated that the results could not be ascribed "to true differences in nursing home quality" and characterized the "low rate" of substandard care citations as "implausible." *Id.* at xi, 39-43.

CMS's Report further asserts that "it is unclear, perhaps doubtful, that the degree of quality improvement observed" could have been responsible for the downward deficiency trend. *Id.* at 41, 540. Essentially, CMS concedes that they have no evidence explaining the results, that

consideration of the possibility that systemic problems with the survey also resulted in inappropriate false positive citations. Every subsequent government sponsored study of the survey protocol would utilize a similar approach, testing the validity of the system based on the incidence of false negative citations.⁸⁶

Nevertheless, despite focusing exclusively on whether the survey failed to identify care delivery problems, this study confirmed that the survey results—both problem identification and classification of severity—were inaccurate and inconsistent.⁸⁷ This conclusion was largely founded on three assumptions: (1) that the new survey should have produced more citations;⁸⁸ (2) that, although nursing home quality had improved, the quality improvements did not explain the observed decrease in citation rates and severity;⁸⁹ and, (3) that deficiency citations and classification made in simultaneous surveys conducted by CMS' contract consultants—notably, using a different, specially designed protocol, not the protocol used by the state surveyors who conducted the certification survey—were the “gold standard” of accurate problem identification and classification.⁹⁰ It bears emphasis that the study provides no empirical support for any of those assumptions.

One year later, and almost ten years after the date on which the statute required the survey protocol be validated, the Office of Inspector General (OIG) of HHS reported:

While it has now been more than a decade later since this legislation [OBRA '87] was passed, there has been... no methodical evaluation of whether the reforms it intended are actually working . . . [T]he lack of a systematic review makes it difficult to determine if this major legislation has been successful in improving nursing home care.⁹¹

Thus, OIG acknowledged, ten years after implementation, that OBRA '87's quality assessment and enforcement system had never been validated. Stated differently, fifteen years after the Tenth Circuit held in the *Smith* litigation that the Secretary had failed to fulfill her clear “duty to promulgate regulations which will enable her to be informed as to whether the nursing facilities

the reasons for the improving survey results are not known, but that the results cannot be accurate. *Id.* at 538.

86. *See infra* pp. 23-26.

87. CMS REPORT, *supra* note 82, at x-xii, 39, 46, 538-40, 542, 549-54, 565-79.

88. *Id.* at x, 40, 41, 538-40.

89. *Id.* at 31, 46, 539-40, 565, 578.

90. *Compare id.* at 20, 43, with 42 U.S.C. § 1395i-3(g)(3)(A) (2012) (requiring federal surveyors assessing the validity of state survey results to utilize “the same survey protocols as the State is required to use.”).

91. OFFICE OF INSPECTOR GEN., U.S. DEPT. OF HEALTH & HUMAN SERVS., OEI-02-99-00060, QUALITY OF CARE IN NURSING HOMES: AN OVERVIEW 4 (1999) [hereinafter OIG OEI-02-99-00060].

receiving federal Medicaid funds are actually providing high quality medical care.”⁹² The Secretary still had no basis to say whether the quality care for which the government was paying was, or was not, being delivered.

The following year, the head of GAO’s health care unit testified before the Senate Special Committee on Aging regarding the effectiveness of the survey process.⁹³ A colloquy between the committee chairman and GAO’s Dr. William Scanlon confirmed, yet again, that the quality assessments made using the survey protocol were not valid:

The Chairman. Dr. Scanlon, do you think the quality of the surveys and the information in the OSCAR data base [in which nation-wide survey results are recorded] is reliable enough to make judgments about the level of quality provided in the nation’s nursing homes?

Dr. Scanlon. Mr. Chairman, I am afraid it is not. I think that the variation that we see across states is troubling in the sense that we do not have confidence that the surveys are being administered consistently We have no confidence that this variation reflects the actual care that is being provided in homes across states.

The Chairman. Do you think that the tools [CMS] and the states are using to assess nursing homes give us information about the quality of care provided in them and what do they tell us about the care provided?

Dr. Scanlon. Mr. Chairman, I do agree with the industry that we are not measuring outcomes of care in the process of the survey⁹⁴

Thus, echoing what had been said in the original evaluation of the survey performed by Abt several years earlier, GAO concluded that survey results were not consistently applied, valid indicators of the quality of care actually provided. Moreover, GAO also concluded that the survey was not measuring the outcomes of resident care—the overarching goal, of OBRA '87.⁹⁵ To put it bluntly, not only were the survey results not valid, whatever the survey was assessing, it was not the performance measurement mandated by the OBRA '87.⁹⁶

92. *Estate of Smith v. Heckler*, 747 F.2d. 583, 591 (10th Cir. 1984).

93. *The Nursing Home Initiative: Two-Year Progress Report Hearing Before the Spec. Comm. on Aging, of the S. 106th Cong. III, 1* (2000) [hereinafter *Scanlon Hearing*].

94. *Id.* at 165-66.

95. *See, e.g.*, 42 C.F.R. § 442.30(a)(5)(ii) (2014) (“The survey process uses resident outcomes as the primary means to establish the compliance status of facilities.”); U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-07-794T, NURSING HOME REFORM: CONTINUED ATTENTION IS NEEDED TO IMPROVE QUALITY OF CARE IN SMALL BUT SIGNIFICANT SHARES OF HOMES: HEARING BEFORE THE SPECIAL COMMITTEE ON AGING OF THE SENATE 1 (2007).

96. *See Scanlon Hearing, supra* note 93, at 166 (contrasting with what seemed to be CMS’s belief, Dr. Scanlon stated that GAO has “never indicated . . . that a majority or even somewhat more than a minority of homes is [sic] providing poor care”).

The frankness of GAO's testimony may have been surprising, but the content should have sounded familiar. In addition to the findings made in the Abt Preliminary and Final Reports, the same points had been made repeatedly by those commenting on CMS' proposed regulations containing the survey's enforcement protocols.

For example, during the rule-making process, a third-party study of the validity and effectiveness of the survey process was submitted to CMS. The study had been conducted by Jean Johnson-Pawlson, Ph.D., a highly regarded nurse clinician, nursing school professor, and consultant to the IOM. That study concluded that survey results were inaccurate, inconsistent, and not based on resident outcomes.⁹⁷

The Johnson-Pawlson study evaluated "the accuracy and consistency of surveyor performance during standard nursing home surveys" based on analysis of official deficiency reports from 420 nursing homes in twenty-one states.⁹⁸ The Johnson-Pawlson Study found that:

- 45% of citations were not related to an actual or potential negative resident outcome[;]
- 42% of citations did not consider the seriousness of findings[;]
- 35% of citations did not consider how widespread (scope) the surveyor findings were in the facility[;]
- 36% of the citations lacked sufficient documentation to support the deficiency[; and]
- 26% of deficiencies cited were not supported by information detailed in the Long Term Care Survey Guidelines.⁹⁹

The data also indicated "considerable variation between states" and concluded that "the probability of these differences being due to chance is zero."¹⁰⁰ The Johnson-Pawlson Study concluded that to assess and enforce quality standards using CMS's on defendants' survey process was "a major flaw:"

The magnitude of the potential inaccuracies is greater than what should be acceptable to consumers, providers and regulatory agencies. With 30% of citations potentially not meeting reasonable standards for determination of deficiencies, the current survey process may not be a valid measure of the extent to which facilities comply with federal requirements.¹⁰¹

97. Rulemaking Record for Survey, Certification, and Enforcement of Skilled Nursing Facilities and Nursing Facilities, 59 Fed. Reg. 56,116 (Nov. 10, 1994) at 5960, 5967 [hereinafter Rulemaking Record].

98. *Id.*

99. *Id.* at 5960. See also Jean Johnson-Pawlson, *Survey Inconsistencies Point To Need For Better Training*, PROVIDER 60 (Jan 1993).

100. Rulemaking Record, *supra* note 97, at 5967-68.

101. *Id.* at 5970.

The reliability of the Johnson-Pawlson study was not questioned; instead, the OIG would soon rely on the study as the basis for its own critique of the survey methodology.

Other national and state organizations also, conducted their own studies, reaching similar—and corroborating—results. The North Dakota Long-Term Care Association submitted “a statewide, six month independent study [concluding] that the current survey system is flawed.”¹⁰² The study was performed by former manager of the North Dakota Health Facility Survey Agency.¹⁰³ The study evaluated deficiency citations for forty-three in-state nursing facilities.¹⁰⁴ Among the study findings were the following:

- | | | |
|---|--------|-------------------|
| • Deficiency not Based on Negative outcomes | 77/226 | 34% |
| • Survey Procedures not Followed | 57/226 | 25% |
| • Deficiency Not Supported by Regulation | 42/226 | 19% |
| • Incorrect Regulation or Tag cited | 21/226 | 9% ¹⁰⁵ |

The Association’s report also concluded that “[t]he study indicated there is inconsistent application of federal laws by surveyors.”¹⁰⁶

A study conducted by the Kentucky Association of Health Care Facilities covered the twelve-month period from October 1990 through September 1991.¹⁰⁷ The study concluded:

152 of 346 deficiencies [were] inappropriately cited under the OBRA survey process The summary found that deficiency content was incomplete and did not address the regulation. This problem was consistent throughout the review. Deficiencies which were found to be incomplete did not reflect that the steps as outlined in the training manuals etc., were followed in order to gather information prior to deficiency formulation.¹⁰⁸

Those findings corroborated what had been said previously and should have been surprising to no one.

Although the problem of inconsistency between states was well known, the Kentucky study also identified problematic inconsistency in deficiency citations within the state as well as among nearby states.¹⁰⁹ Tracking the deficiencies cited by the state surveyors revealed “no pattern of what type deficiencies are cited per region” among the four regions in the state.¹¹⁰ The Association noted that “[e]ach office seems to vary not only in which area of

102. *Id.* at 020134-48, 020134.

103. *Id.* at 020134.

104. *Id.* at 020141.

105. Rulemaking Record, *supra* note 97, at 020135.

106. *Id.*

107. *Id.* at 4110-13.

108. *Id.* at 4111.

109. *Id.* at 4111-12.

110. Rulemaking Record, *supra* note 97, at 4111.

the regulations the problem lies, but in the number of deficiencies being cited per region.”¹¹¹ Moreover, “[i]n reviewing the deficiency profiles compiled by [CMS’s], Region IV, there is much discrepancy among the eight [nearby] states, as to the [seriousness of] . . . deficiencies and the area of concentration.”¹¹²

An analysis of survey results by the Georgia Nursing Home Association identified “about a 30% error rate.”¹¹³ Further,

[t]he inconsistency in our State and the HCFA Region show a variance between states ranging from 4% to 54% for the same tag number. OSCAR reports [a computerized CMS database compiling survey results] show Georgia leading the nation in some program requirements and neighboring states near the bottom. Georgia also leads the nation in [the percentage of facilities cited at the most serious level of] deficiencies with 14% of the facilities receiving [such] citations while the national average is 6%.¹¹⁴

Data submitted by the Louisiana Nursing Home Association likewise found intra-state variation in citations. For example, one regulatory violation was cited fourteen times in one of the seven regions in the state, whereas another region cited the same violation only once.¹¹⁵

Whatever discount might arguably be necessary for comments offered by provider representatives, should have been offset because the concerns were corroborated by government officials directly responsible for the survey and certification program. For example, one of CMS’s Regional Offices was critical of the inconsistency of survey results in the same time frame.

We have found the new long term survey protocols and guidelines do not provide a method for producing consistent survey decisions. Therefore, a State agency that cites more deficiencies will have more facilities facing alternate remedies than states who historically cite fewer deficiencies. [The agency’s] OSCAR report #19 [recording all deficiency citations] gives evidence of this by showing the number of deficiencies cited that are above or below two standard deviations of the national mean.¹¹⁶

Echoing what Abt had said years earlier, the Regional Office continues, explaining that:

CMS had not determined a precise definition of individual data tags in LTC regulations. In other words, when a LTC survey team collects evidence and determines there is a deficiency, it is not clear which data tag should be cited.

111. *Id.*

112. *Id.* at 4112.

113. *Id.* at 6825.

114. *Id.*

115. Rulemaking Record, *supra* note 97, at 009174.

116. *Id.* at 032815 (Letter from the Associate Regional Administrator, Division of Health Standards and Quality, HCFA, Region X).

There are in fact, multiple, correct, data tags that can be cited for the same deficiency. The danger in this situation is obviously lack of surveyor consistency on a national scale. The application of sanctions will be inconsistent.¹¹⁷

That the criticisms were made in writing by an Associate Regional Administrator of Division of Health Standards and Quality—the division of CMS responsible for day-to-day oversight of the survey—should have been unsettling. Individual states also objected that CMS did not have “consistently applied and meaningful standards.”¹¹⁸ Inconsistency, according to another state, was inevitable because “HCFA does not have a defined threshold for deficiencies and admits that deficiencies are based on surveyor judgment.”¹¹⁹ States also criticized the proposed rule for ignoring resident outcomes.¹²⁰ The State Medicaid Directors’ Association criticized the proposed rule for not “focus[ing] on resident outcomes.”¹²¹

In May 1993, the OIG issued a report “State Progress in Carrying Out the Nursing Home Survey Reforms.”¹²² The OIG Report found that the nineteen states it had studied “are facing implementation problems that could jeopardize the intent of the nursing home reforms.”¹²³ Relying on the Johnson-Pawlson Study, OIG warned:

That study raises questions about the extent to which surveyors . . . have adjusted to the new outcome focus: 45 percent of the deficiency statements failed to consider either actual or potential negative outcomes. When these deficiencies related more to a home’s structure than to a resident’s outcome were excluded, 43 percent still failed to consider actual or potential outcome.¹²⁴

OIG’s fear that the intent of the nursing home reforms was in jeopardy would be confirmed within two months by “a report for which defendants contracted

117. *Id.* at 032817.

118. *Id.* at 032080 (citing Letter from Ga. Dep’t. of Human Resources).

119. *Id.* at 012159 (citing Letter from State of Me. Dep’t. of Human Services); *see also id.* at 032082-83 (citing Letter from Ga. Dep’t. of Human Resources criticizing the enforcement regulations for “not giv[ing] notice of proscribed conduct” to the extent “that men of common intelligence must necessarily guess at its meaning and differ as to its application”).

120. *Id.* at 013809 (citing Letter from Ark. Dep’t. of Human Services).

121. *Id.* at 013389.

122. Transcript of Administrative Record at 14685, Beverly Health and Rehabilitation-Spring Hill v. Health Care Fin. Admin., Dec. No. CR553 (Department Appeals Board, Civil Remedies Div. 1998), *aff’d* Dec. No. 1696 (Department Appeals Board, App. Div. 1999) [hereinafter Administrative Record].

123. *Id.* at 14686; OFFICE OF INSPECTOR GEN., OIG OEI 01-91-01580, STATE PROGRESS IN CARRYING OUT THE NURSING HOME SURVEY REFORMS at ii (1993) [hereinafter OIG OEI 01-91-01580].

124. Administrative Record, *supra* note 122, at 14689; OIG OEI 01-91-01580, *supra* note 123, at 4.

on surveyor decision making and consistency as well as quality of care and quality of life measurement.”¹²⁵

Many of the problems identified by GAO, OIG and the commentary on the proposed enforcement rule also were largely confirmed by IOM in a 2001 follow-up to the 1986 report that had been the foundation for OBRA '87's survey and enforcement legislation.¹²⁶ The 2001 IOM report found that:

Several studies support the conclusion that the current survey process fails to identify important quality of care problems

[States] very substantially in their survey findings and no evidence suggests that this variation is a function of corresponding variation in the quality of care provided in the states.¹²⁷

In other words, whether the care provided by a nursing facility was found acceptable or unacceptable depended on where it was located, not on the actual quality of the care delivered.

One year later, in 2002, CMS sought proposals to redesign the survey methodology. According to the Request for Proposals (RFP), CMS sought to redesign the survey process because:

There is substantial State variation in nursing home deficiency rates, and scope and severity determinations for identified deficiencies. . . . Some of these differences might be accounted for by real quality differences among the nursing homes, but it is extremely unlikely that average differences in this great of magnitude for entire states can be explained in any significant degree by real quality of care differences. . . . [N]or can possible state differences in morbidity of nursing home populations' account for these differences in citation patterns.¹²⁸

Further, according to CMS, “[t]he State-to-State variability in magnitude and pattern of various kinds of deficiency citation rates indicate that there is inconsistency in the survey process and raises doubts about the appropriateness in some deficiencies of either over or under identification of problems.”¹²⁹

Despite acknowledging that: (1) it was “extremely unlikely” that survey results reflected “real quality of care;” and (2) doubts about “the

125. Administrative Record, *supra* note 122, at 14684; OIG OEI 01-91-01580, *supra* note 123, at 10. OIG appears to be referring to the Abt Associates' preliminary reports discussed previously, which were presented to HCFA in July, 1993. Administrative Record, *supra* note 122, at 14303-14312; BRIEFING POINTS, *supra* note 68, at 1-9.

126. DIV. OF HEALTH CARE SERVS., INST. OF MED., IMPROVING THE QUALITY OF LONG-TERM CARE 77-79, (Gooloo S. Wunderlich and Peter O. Kohler, eds., 2001).

127. *Id.* at 145, 151.

128. CTRS. FOR MEDICARE & MEDICAID SERVS., REQUEST FOR PROPOSAL, No. CMS-02-017/JB, IMPROVING THE CONSISTENCY OF THE NURSING HOME SURVEY PROCESS 3-4 (2002) [hereinafter IMPROVING THE CONSISTENCY].

129. *Id.* at 4, 10.

appropriateness,” i.e., accuracy of problem identification,¹³⁰ CMS continued to utilize the survey results to determine facilities’ compliance with Medicare and Medicaid requirements and as a basis for imposing sanctions on facilities with no accommodation for the admitted concerns about the validity and accuracy of survey results.¹³¹ To the contrary, rather than acknowledge the potential impact on individual providers, CMS, as a practical matter, had insulated its failure to validate the survey protocol from challenge. CMS’ regulations prohibited facilities from challenging survey results or enforcement actions in appeals before Administrative Law Judges based on defects in the survey protocol from oversight.¹³²

It is fair to ask what the consequences might be for a private litigant who offered or relied on evidence acquired in violation of a statute or regulation or on evidence developed using non-standard protocols and methodologies.¹³³ It

130. *Id.* at 3, 9.

131. US GOV’T ACCOUNTABILITY OFFICE, GAO-03-561, NURSING HOME QUALITY: PREVALENCE OF SERIOUS PROBLEMS, WHILE DECLINING, REINFORCES IMPORTANCE OF ENHANCED OVERSIGHT 4-5 (2003) [hereinafter PREVALENCE OF SERIOUS PROBLEMS].

132. 42 C.F.R. § 488.305(b) (2014) (“The State survey agency’s failure to follow the procedures set forth in this section [regarding the resident sample, surveying the quality of care, auditing assessments and care plans and assessing compliance with resident’s rights requirements] will not invalidate otherwise legitimate determinations that a facility’s deficiencies exist.”); 42 C.F.R. § 488.318 (Failure to “cite only valid deficiencies,” “conduct surveys in accordance with [regulations]” or use the forms, procedures, protocols, standards and methods specified by CMS “does not . . . invalidate adequately documented deficiencies.”). Note, however, that a state may be denied federal funding if it conducts surveys in violation of such requirements 42 C.F.R. § 442.30(a)(1) (2014). Moreover, a provider agreement is not valid evidence of compliance if the survey is not performed in accordance with federal requirements. 42 C.F.R. § 442.30(a)(5).

133. *Cf. e.g., Daubert v. Merrell Dow Pharm. Inc.*, 509 U.S. 579, 594, 600 (1993) (Overarching test of admissibility of evidence under Rule 702 is “scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission.” Court must assess “whether the reasoning or methodology underlying the testimony is scientifically valid,” by considering, among other things “whether it can be (and has been) tested” empirically, “the known or potential rate of error . . . and the existence and maintenance of standards controlling the technique’s operation”); *Schering Corp. v. Pfizer*, 189 F.3d 218, 234 (2d Cir. 1999) (noting that courts deciding whether to admit surveys in these circumstances should therefore examine their trustworthiness—and ultimately their weight, if admitted—both in terms of their methodological strengths and in terms of their proneness to faulty memory and perception.); *Toys “R” Us, Inc. v. Canarsie Kiddie Shop, Inc.*, 559 F. Supp. 1189, 1205 (E.D.N.Y. 1988) (“The trustworthiness of surveys depend upon foundation evidence that (1) the “universe” was properly defined, (2) a representative sample the universe selected, (3) the questions to be asked of interviewees were framed in a clear, precise and non-leading manner, (4) sound interview procedures were followed by competent interviewers who had no knowledge of the litigation or the purpose for which the survey was conducted, (5) the data gathered was accurately reported, (6) the data was analyzed in accordance with accepted statistical principles and (7) objectivity of

is also fair to ask whether a private litigant asserting breach of contract would be permitted to prove such a claim using evidence developed in derogation of the breach assessment procedures required by the contract.

Significantly in light of later events, CMS' RFP discussed the agency's view of the relative importance of consistent survey results and accurate survey results. Recognizing that survey results might be consistent but inaccurate, CMS made the seemingly obvious point that accurate survey results were the prime objective, and consistency, although important, was secondary:

In general, we would expect a more consistent survey process to result in a more accurate survey process. However, the goal of improving the consistency of the survey process must not be attained at the expense of incurring a less accurate survey.¹³⁴

CMS' actions would belie that assertion before the redesign of the survey protocol was complete.

In the same year that the RFP was issued by CMS, GAO again confirmed that deficiency findings were unacceptably inconsistent. GAO also reiterated that such inconsistency could not be explained by differences in quality:

[W]e have found considerable variation nationwide in the reporting of deficiencies . . . such differences in reporting make comparisons across states difficult since it cannot be determined whether observed differences are due to real variations in quality or to inconsistent application of standards.¹³⁵

Thus, fifteen years after OBRA '87 was adopted and twelve years after the date by which the protocol was mandated to be tested and validated, it could not be said that protocol accurately assessed quality. In 2002, as in 1984, when the Tenth Circuit ordered the Secretary to promulgate an effective survey methodology, the Secretary still had no basis to conclude that the government was getting the quality care for which it was paying.¹³⁶

Also in 2002, another GAO study reported that, when federal surveyors conducted oversight surveys of state survey activities, the federal surveys found substantially more serious problems than were found by the state:

[T]he subjective nature of the survey process means that the surveyors may apply standards unevenly. Indeed we previously have reported that during attempts to validate the findings of state surveyors, federal surveyors have

the entire process was assured. Failure to satisfy one or more of these criteria may lead to exclusion of the survey").

134. IMPROVING THE CONSISTENCY, *supra* note 128, at 4.

135. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-02-431R, NURSING HOMES: EXPENDITURES AND QUALITY 4 (2002) [hereinafter EXPENDITURES AND QUALITY].

136. Estate of Smith v. Heckler, 747 F.2d 583, 591 (10th Cir. 1984); EXPENDITURES AND QUALITY, *supra* note 135, at 11-12.

found more than three times the number of serious care problems recorded by state surveyors.¹³⁷

GAO failed to note, however, that the federal surveyors are required by law: (1) to conduct the oversight survey in roughly the same time frame as the state survey; and (2) to use the same survey tools and methodologies as the state surveyors when conducting oversight surveys.¹³⁸

In other words, the federal surveyors and the state surveyors were mandated to conduct the nursing facility survey in precisely the same time frame and in the same way. It would be reasonable to expect, therefore, that the state survey and the federal survey would generate substantially similar, if not approximately the same, results. Nonetheless, the federal and the state surveyors obtained significantly different results despite using the same nursing facility survey protocol. The significance of the discordance between federal and state survey results ought not be minimized. Nor should it be overlooked, that GAO's report—as do all other federal studies—presumes without either evidence or explanation that the federal surveyors' results are superior to those reported by the same state surveyors.

It is one thing to say that surveyors in different states and in different regions of the country produce inconsistent results using the same survey protocol. Although there is no evidence that explains inter-state differences, in theory, at least, such differences might be attributed to—though not excused by—differing standards of care, differences between facilities and resident populations, differences in resources, differences in time and circumstances of the survey, differences in public health status and other factors.

The difference between state and federal results, however, cannot be explained by such factors. To the contrary, while inter-state comparisons are affected by numerous variables, many of the circumstances intuitively most likely to generate different results—inspection methodology, time, place, residents and caregivers—are held relatively constant when federal and state surveys are compared. It would seem that the federal-state differences are indicative of the dominant role of surveyor discretion compounded by the fact that, as the Abt Final Report concluded, the survey's reliance on professional judgment is misplaced because nursing judgment is not adequate to make the complex decisions required by an outcome-based survey¹³⁹

137. EXPENDITURES AND QUALITY, *supra* note 135, at 4.

138. 42 U.S.C. § 1395i-3(g)(3)(A) (2012).

139. *See generally* FINAL REPORT, *supra* note 66.

Moreover, it ought to be noted that there is no guarantee that the complex outcome analysis required by the survey process will be made by a nurse. The regulations require that the team that conducts a standard survey include one registered nurse. However, the remainder of the team is composed of interdisciplinary professionals. "Examples of professionals include physicians, physician assistants, nurse practitioners, physical, speech or occupational therapists,

One year later, in 2003, another GAO report addressed the reliability of the nursing facility survey process. GAO also addressed actions proposed by CMS to improve the validity and reliability of the survey process:

The agency [i.e., CMS] is committed to implementing only those portions of the new methodology that are proven to be significantly more effective than the current survey methodology... We continue to believe that redesign of the survey methodology, underway since 1998, is necessary for CMS to fully respond to our past recommendation to improve the ability of surveys to effectively identify the existence and extent of deficiencies.¹⁴⁰

In short, GAO acknowledged that the redesign of the survey process has been underway for five years and, during that time that surveys have lacked the ability “to effectively identify the existence and extent of deficiencies.”¹⁴¹

GAO treats the problem as a sterile administrative issue and ignores the real-world consequences. There is no sense of urgency in GAO’s assessment. More to the point, GAO fails to acknowledge that for thirteen years, consumers have been relying on the survey to assure their health, safety, and welfare. Likewise, care givers and providers—including individual nursing home administrators and nurses—have had their performance evaluated and their professional status determined based on a process repeatedly determined to be ineffective and inaccurate. In other words, the Secretary still does not have a means of determining whether federal dollars are paying for quality care. Yet, for those “real people” who are most dependent on the survey process, the issue is not economic, it is personal.

GAO’s critique was far from singular. In 2003 OIG again reiterated:

Our review of the survey process reveals that states differ in how they determine both the number and type of deficiencies... As a result, we conclude that nursing home survey results are not always consistent among states, therefore limiting the comparability of the data. Further, we cannot conclude whether trends in deficiencies are due to deteriorating care, variations in the survey process, and/or increased enforcement.¹⁴²

Remarkably, although OIG alluded to “deteriorating care,” the number of deficiency citations alleging poor care and citations at the highest severity levels had both been dropping for several years.¹⁴³ There was, in fact, no

registered professional nurses, dietitians, sanitarians, engineers, licensed practical nurses or social worker.” 42 C.F.R. § 488.314(a)(2) (2014). Ultimately, however, the state decides “what constitutes a professional subject to CMS’s approval.” 42 C.F.R. § 488.314(a)(3). CMS’s position is that any trained surveyor can make any decision required by the survey protocol.

140. PREVALENCE OF SERIOUS PROBLEMS, *supra* note 131, at 38 n.50, 39.

141. *Id.* at 39.

142. U.S. OFFICE OF INSPECTOR GEN., OEI-02-01-00600, NURSING HOME DEFICIENCY TRENDS AND SURVEY AND CERTIFICATION PROCESS CONSISTENCY 7 (2003).

143. *Id.* at 10-11.

empirical evidence that care was “deteriorating.” To the contrary, the only basis for that assertion was speculation by CMS and other government agencies that their actions should have resulted in greater numbers of deficiencies.¹⁴⁴ The possibility that efforts by regulators and the industry were producing positive results seems to have received little serious consideration. Like GAO, OIG made no comment on the significance of the problems with the survey to nursing facility residents, on the propriety or legitimacy of continuing to use such survey results to sanction providers or with respect to whether the survey results could be used to distinguish between “good” and “poor” facilities in the interim.¹⁴⁵

A couple of years later, in 2005, GAO again emphasized problems with the nursing facility nursing process and with the results of that process:

Inconsistency in states’ surveys is demonstrated by... continued wide interstate variability in the proportion of homes found to have serious deficiencies. For example, in the most recent time period, one state found such deficiencies in about 6 percent of homes, whereas another state found them in about 54 percent of homes. . . . In addition, state surveyors continue to understate serious deficiencies as shown by the larger number of serious deficiencies identified in federal comparative surveys than in state surveys of the same home.¹⁴⁶

Once again, GAO failed to acknowledge that the state and federal surveyors were required to conduct surveys of the same facilities in the same timeframe and to use the same survey methodology and investigative protocol.

The obvious question is: on what basis can one conclude that a protocol, which produces inaccurate results when utilized by trained professionals (i.e., state) surveyors, produces accurate results when utilized by similarly trained, professional federal surveyors to assess the quality of care delivered to the same patients by the same caregivers in approximately the same timeframe? Arguably, those variations are indications of profound systemic defects in the methodology used to assess quality by both state and federal surveyors, and not necessarily to differences between individual surveyors. GAO simply assumes,

144. *Id.* at 2-3.

145. *Id.* at 2.

146. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-06-117, NURSING HOMES: DESPITE INCREASED OVERSIGHT, CHALLENGES REMAIN IN ENSURING HIGH-QUALITY CARE AND RESIDENT SAFETY 4 (2005) [hereinafter CHALLENGES REMAIN]. Below is the percentage of nursing homes identified as having deficiencies discovered during state survey processes:

<i>Percentage of Homes with Serious Deficiencies</i>	<i>Number of States</i>
More Than 20 Percent	15
10 to 20 Percent	26
Less than 10 Percent	10

Id. at 4.

again without explanation, that results reported by the federal survey teams were accurate and that the states' survey teams results were inaccurate. That may, or may not, be correct, but no evidence is proffered supporting the assumption.

In 2005, GAO also emphasized the point that IOM had made in 1986 regarding the importance of an accurate and reliable survey methodology:

Given the pivotal role played by surveys in helping to ensure nursing homes residents receive high-quality care, the development and implementation of a more rigorous survey methodology is one of the most important contributions one can make to addressing oversight weaknesses.¹⁴⁷

The irony of GAO's emphasis on the priority that ought to be given to assuring an accurate and reliable survey methodology is made poignant by the context in which the statement was made.

Specifically, CMS had been engaged in redesign of the survey methodology to address the numerous criticisms made by government studies since at least 1998, more than ten years after OBRA '87 had mandated use of a validated survey protocol. In 2005, however, CMS was not focused on fixing the survey process. Instead, the development of a valid survey protocol took a backseat to CMS' initiative to roll out its Nursing Home Compare website.

Despite the obvious need for the long overdue accurate and valid survey process—even ignoring that a valid survey protocol had been a statutory priority since 1987—no justification was ever offered by CMS for prioritizing the expenditure of time and capital on Nursing Home Compare. Indeed, CMS seemed to perceive no need to explain the source of its claimed authority to set priorities other than those dictated by Congress.

The Nursing Home Compare website—and the Five Star Quality Rating System added to Nursing Home Compare in 2008—were designed to provide consumers with information to allow comparison of the relative merits of one nursing facility versus another. Both Nursing Home Compare and the Five Star Quality Rating System rely to a significant extent on the very survey results that GAO, OIG, and others had criticized as unreliable and inconsistent.¹⁴⁸

147. *Id.* at 45.

148. See generally CHRISTIANNA WILLIAMS ET. AL., MEASURING NURSING HOME QUALITY – THE FIVE-STAR RATING SYSTEM (2010). The survey results were not the only questionable data utilized in CMS's Five-Star Rating calculation. GAO emphatically pointed out that the Quality Measures were not valid measures of quality and, indeed, were not derived from a valid quality measurement assessment: "[I]t is not clear whether a resident who triggers a QM . . . is actually receiving poor care. The lack of correlation among the QMs—a home may score well on some QMs and poorly on others—also calls into question their validity as measures of overall quality." U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-07-373, NURSING HOMES: FEDERAL ACTIONS NEED TO IMPROVE TARGETING AND EVALUATION OF ASSISTANCE BY QUALITY IMPROVEMENT ORGANIZATIONS 41-2 (2007) (footnote omitted). "The use of MDS data to measure the quality of

Thus, despite the long-standing recognition that the survey results were inaccurate and not clearly indicative of quality, CMS was engaged in a process to expand consumer reliance on survey results by individuals making critical decisions about where to obtain nursing facility care.

In fact, the “Technical User’s Guide” for the Nursing Home Compare Five-Star Quality Rating System states that “[t]he survey rating is the most important dimension in determining the [facility’s] overall rating.”¹⁴⁹ According to the individuals who designed the Nursing Home Compare Five-Star Quality Rating System, survey results were given disproportionately heavier weight in the Five-Star calculation because CMS wanted it that way:

CMS wanted the health inspection domain to play a predominant role in determining the overall 5-star rating. The main reason for this is that, unlike the staffing and quality measure domains, which are based on data self-reported by nursing homes, independent surveyors carry out the health inspection; thus, the data should be more objective and unbiased.¹⁵⁰

CMS, of course, knew that the data was neither “objective” nor “unbiased.” CMS had over ten years of government and government-sponsored reports to that effect beginning with the Preliminary Report issued by CMS’ contractor, Abt Associates in 1993.¹⁵¹

To be clear, CMS insisted that survey results play a “predominant role” in the ratings published on Nursing Home Care, and it urged consumers to use Nursing Home Care in making care decisions despite certain knowledge that survey results were inaccurate and that variation in survey results could not be attributed to quality difference between facilities. Indeed, those inaccuracies were so deeply embedded in the survey system that CMS had been engaged in redesigning the survey protocol for approximately seven years.¹⁵² The problems with the survey system were serious enough that CMS had spent almost five million dollars on the redesign in only the first four years of the eight-year project.¹⁵³ Moreover, rather than prioritize the development of a valid survey methodology, CMS insisted on moving forward and making the

care in nursing home is problematic because the MDS was not designed as a quality measurement tool and does not reflect advances in clinical practice.” *Id.* at 43.

149. CTRS. FOR MEDICARE AND MEDICAID SERVS., DESIGN FOR NURSING HOME COMPARE FIVE-STAR QUALITY RATING SYSTEM: TECHNICAL USER’S GUIDE 15 (2012).

150. See WILLIAMS ET AL., *supra* note 148 at 4.

151. See generally Administrative Record, *supra* note 122. See also FINAL REPORT, *supra* note 66, at 1 (discussing significant concerns with surveyor discretion); CMS REPORT, *supra* note 82, at 75 (published in 1998 and discussing “self-report bias” among nursing home administrators); OIG OEI-02-99-00060, *supra* note 91, at 4 (published in 1999 and discussing the HCFA’s reliance on the Online Survey Certification and Reporting System and its self-reported data).

152. CHALLENGES REMAIN, *supra* note 146, at 45 (2005).

153. *Id.* at 42.

Nursing Home Compare Website available to consumers despite GAO's objection that the data utilized in the website was misleading and inaccurate.¹⁵⁴

In 2006, CMS announced an improved survey process called the Quality Indicator Survey, or "QIS." CMS' memorandum introducing the QIS stated:

The CMS recognized the need to improve the consistency and effectiveness of the nursing home standard survey process and awarded a contract in 1998 to develop improvements to that process. . . . Improvements to the survey process are incorporated into fundamentally revised survey called the Quality Indicator Survey (QIS). The objectives of the QIS process include: Improve consistency and accuracy of quality of care and quality of life problem identification by using a more structured process. . . .¹⁵⁵

Thus, according to CMS, the new QIS would improve both consistency and accuracy of survey findings.

According to CMS, one innovation in the QIS Survey was the identification of "a set of Quality of Care Indicators (QCIs) specific to each surveyed facility, which are compared to national norms and used to identify care areas for systematic investigations."¹⁵⁶ Testing of the new QIS process was still underway. Nonetheless, CMS stated that "[a]lthough we have not received final results from the evaluation of the QIS, the preliminary information received is very positive and indicates that it is time to strategically move forward with plans to implement the QIS process in a estimated eight to ten . . . selected States[]" in addition to the five states where QIS already had been implemented.¹⁵⁷

154. See U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-03-187, NURSING HOMES: PUBLIC REPORTING OF QUALITY INDICATORS HAS MERIT, BUT NATIONAL IMPLEMENTATION IS PREMATURE 4 (2002). Notably GAO's objection to publication was not based on the well-documented inaccuracies and inconsistency in the survey findings that CMS made a dominant factor in the analysis. Instead, GAO questioned the accuracy of the data used to determine another element in the calculation, the quality measures. *Id.* at 3. GAO also questioned whether the quality indicators had been appropriately validated. *Id.* at 21.

155. SURVEY & CERTIFICATION GROUP, CTRS. FOR MEDICARE & MEDICAID SRVS., S&C-07-09, SOLICITATION OF STATE SURVEY AGENCY PARTICIPATION IN QUALITY INDICATOR SURVEY (QIS) IMPLEMENTATION 1-2 (2006).

156. *Id.* at 2.

157. *Id.* at 3. See also SURVEY & CERTIFICATION GROUP, CTRS. FOR MEDICARE & MEDICAID SRVS., S&C-07-09, FYI - RELEASE OF BROCHURE DESCRIBING THE QUALITY INDICATOR SURVEY (QIS) DEMONSTRATION PROJECT 1 (2005) ("Initial testing of the QIS process has revealed it yields increased consistency and improved documentation of survey findings"). Note that CMS made no claim that initial testing reported improved accuracy, despite that "improve[d] . . . accuracy of quality of care and quality of life problem identification" was one of the prime objectives of the redesign of the survey process. *Id.* See also ABT ASSOCIATES & VANDERBILT UNIVERSITY, EVALUATION OF THE QUALITY INDICATOR SURVEY (QIS): FINAL REPORT at ii (2007) [hereinafter QIS EVALUATION REPORT] ("Improved accuracy of quality of care and quality of life problem identification is one of the objectives of the QIS").

Although CMS characterized the QIS's reliance on quality indicators as a "fundamentally revised survey," the requirement to conduct the survey based on quality indicators was long-standing: IOM's 1986 Report had recommended and OBRA '87 required "a survey of the quality of care furnished, as measured by indicators of medical, nursing, and rehabilitative care, dietary and nutrition services, activities and social participation, and sanitation, infection control, and the physical environment."¹⁵⁸ OBRA '87 incorporated the recommendations requiring "a survey of the quality of care furnished, as measured by indicators of Medicaid nursing and rehabilitative care . . . among others."¹⁵⁹

It soon became apparent that CMS had acted and spoken too soon when promising improved accuracy and consistency. Significant problems with both accuracy and consistency were observed when the QIS process was fully tested.

The Final Report of the "Evaluation of the Quality Indicator Survey (QIS)" (QIS Evaluation Report) was published in December 2007.¹⁶⁰ The QIS Evaluation Report first described the purposes, objective, and methodology of the evaluation and then the results.

According to the QIS Evaluation Report:

The QIS is a revised long-term care survey process that was developed under the Centers for Medicare & Medicaid Services (CMS) oversight through a multi-year contract. It represents an effort to standardize how the survey process measures nursing home compliance with federal standards and the interpretative guidelines that define those standards. The QIS demonstration represents the culmination of 15 years of CMS-sponsored research and development aimed at addressing criticisms of long-term care survey process raised by consumers, providers, the General Accounting Office, Congress, survey agencies, and CMS Central Office.¹⁶¹

Among other things, "[i]mproved accuracy of quality of care and quality of life problem identification [was] one of the objectives of the QIS."¹⁶² QIS was

158. Compare IOM REPORT, *supra* note 6, at 34 ("[t]he standard survey should rely on 'key indicators' of quality of resident life and care . . ."), and *id.* at 113-14 ("[t]he standard survey would be designed to use 'key indicators' of performance to identify facilities with poor resident outcomes that might have resulted from substandard nursing home performance[]"), and *id.* at 116-17 (discussing use of key indicators), and *id.* at 378-88 (discussing key indicators and operational usage), with 42 U.S.C. § 1395i-3(g)(2)(A)(ii) (2012) (specifying content of standard survey and requiring use quality indicators), and 42 U.S.C. § 1396r(g)(2)(A)(ii) (2012) (also specifying content of standard survey and requiring use quality indicators).

159. 42 U.S.C. § 1396r(g)(2)(A)(ii)(I).

160. See QIS EVALUATION REPORT, *supra* note 157.

161. *Id.* at i.

162. *Id.* at ii.

“expected to yield more valid inferences about the care provided to residents and systems of care.”¹⁶³

The QIS Evaluation Report assessed whether the QIS was more accurate than the standard survey by comparing facility quality measured by QIS with quality as measured by findings made using the traditional standard survey protocol.¹⁶⁴ Evaluation researchers visited a sample of nursing facilities to collect data at the same time that the facility was undergoing a standard survey.¹⁶⁵ The Evaluation researchers worked independently of the surveyors to collect data regarding seventy-five “Care Indicators” (CIs) addressing five quality domains: incontinence, nutrition, pressure ulcers, choice, and activities.¹⁶⁶ In contrast, OBRA '87 requires use of quality indicators assessing thirteen quality domains: “medical, nursing, and rehabilitative care, dietary and nutrition services, activities and social participation, and sanitation, infection control, and the physical environment,” as well as the accuracy of resident assessments, adequacy of care plans, and resident rights.¹⁶⁷ After completion of data collection, the QIS Evaluation assessed:

How well the QIS achieved its primary objective of improving the accuracy, consistency, and documentation of the nursing home survey process within existing survey resources. The CIs addressed [five] research questions:

- *Does the QIS lead to increased accuracy? . . .* [The CIs] give measures of the quality of care provided by the nursing home that can be compared to surveyor findings to determine whether there is more agreement between surveyor findings and the quality indicators for QIS or Standard surveys. If the QIS is more accurate than the Standard survey, then we expect that there would be a higher level of agreement between QIS survey results and the quality indicators than for Standard surveys.
- *Does the QIS result in improved documentation of survey deficiencies? . . .* We used content analysis to compare the quality of documentation . . . for a sample of CMS 2567 [Survey Report] Forms from QIS and Standard surveys, using a blind review process to measure whether the documentation supports the specific F-tag, scope and severity that was cited.
- *How does the time required to complete the QIS compare to the time required for the current survey?* A key research question is whether the QIS requires more surveyor time than the current survey process. Using data from the CMS-670 form, we analyzed how QIS time compared to the time

163. *Id.*

164. *Id.* at iii.

165. *Id.* at ii, iv (explaining that the data was collected by reviewing the resident’s medical record, interviewing the resident and observing the resident in a variety of care-related activities).

166. *Id.* at ii.

167. 42 U.S.C. § 1395i-3(g)(2)(A)(ii)(I)-(II) (2012).

for the facility's prior survey and to Standard surveys at similar facilities. . . .

- *How does the QIS impact the number and types of deficiencies that are cited?* Analysis of survey deficiencies was used to examine the impact of the QIS on the number and scope/severity of deficiencies cited and also whether the QIS is associated with changes in the types of regulatory care issues that are cited.
- *Does the QIS improve surveyor efficiency?* One of the objectives of the QIS is to improve the efficiency of surveyors We analyze the relationship between time and outcomes to measure whether the QIS is associated with changes in how well surveyor time is targeted to facilities with more quality problems.

The answers to the research questions were not what CMS had seemed to promise when it asserted that “it [was] time to strategically move forward with plans to implement the QIS process.”¹⁶⁸

The QIS Evaluation Report concluded that, with respect to at least four of the five areas studied, the QIS was not an improvement over the standard survey.¹⁶⁹ The QIS Evaluation Report summarized the evaluation's conclusions separately for each of the five questions investigated:

- *Does the QIS lead to increased accuracy?* [W]e did not find evidence that the QIS was more accurate than the Standard survey. . . . Ultimately, under both types of surveys, there appears to be a great deal of surveyor discretion and judgment that influences the decision to cite.¹⁷⁰

168. CTRS. FOR MEDICARE & MEDICAID SERVS., SOLICITATION OF STATE SURVEY AGENCY PARTICIPATION IN QUALITY INDICATOR SURVEY (QIS) IMPLEMENTATION MEMORANDUM 3 (Dec. 21, 2006) [hereinafter CMS QIS MEMORANDUM]. Although CMS asserted in the QIS Memorandum that preliminary results were “very positive,” that is hard to believe in light of the final results. *Id.* at 116. Moreover, as discussed earlier, the roll out of Nursing Home Compare demonstrates CMS's willingness to take actions despite empirical evidence.

169. QIS EVALUATION REPORT, *supra* note 157, at 13-14 The conclusion section of the QIS Evaluation Report opens with the sentence: “The results of the evaluation were mixed and do not lead to firm conclusions about the effectiveness of the QIS.” That statement is belied by the Report's actual answers to each of the questions and by the Report's discussion of the implications of those answers. For example, with respect to the accuracy of the QIS compared to the standard survey, the QIS Evaluation Report is unequivocal: “We did not find evidence that the QIS is more accurate than the Standard Survey. . . .”

170. The specific findings and implications reported by the QIS Evaluation Report with respect to the accuracy of the QIS survey compared to the standard survey are excerpted below:
Results

“In general, we did not find any difference in accuracy for the QIS and Standard surveys.

- *The QIS and Standard survey samples were comparable with respect to overall quality and survey deficiencies cited.* The two groups were also similar with respect to the frequency of Ftag citations of a scope beyond isolated . . .

- *The overall failure rate on the CIs was high.* The overall fail rate across all care areas was 44 percent for the Standard sample and 45 percent for the QIS sample. . . . Across nursing homes, there was significant variability between CI fail rates for all care areas that provided the opportunity to judge how the two survey types discriminated quality with the Ftag citation system.
- *Overall, the relationship between quality and survey deficiencies was low.* We found a positive but low correlation between the overall number of related Ftags and quality as measured by the CI fail rate for both Standard and QIS surveyed nursing homes. Neither QIS nor Standard surveys consistently documented that providers failed to implement many of the care indicators recommended in the investigative protocols. Some recommended quality measures were never or rarely documented by either survey team in an Ftag statement despite the fact that the majority of residents who were eligible for the care described in the guidelines were found in this evaluation to not receive that care.
- *There was no evidence of a stronger relationship between quality and deficiencies for QIS surveys.* We found that there were no differences in the ability of the QIS and Standard surveys to detect quality problems . . .
- *Findings suggested that more survey deficiencies with scope greater than isolated could have been cited for both QIS and Standard surveys . . .* Both types of surveys failed to detect more than isolated problems in many facilities.
- *There was no evidence that the QIS was more accurate with respect to survey deficiencies with scope greater than isolated.* The CIs identified 29 occasions that would justify an Ftag greater than isolated across the Standard survey nursing homes and 31 opportunities for QIS nursing homes. The Standard Survey teams wrote greater than isolated Ftags in 15 out of these 29 opportunities (52 percent) vs. 11 out of the 31 opportunities (35 percent) for QIS surveys.
- *Both types of surveys failed to detect many residents with poor pressure ulcer and weight loss outcomes. . . .* Both types of surveys failed to detect many residents who have poor pressure ulcer and weight loss outcomes and who also receive poor care according to multiple data sources.”

Implications

“We did not find evidence that the QIS is more accurate than the Standard survey, despite the fact that it has started the process of making the survey process more specific and focused with its Stage I protocols and automated data entry system. . . . [W]e do not believe that a larger sample size would produce dramatically different results until further refinements are made in the basic concepts that underlie the QIS and which make it different from the Standard survey process . . . [W]e believe that the best explanation for the lack of differences between the two survey methods is related to two issues: 1) the specificity of the investigative guidelines and the critical element pathways, and 2) how feasible or “user friendly” the critical element pathways and interpretative guidelines are to implement. The various investigative documents used by both survey types vary in specificity such that there is much interpretation left to the discretionary judgment of surveyors. . . .”

Id.

- *Does the QIS result in improved documentation of survey deficiencies?* We found essentially no difference in documentation quality associated with the QIS¹⁷¹
- *How does the time required to complete the QIS compare to the time required for the current survey?* We found that the QIS took considerably longer to complete than Standard surveys in two of the five demonstration states; two states consumed about the same amount of time and one state's time was open to different interpretations.¹⁷²

171. The specific findings and implications reported by the QIS Evaluation Report with respect to documentation of survey findings in QIS surveys compared to standard survey are excerpted below:

Results

"[O]verall there was essentially no evidence that the QIS leads to higher quality deficiency documentation. Nor was there any evidence that the QIS led to an overall increase in the citation of related Ftags:

- *Standard surveys were more likely to include both a deficiency statement and a related outcome.* Overall, 33 percent of Standard survey Ftags reviewed [were] noted to include both a deficient statement and a related outcome, while 21 percent of QIS Ftags reviewed met this standard. QIS survey *process* deficiencies, such as assessment (F272) and care planning (F279) deficiencies, were more frequently accompanied by their related outcome tags than were Standard survey deficiencies, but, for outcome deficiencies (e.g., pressure ulcer development, F314), Standard surveys were more frequently accompanied by related process deficiencies than were QIS outcome deficiencies.
- *QIS deficiencies tended to cite more types of evidence than Standard deficiencies.* QIS Ftags reviewed cited as many or more types of evidence in general, for both process and outcome tags.
- *There was little difference with respect to the quantity of evidence cited.* . . . California and Ohio QIS surveys referenced a higher number of data points than their Standard survey counterparts, while fewer data points were cited on Louisiana QIS surveys.
- *There was no evidence that the QIS was associated with citation of additional related Ftags.* The review of [Critical Element] Pathways [guiding QIS surveyors' reviews] did not reveal significant differences between the "related Ftags" cited on QIS vs. Standard surveys. . .

Implications

"While there is no concrete evidence that a reasonable level of inter-rater reliability was achieved prior to the review nor sustained during the review, the reviewers were experienced surveyors who participated in the development of the review protocol and guidance. The lack of a systematic difference[] in documentation quality may reflect the variable knowledge and skill of the surveyors under both the QIS and Standard survey, which likely influences both the decision to cite and the supporting documentation[.]"

Id.

172. *Id.* at 13. The specific findings and implications reported by the QIS Evaluation Report with respect to the "major evaluation question" (because implementation of QIS must be resource neutral) of how the time to complete a QIS compares to the time required to complete the current survey are excerpted below:

- *How does the QIS impact the number and types of deficiencies that are cited?* The results of this evaluation clearly indicated that the QIS cites more deficiencies, at higher levels, and more in these usually under-cited areas.¹⁷³

Results

“Results varied across States. . .

- For three states (California, Kansas, Ohio), the QIS took longer than the prior Standard survey at the same facilities – the pre-post differences were especially large in California and Kansas. . .
- In Connecticut and Louisiana, the QIS was completed more quickly than the prior survey at the facility. . .
- Exclusion of outliers does not change basic conclusions regarding QIS completion time. There were some QIS surveys that took an extraordinarily long time to complete, in some cases 200 or more hours. All but one of these surveys was in California and Kansas. The explanation for the high completion times varied. . .
- The changes to the QIS implemented after the formative evaluation reduced QIS time requirements. The changes to the QIS implemented after the formative evaluation appeared to lead to modest reductions in the time required to complete QIS surveys.”

Id. at 8–9.

Implications

“. . . [W]hile we do have conclusions about how long QIS surveys took to complete in each of the demonstration states, we do not offer conclusions about the time requirements of the QIS in other states. The likelihood is that there will be some states for which the QIS does not take any longer to complete than Standard surveys and others that struggle to implement the QIS and find that it does take longer.”

Id. at 10.

173. *Id.* at 14. The specific findings and implications reported by the QIS Evaluation Report with respect to how the QIS impacted the numbers and types of deficiencies cited are excerpted below:

“Results

- *The QIS was associated with an increase in the number of survey deficiencies:* . . . [W]e estimate that the QIS was associated with 1.6 additional deficiencies in California (a 14 percent increase), 0.6 fewer deficiencies in Connecticut (a 9 percent decrease), 9.4 additional deficiencies in Kansas (a 99 percent increase), 1.9 additional deficiencies in Louisiana (a 29 percent increase), and 2.4 additional deficiencies in Ohio (a 52 percent increase).
- *The QIS was associated with an increase in G-level deficiencies:* We also examined the impact of the QIS on deficiencies cited at the G-level or above (G, H, I, J, K, L). The rate of G-level deficiencies was relatively small for both types of surveys, but the QIS was associated with large increases in Kansas, and Ohio and a large decline in Connecticut. There was a slight increase in California and relatively little change in Louisiana.
- *The QIS was associated with an increase in the regulatory care areas cited:* One of the objectives of the QIS is to comprehensively review a wide range of regulatory care areas. In all five states, there were more regulatory care areas cited on QIS survey than on the prior survey at the facility. For all

- *Does the QIS improve surveyor efficiency?* Ohio was the only state for which the QIS was associated with an increase in surveyor efficiency.¹⁷⁴

To summarize, after almost ten years and millions of dollars had been invested in developing a new survey methodology, the QIS survey did not result in more accurate and valid survey determinations than the traditional survey. Further, the QIS did not improve the survey process in most other areas studied compared to the current standard survey.

Remarkably, without acknowledging the overlap, the QIS Evaluation Report concluded that “the best explanation for the lack of differences [in accuracy and validity] between the two survey methods” lay in three problems—all of which Abt had identified as systemic defects in the original survey methodology in its 1996 Final Report.¹⁷⁵ First, the “investigative

regulatory care areas except for infection, facilities were more likely to receive a deficiency with the QIS survey than with their prior survey. For some regulatory care areas (resident rights, quality of life, dietary care, physician services, dental care, physical environment), the differences were substantial.”

Id. at 11.

Implications

“...The analysis provides strong support for the hypothesis that the QIS leads to an increase in the number of survey deficiencies and an increase in the regulatory care areas that surveyors cite, supporting expectations about the QIS. These are an important finding given the studies by the General Accounting Office (GAO) and Office of the Inspector General (OIG) that have found that the Standard survey under reports deficiencies, harm-level deficiencies, quality of life, resident rights, and dental deficiencies. As a practical matter it would be difficult to implement any system that results in several fold increases in deficiencies, but this is not in general the case although the increase observed for Kansas QIS surveys may be a reason for concern.

“A potential limitation of this analysis is that we are unable to control for surveyor quality. QIS surveyors were chosen to participate in the demonstration because of their experience. It may be that the QIS teams have higher citation rates than other survey teams, and that this may explain some of the increase in survey deficiencies that we observed for QIS surveys.”

Id. at 12.

174. *Id.* at 14. The specific findings and implications reported by the QIS Evaluation Report with respect to whether the QIS improved surveyor efficiency compared to surveyor performance when conducting the standard survey are excerpted below:

Results

“...[T]he only state in which the QIS was associated with an increase in surveyor efficiency (as measured by the relationship between time and survey outcomes) was Ohio. . . The patterns that we observed on Ohio suggest that the QIS has the potential to improve surveyor targeting to facilities with the most quality problems, but the experience of the other states suggests that this need not be the case.”

Id. at 13.

175. *Id.*, at 5; *see also* FINAL REPORT., *supra* note 66, at 222–23.

guidelines and critical element pathways” intended to guide surveyor decision-making were not sufficiently specific. Second, the guidelines and pathways were difficult to implement. Finally, and as a result “there is much interpretation left to the discretionary judgment of the surveyors.”¹⁷⁶ The unvarnished truth was that, despite being repeatedly told that the survey guidelines lacked specificity, were difficult to apply and left complex outcome-related decisions to the discretion and “professional judgment” of surveyors who lacked the ability to make such decisions, CMS officials later admitted that “surveyor guidelines on identifying deficiencies were not changed due to the implementation of the QIS.”¹⁷⁷

Although CMS relied on the Evaluation—albeit the Evaluation’s preliminary results—as a basis to move forward with implementation of QIS and away from the standard survey, the final QIS Evaluation Report actually rejected CMS’ approach. Rather than supporting a change in survey processes, the QIS Evaluation Report stated that its findings were “limited by a small sample size; thus the data we provide are best used for survey improvement purposes rather than to inform a decision about what type of survey process to use.”¹⁷⁸ Just as it had ignored GAO and proceeded with Nursing Home Compare despite its reliance on flawed data, CMS proceeded to implement QIS notwithstanding the flaws identified in the QIS Evaluation Report.

Despite having previously acknowledged the importance of accurate and consistent survey results and despite having assured that the QIS process would improve both consistency and accuracy, CMS reconsidered and retreated following publication of the QIS Evaluation Report in December 2007. In 2008, CMS told GAO that:

The QIS is not likely to increase the accuracy of deficiency identification. . . .
But we do expect the QIS will increase the consistency of survey process both between States and within States. . . .¹⁷⁹

CMS’ underlying assumption, of course, was directly contrary to what CMS had previously said: that accuracy and consistency of survey results both were important, but accuracy was the overarching goal. Now, however, CMS claimed that improved consistency in survey results made the process significantly better even if there was no improvement in accuracy of results.

176. QIS EVALUATION REPORT, *supra* note 157, at v.

177. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-11-403R, NURSING HOME QUALITY: IMPLEMENTATION OF THE QUALITY INDICATOR SURVEY 19 (2011) [hereinafter SURVEY IMPLEMENTATION].

178. QIS EVALUATION REPORT, *supra* note 157, at v.

179. Letter from Vincent J. Vehtimiglia, Jr. to John Dicken (May 2, 2008), in U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-08-517, NURSING HOMES: FEDERAL MONITORING SURVEY DEMONSTRATE CONTINUED SERIOUS CARE PROBLEMS AND CMS OVERSIGHT WEAKNESSES (2011) [hereinafter OVERSIGHT WEAKNESSES].

The following year GAO rejected CMS' assumption that improving consistency without improving accuracy was an acceptable outcome. To the contrary, GAO reiterated that consistency might be important, but the QIS must be accurate:

[T]he evaluation [of the QIS methodology] did not find that QIS methodology increased survey accuracy, noting that QIS and traditional survey samples were comparable in overall quality and in the frequency of standards cited for deficiencies with either a pattern or widespread scope. . . . While improving the consistency of the survey process is important, CMS must also focus on addressing the accuracy of QIS surveys.¹⁸⁰

GAO continued, emphasizing that:

Survey methodology and guidance are integral to reliable nursing home surveys, and we found that weaknesses in these areas were linked to understatement . . . surveyors noted that surveys were too lengthy, complex, and subjective.¹⁸¹

GAO might be entitled to claim credit for the insight if it were not for the fact that the exact same criticisms of the survey process had been leveled years prior to the adoption of OBRA '87.¹⁸²

The OBRA '87 mandate that surveys be conducted using a protocol that had been designed, tested, and validated prior to use was specifically directed to eliminate one of the major causes of inconsistent and inaccurate survey performance. Moreover, GAO exhibits the same myopia that increasingly affects those evaluating the survey process: GAO asserts that problems in the survey process result in understatement of quality problems, but ignores the possibility that systemic methodological defects also result in overstatement of deficiencies. An accurate survey process must both "identify deficiencies" and "[c]ite only valid deficiencies."¹⁸³

Indeed, at the same time that GAO, OIG, and CMS' consultants were focusing on under-citation, experience strongly suggested that the systemic issues were pervasive because false-positive citations were evident at the individual facility and state level. By law, nursing facilities issued deficiencies during a survey have a right to ask state survey agencies to reconsider such citations.¹⁸⁴ The process, known as Informal Dispute Resolution (IDR) can take various forms. However, at the first level, the IDR is non-adversarial. Nevertheless, government studies of the results of the IDR process found that a

180. SURVEY IMPLEMENTATION, *supra* note 177, at 19.

181. *Id.*

182. See, Patricia A. Butler, *Assuring the Quality of Care in Nursing Homes*, 57 N. CAR. L. REV. 1318, 1327-337 (1979); Robert N. Brown, *An Appraisal of the Nursing Home Enforcement Process*, 17 ARIZ. L. REV. 304, 314-20, 324-38 (1975).

183. 42 C.F.R. § 488.318 (2014).

184. 42 C.F.R. § 488.331 (2014).

significant numbers of deficiency citations were either eliminated or modified at the first level review.

Specifically, a 2009 GAO study found:

16 IDRs were requested per 100 homes in fiscal year 2007, with this number ranging among states from 0 to 57 per 100 homes. For IDRs occurring in fiscal year 2007, 20 percent of disputed deficiencies were deleted and 7 percent were downgraded in scope or severity, but in 4 states at least 40 percent of disputed deficiencies were deleted through this process.¹⁸⁵

GAO's findings were generally consistent with the findings of a 2005 analysis performed by the OIG.¹⁸⁶ OIG found that forty-five percent of 1,211 disputed deficiencies issued in fourteen states during 2002 were changed as a result of IDR.¹⁸⁷ Those changes included modification of the written citation (nineteen percent), deletion of the citation (nineteen percent), and decrease in scope or severity level (six percent).¹⁸⁸

Thus, if nursing facilities were willing to challenge the deficiencies issued in the survey process—and many facilities do not pursue such challenges for reasons unrelated to their perception of the validity of the citation—between twenty-five and forty percent of such deficiencies were either eliminated, modified or downgraded.¹⁸⁹ It bears emphasis that deficiencies are eliminated or downgraded in significant numbers: (1) in a largely non-adversarial process; and (2) generally by the state survey agency that issued the citation in the first place. The obvious question is whether an average error rate of approximately twenty-five to forty percent would be acceptable—or permitted—in any other law enforcement environment. There is irony in that, despite findings

185. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-10-70, NURSING HOMES: ADDRESSING THE FACTORS UNDERLYING UNDERSTATEMENT OF SERIOUS CARE PROBLEMS REQUIRES SUSTAINED CMS & STATE COMMITMENT 42-43 (2009).

186. See OFFICE OF INSPECTOR GEN., U.S. DEPT. OF HEALTH & HUMAN SERVS., OEI-06-02-00750, INFORMAL DISPUTE RESOLUTION FOR NURSING FACILITIES (2005) [hereinafter INFORMAL DISPUTE RESOLUTION].

187. *Id.* at 12.

188. *Id.* See Dana B. Mukamel et al., *Nursing Homes Appeals of Deficiencies: The Informal Dispute Resolution Process* 13 J. AM. MED. DIRS. ASSOC. 512, 514 (2012) (finding that in the period 2005-2008, 9.8 percent of all annual surveys and 9.1 percent of all complaints resulted in an IDR request but that there was large variability across states in the rate of which IDR was requested with 30 percent of surveys resulting in an IDR request in some states).

CMS also has reported: "The quality and/or level of deficiency review may also affect the deficiencies cited by various States. . . . Some States have an intense deficiency review and carefully scrutinize each deficiency. The percentage of deficiencies rewritten as a result of the State survey agency review process ranged from 1 percent to 90 percent. Most States noted that this was a guess. Forty-four States reported the percentage of deficiencies that were withdrawn by their review mechanism in calendar year 1996. The percentages ranged from 1 percent to 25 percent." CMS REPORT, *supra* note 82, at 551-52.

189. INFORMAL DISPUTE RESOLUTION., *supra* note 186, at iii, 22.

suggesting that significant over-citation might be a problem as well as an understatement, GAO titled its report “Nursing Homes: Addressing the Factors Underlying *Understatement* of Serious Care Problems Requires Sustained CMS and State Commitment.”

Some in Congress also began to notice. In 2010, a bill titled “Improving the Quality of Care in Nursing Home Act of 2010” was introduced in Congress. Although the bill was never adopted, the findings section of the proposed legislation is instructive:

Sec 2 Findings: Since the enactment of the Omnibus Budget Reconciliation Act of 1987, there has been little systematic evaluation or review of the effectiveness of the survey and certification system in measuring and improving the quality of care for nursing home residents as well as ensuring compliance with the requirements of participation by nursing homes participating in Medicare or Medicaid.¹⁹⁰

The bill’s assertions regarding the failure to evaluate the survey system—despite OBRA '87’s mandate to validate the protocol and for the Secretary and the states to implement programs to improve consistency—is, to say the least, ironic.

That irony was made even more poignant by GAO’s April 2010 update¹⁹¹ of its 2008 study, “Nursing Homes: Federal Monitoring Surveys Demonstrate Continued Understatement of Serious Care Problems and CMS Oversight Weaknesses.”¹⁹² Although the update, like the original study, assumes that differences in the results of state and federal surveys were attributable to poor performance by state surveyors and that understatement of deficiencies¹⁹³ was the only problem with the survey process, the data on which the update relies confirms continued, profoundly serious and significant inconsistencies in survey results. Indeed, the data highlights that the inconsistencies were not merely between federal and state survey results, but also prevalent between states. Moreover, GAO states that the range of fluctuation in the results over several years makes “the longer-term trend unclear.”¹⁹⁴

For example, GAO’s 2010 update reported that, when federal government inspectors visited the same nursing facilities recently inspected by state inspectors and conducted a second “comparative” assessment of the quality of

190. Improving the Quality of Care in Nursing Homes Act of 2010, S. 3407, 111th Cong. § 2(3) (2010).

191. See Letter from John E. Dicken, Director of Health Care, United States Gov’t Accountability Office, to Senators Herb Kohl and Charles E. Greasley, GAO-10-434R, Nursing Homes: Some Improvement Seen in Understatement of Serious Deficiencies, but Implications for the Longer-Term Trend Are Unclear (Apr. 28, 2010) [hereinafter “GAO Update”].

192. See generally OVERSIGHT WEAKNESSES, *supra* note 179.

193. Defined by GAO as: (1) failure to cite problems; and (2) understatement of scope or severity. GAO Update, *supra* note 191, at 1.

194. *Id.* at 2.

care, the percentage differences in serious deficiency violations found by federal versus state inspectors nationally between 2002 and 2008 fluctuated “from as low as 11.1 percent to as high as 17.5 percent . . .” with 12.3 percent of the state surveys failing to cite at least one serious deficiency identified by the federal comparative survey.¹⁹⁵ Moreover, differences with respect to less serious deficiencies existed, on average, in about seventy percent of the federal surveys in the period.¹⁹⁶ In 2008, however, differences in less serious deficiency citations were observed in 74.8 percent of the state inspections compared with the federal comparative survey.¹⁹⁷

A wide range of variation in the alleged under-citation rates between individual states also is evident in the data, for both serious and less serious deficiencies. The range of variation in the federal and state inspections with respect to serious deficiency citations between 2002-2008 was reported as running from a low of zero in six states (Oregon, Alaska, Idaho, North Dakota, Maine and Vermont) to highs of twenty-eight to thirty-five percent in three states (Missouri, South Carolina and South Dakota).¹⁹⁸ Nineteen states failed to cite serious deficiencies allegedly identified by the federal surveyors in less than ten percent of the state surveys, while thirteen states failed to cite such deficiencies in twenty percent or more of the surveys.¹⁹⁹ With respect to less serious deficiencies in the same period, the state inspections cited fewer such deficiencies, on the low end, in twenty percent of the surveys in West Virginia, 38.5 percent of the surveys in Alaska, and 30.8 percent of the surveys in Wisconsin, to a high of 100 percent of the inspections conducted in Montana, South Dakota, and Utah.²⁰⁰ Twenty-one states cited fewer less serious deficiencies than the comparative federal survey in seventy-five or greater percent of their surveys.²⁰¹

Discussing the comparability of federal and state survey results, GAO noted that:

By statute, comparative [federal] surveys must be conducted within 2 months of the completion of the state survey. However, differences in timing, selection of residents for the survey sample, and staffing can make analysis of differences between the state and federal comparative surveys difficult. On the basis of our prior recommendations, CMS now calls for the length of time between the state and federal surveys to be between 10 and 30 working days and requires federal surveyors conducting a comparative survey to include at least half of the state survey’s sample of residents from that nursing home in

195. *Id.*

196. *Id.* at 3.

197. *Id.* at 2.

198. GAO Update, *supra* note 190, at 10-11.

199. *Id.*

200. *Id.* at 14-15.

201. *Id.*

the comparative survey sample, making it easier to determine whether state surveyors missed a deficiency. Furthermore, federal comparative survey teams are expected to mimic the number of staff assigned to the state survey.²⁰²

GAO did not mention, however, that federal and state surveyors receive the same training, that the state surveyors are equally qualified professionals, and that the same survey procedures, forms, and protocols used by the state surveyors must, by law, be used to conduct the federal comparative surveys.²⁰³

It is, of course, possible that the variation in the state survey results and the results of federal comparative surveys exists despite that the federal and state surveyors have the same training and qualifications, survey the same facilities, conduct the survey within days of one another, assess the care provided by roughly the same care givers to roughly the same residents,²⁰⁴ and utilize the same survey procedures, forms, and protocols to conduct the survey due to failings on the part of state surveyors and that the federal surveyors are the so-called “gold standard” of nursing facility survey performance. There is no evidence that is the case, however. Intuitively, one could justifiably wonder if the range of discrepancies between federal and state survey results suggests that it is the federal survey results that are problematic. Indeed, superficially, the state survey results seem to align with one another better than they do with the results of the federal comparative surveys. Alternatively, an objective observer might legitimately ask if the real cause of the variation is a systemic problem with the survey system’s assessment tools that assures that survey findings are not reproducible.

In a February 2012 report, GAO again made several points regarding the implementation of the QIS. GAO noted that CMS had developed the QIS to address weaknesses—including problems with accurate assessment of quality—in the traditional survey:

According to CMS, the QIS was developed to help the agency meet several objectives, including improving surveyors’ documentation of quality concerns; improving the efficiency of the survey process by focusing resources on facilities, and on areas within facilities, with the greatest quality concerns; and improving the accuracy and consistency with which surveyors identify deficiencies.²⁰⁵

202. *Id.* at 4.

203. 42 U.S.C. § 1395i-3(g)(2)(D) (2012).

204. Average length of stay for short-term residents (approximately fifty percent of residents) is approximately 23 days, for current residents, approximately 835 days, and for discharged residents approximately 270 days. *The National Nursing Home Survey: 2004 Overview*, 13 CTR. FOR DISEASE CONTROL & PREVENTION, June 2009, at 4, available at www.cdc.gov/nchs/data/series/sr_13/sr_13_167.pdf.

205. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-12-214, NURSING HOME QUALITY: CMS SHOULD IMPROVE EFFORTS TO MONITOR IMPLEMENTATION OF THE QUALITY INDICATOR SURVEY 2 (2012) [hereinafter QUALITY INDICATOR SURVEY].

GAO then reviewed the results of the 2007 QIS Evaluation concluding:

In terms of whether the QIS led to improved accuracy or improved documentation of deficiencies, the report found that the QIS-based routine survey and the traditional survey approach generally led to comparable results. The researchers, therefore, concluded that the QIS did not significantly enhance or diminish surveyors' ability to accurately identify or document deficiencies. Similarly, the study suggested that in general, surveyors using the QIS tool completed routine surveys as efficiently, as surveyors using the traditional survey methodology.²⁰⁶

GAO also noted that in 2009 CMS had commissioned a third study to evaluate the QIS. That study, completed in 2011, "was intended to identify aspects of the QIS that could affect how consistently surveyors identify quality problems" ²⁰⁷ According to GAO:

The study found that various aspects of the QIS process, including the way resident interviews were conducted and the surveyor initiation process (in which surveyors identify additional problem areas, not identified by the QIS software, for further investigation during surveys), might affect the consistency with which surveyors identify quality problems during nursing home inspections."²⁰⁸

Stated more directly, the absence of specific guidelines governing surveyor performance required the exercise of surveyor discretion to make important decisions regarding how quality was assessed in the survey process—once again, precisely the same issues Abt had originally raised in its 1993 and 1996 reports.

GAO noted that CMS had begun to implement the QIS nationally following the 2007 QIS Evaluation Report. CMS "told [GAO that] they worked to improve aspects of the QIS process by taking steps to address the study's findings and recommendations."²⁰⁹ "CMS officials" likewise told GAO "that they intend to consider the recommendations made in [the third] study [completed in 2011] in their efforts to continue improving the QIS."²¹⁰

The remainder of GAO's report makes clear that CMS had not taken the steps necessary to keep those commitments. Noting that whether the QIS is meeting its objectives requires performance monitoring—"a key component of federal internal control standards"—GAO stated that CMS "acknowledged the need for certain performance measures—specifically with regard to examining the effect of the QIS on surveyor consistency—but noted that they did not have

206. *Id.* at 8.

207. *Id.* at 8-9.

208. *Id.* at 9.

209. *Id.* at 8.

210. QUALITY INDICATOR SURVEY, *supra* note 205, at 9.

performance goals or measure in place as of December 2011.”²¹¹ GAO summarized:

[T]he agency has not established the means—such as performance goals and measures—that would allow for routine, ongoing monitoring of the extent to which the QIS is helping CMS improve the survey process as intended While studies commissioned at key points can be valuable, the information CMS would obtain on the QIS through routine monitoring is needed for this ongoing effort.²¹²

Even more remarkable, GAO found that CMS had no systemic process of assessing the accuracy and consistency of QIS results despite the requirements, included in the Medicare and Medicaid Acts since 1987, that:

Each State and the Secretary shall implement programs to measure and reduce inconsistency in the application of survey results among surveyors.²¹³

211. *Id.* at 9-10. In classic understatement, “performance goals and measures can affect CMS’s ability to effectively monitor the OIS.” *Id.* at 12.

212. *Id.* at 20-21.

213. 42 U.S.C. § 1395i-3 (g)(2)(D) (2012). It bears some emphasis that when government studies refer to inconsistency of survey results, the reference is to state-to-state variation. There is no government study that examines whether inconsistency is a problem between surveys in the same state, i.e., intra-state inconsistency. One non-government study found substantial intra-state variations:

“Variation also exists within states. For example, the state of Kansas is composed of 6 survey regions. In 2001 facilities in the Northeast Region averaged 11.64 deficiencies, nearly three times as many as facilities in the West Region (3.69 deficiencies) . . . Administrators and directors of nursing tended to think this heterogeneity reflected differences in the survey process; surveyors thought it reflected differences in facility characteristics.”

Robert H. Lee et al., *Reliability of the Nursing Home Survey Process: A Simultaneous Survey Approach*, 46 GERONTOLOGIST 772, 772 (2006). The same study also reported substantial differences between two different survey teams: looking at the same evidence:

“The variability interpretation notes that the two survey teams often reached different conclusions about whether a deficiency existed, what regulation had been breached, the scope of the deficiency, or the severity of the deficiency. These differences, furthermore, might well have consequences. The penalties imposed by the survey agency, the career prospects of facility managers, and the response of consumers are likely to be different for a nursing home that gets 7 D deficiencies than for a nursing home that gets 12 D deficiencies and 1 G.

“The same process can draw no deficiencies from one survey team and multiple deficiencies from another. As a result, nursing home administrators and directors of nursing cannot be confident that a good survey means that a process works well. . . Improvement efforts are inhibited by a survey process that falls short of systematic, replaceable data gathering and analysis.

“The variability of the survey also reduces its value to regulator and policy makers. . . The many disagreements of these two teams about whether a regulation has been breached, which regulation had been breached, and how serious the breach was cannot make federal or state officials comfortable.

In a March 2012 report, GAO again reiterated that there was no significant difference between the results of the QIS process and the standard survey originally adopted in the early 1990s to implement OBRA '87.²¹⁴ According to GAO, following introduction of QIS in some states, CMS considered whether to differentiate in its Nursing Home Compare Five-Star Quality Rating System calculation between nursing facilities surveyed by QIS and those inspected using the traditional methodology.²¹⁵ CMS decided to make no distinction because:

CMS examined whether the health inspection rating of facilities that were inspected using the quality indicator survey differed from those that were inspected with the traditional . . . survey. No significant difference was found between the two approaches in conducting the survey. . . .²¹⁶

GAO noted that CMS justified use of the survey data in the Five-Star Quality Rating calculation claiming, “the strengths of the surveys are that they are conducted by trained individuals and follow national standards.”²¹⁷ Of course, years of GAO studies demonstrate that, training and national standards notwithstanding, twenty-five years after the adoption of OBRA '87 and the mandate that the survey protocol was to be evidence-based, accurate and reliable, those goals remain aspirational.

IV. BACK TO THE FUTURE AND AT THE END OF THE DAY

A. *Déjà Vu All Over Again*

History gives those who hope to see CMS develop a survey protocol that validly and accurately assesses quality little reason to be optimistic. The need to revamp the survey process so that it validly and accurately assesses quality of nursing facility care has been recognized for decades, long before OBRA '87, and efforts to develop such an assessment tool have failed repeatedly. For years, deep-seated political, methodological, philosophic, and other conflicts

This article suggests that surveyors need more specific criteria, in the form of decision-making algorithms, to reduce the influence of individual perceptions. These findings concur with other valuations of survey consistency (GAO, 2003b, Office of the Inspector General, 2003, 2004).

Id. at 778.

214. QUALITY INDICATOR SURVEY, *supra* note 204, at 1; *See generally* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-12-390, NURSING HOMES: CMS NEEDS MILESTONES & TIMELINES TO ENSURE GOALS FOR THE FIVE-STAR QUALITY RATING SYS. ARE MET (2012) [hereinafter CMS MILESTONES].

215. QUALITY INDICATOR SURVEY, *supra* note 205, at 16.

216. CMS MILESTONES, *supra* note 214, at 16.

217. *Id.* at 6. (“The primary goal of the Five-Star System is to help consumers make informed decisions about their care by providing understandable and useful information on nursing home quality.”)

among those most directly affected have precluded development of an effective quality assessment tool.²¹⁸

Development of a valid and accurate survey protocol has been a “third rail,” so politically sensitive that it has been unapproachable and certainly untouchable. As a result, various interest groups tolerate a survey process that, almost all agree, is ineffective because reform is so unlikely that commitment of the required resources and political capital would be the paradigm of throwing good money after bad. In the meantime, CMS tinkers with the Nursing Home Compare Five-Star Quality Rating System, continuing to assert that the System relies on objective and valid survey data²¹⁹ and announces initiatives to tie Medicare payment to quality of care,²²⁰ despite that it cannot effectively measure quality, at least in nursing facilities.

To be sure, in the early years after the adoption of the Medicare and Medicaid Acts, there were serious obstacles impeding the design and development of an effective means of assessing and enforcing quality in nursing facilities.²²¹ As a federal official directly involved in the effort explained around the time that OBRA '87 was adopted:

During the early part of the decade, congressional hearings, research, and discussion contributed to a growing body of concern and knowledge about the nature and regulation of nursing homes. Concurrently, the Federal Government was attempting to develop new sets of regulations. Two particular efforts are worthy of note: the attempt in 1980 to issue regulations that contained new standards for facilities to meet and the attempt in 1982 to issue new enforcement regulations.

[A] detailed explanation of all of the provisions of these regulatory proposals is of little interest. The significance of these proposals is in their failure. They proved that the Federal Government, for a myriad of reasons, was unable to establish sufficient consensus to reregulate [sic] the nursing home industry. The various conflicting forces, including consumer, industry, and State government, made it, in effect, impossible for the Federal Government to establish a reasonable, respectable set of proposed regulations.²²²

More bluntly, the needs of nursing facility residents took a backseat to CMS' institutional political considerations and to the agency's inability—or unwillingness—to lead from the front.

218. See Morford, *supra* note 8, at 129 -130.

219. See, e.g., Ctrs. for Medicare & Medicaid Servs., Fact Sheet: Nursing Home Compare 3.0: Revisions to the Nursing Home Compare 5-Star Rating System (Feb. 12, 2015).

220. See Sylvia Matthews Burwell, *Progress Towards Better Care, Smarter Spending, Healthier People* (Jan.26, 2015) <http://www.hhs.gov/blog/2015/01/26/progress-towards-better-care-smarter-spending-healthier-people.html>.

221. See, e.g., Morford, *supra* note 8, at 129.

222. See *Id.* at 129 -130.

Some nursing facility residents, however, had grown tired of CMS playing political “rope-a-dope” with the survey process. Federal litigation challenging the legality of the nursing facility inspection process was instituted in the early 1980s, ultimately leading to several court orders that CMS develop and implement a survey process that accurately assessed the quality of nursing facility care.

In what might, in retrospect, be seen as a prequel to the last couple of decades, CMS repeatedly refused to comply with the court orders. Indeed, the agency’s persistent refusal to develop and implement a valid nursing facility quality assessment tool led to the Secretary of HHS being held in contempt of court.²²³ The case is a cautionary tale for anyone who hopes that CMS will voluntarily fulfill its obligations to test and validate any time soon.

Initially, the District Court dismissed the case, holding that it lacked authority to compel defendants to develop and implement a new inspection methodology.²²⁴ In 1984, the Tenth Circuit Court of Appeals reversed.

Based upon the Court’s detailed review of the Medicaid Act as a whole, the Tenth Circuit held that the Act imposed a duty on the Secretary of HHS “to promulgate regulations which will enable her to be informed as to whether facilities receiving federal Medicaid funds are actually providing high quality medical care.”²²⁵ The Tenth Circuit further held that the duty to develop and utilize a survey system that validly assessed the actual quality of nursing home care was so clear and unequivocal that mandamus was the appropriate remedy.²²⁶

In 1985, the District Court, on remand, ordered CMS to use the formal notice and comment rulemaking procedures provided by the Administrative Procedure Act (APA) to adopt a survey process that assessed the quality of the care actually provided.²²⁷ On June 13, 1986, CMS promulgated the first of a series of “final” regulations.²²⁸ The new Patient Care and Services (PaCS) survey was described as outcome-oriented because compliance with federal standards was determined by the “facility’s success in providing . . . care.”²²⁹

The *Smith* plaintiffs challenged the “final” survey rule, arguing that CMS had not complied with the Tenth Circuit’s mandate or with the District Court’s

223. See *Quality of Long-Term Care: Hearing on S. 1108 Before the Senate Subcomm. on Health of the Comm. on Finance*, 100th Cong. 51, 52 (1987) [hereinafter *Hearing on S. 1108*]; see also *Estate of Smith v. Heckler*, 557 F. Supp. 289 (D. Colo. 1983), *rev’d*, 747 F.2d 583, 591 (10th Cir. 1984), *on remand* 622 F. Supp. 403 (D. Colo. 1987).

224. *Heckler*, 557 F. Supp. at 299.

225. *Heckler*, 747 F. 2d at 591.

226. *Id.*

227. *Heckler*, 622 F.Supp. at 412.

228. See 51 Fed. Reg. 21,550 (June 13, 1986).

229. See, e.g., 59 Fed. Reg. at 56,231.

order to publish a valid survey protocol in binding regulations.²³⁰ Because the regulations did not include “the guidelines and forms which constitute the [survey] system,” or the “details of the [survey] methodology,” the District Court agreed and held that defendants had not complied, substantively or procedurally, with the Tenth Circuit’s mandate.²³¹ The Court held that the Secretary’s refusal to be bound by specific, defined survey procedures, guidelines and forms was a dereliction of the duty imposed by the Medicaid Act.²³²

The District Court further found:

There is no legislative definition of ‘quality health care,’ and there can be none. It is something which emanates from the process of regulation. The methodology prescribed is the vehicle by which the Secretary will become ‘informed as to whether the nursing facilities are actually providing high quality medical care.’ Thus, the method is the medium both for defining the expected level of care and for determining performance

The Secretary needs to adequately define and guide the state surveyors’ functions to ensure that there is uniformity in evaluating the delivery of health care. Uniform guidelines are imperative for achieving that goal

The failure of the Rule to include the specifics of the PaCS survey system also raises a procedural due process concern [A] facility adversely affected by findings of deficiency may successfully challenge enforcement because of a failure to give adequate notice of what is required or expected. There is merit in the argument that the failure to include sufficient detail in the regulation may preclude adequate enforcement under commonly accepted principles of fundamental fairness.²³³

The court once again ordered CMS to engage in formal APA rulemaking.²³⁴

Notably, the Tenth Circuit and the District Court were not alone in perceiving the need for binding survey regulations. During congressional hearings, in contrast to the Secretary’s long-standing opposition to promulgating the survey protocol as a formal rule,²³⁵ the IOM advised that, to promote consistency in survey results, the survey protocols and methodologies should be issued in regulations and consolidated in one location in the Code of Federal Regulations.²³⁶

230. *See Smith v. Bowen*, 656 F. Supp. 1093, 1094 (D. Colo. 1987).

231. *Id.* at 1099.

232. *Id.* at 1096.

233. *Id.* at 1097 (citations omitted).

234. *Id.*

235. *See, e.g.*, Medicare & Medicaid Programs; Survey, Certification, and Enforcement of Skilled Nursing Facilities and Nursing Facilities, 59 Fed. Reg. at 56,120.

236. IOM REPORT, *supra* note 6, at 32, 111. *See also Hearing on S. 1108*, *supra* note 223, at 23.

In July 1987, CMS published the second “final” rule in the series.²³⁷ The survey procedures, forms and guidelines were offered for comment in an appendix, but, despite the courts’ orders, still were not part of the binding, final rule and would not be published in the Code of Federal Regulations.²³⁸ Despite several court orders to the contrary, CMS continued to maintain that adoption of the forms, procedures and standards as a rule was undesirable because modifying them to reflect “changes and improvements” would be “slow and difficult.”²³⁹

The court again disagreed. Reiterating the earlier ruling of the Tenth Circuit, and echoing IOM’s concern for consistency, the District Court again held:

To become and remain informed, the Secretary must establish uniform standards for facility performance and a uniform methodology for evaluating that performance to ensure the delivery of high quality health care. Thus, the regulations required for these purposes must be prescriptive and legislative.²⁴⁰

The District Court this time found the Secretary in contempt of court. The defendants were ordered to “take new action to promulgate a rule establishing standards of care and the methodology, forms and directions for the states’ survey certification process.”²⁴¹

In December 1987, seemingly despairing of CMS’ ability to reform the nursing home regulatory system and expressing skepticism about whether defendants would “eventually publish new regulations,” Congress adopted the Nursing Home Reform Act that was included in OBRA ’87.²⁴² OBRA ’87 enacted into substantive law many of IOM’s recommendations.²⁴³

CMS treated the adoption of OBRA ’87 as an opportunity to avoid compliance with the *Smith* courts’ orders and moved the court for relief from the prior orders. The District Court disagreed, however, ruling that the orders requiring that the inspection procedures, forms, and methodology must be promulgated as binding regulations, “are not in any way inconsistent with the Omnibus Budget Reconciliation Act of 1987[.]”²⁴⁴ According to the District

237. 52 Fed. Reg. 24,752 (July 1, 1987).

238. See 52 Fed. Reg. at 24,760.

239. See *Smith v. Bowen*, 675 F. Supp. 586, 589 (D. Colo. 1987).

240. *Id.* at 589.

241. *Id.* at 590.

242. H.R. Rep. No. 100-391(I), at 452 (1987), reprinted in 1987 U.S.C.C.A.N. 2313-1, 2313-272. See also *Hearing on S. 1108, supra* note 223, at 51, 52; Ex. 30; Morford, *supra* note 8, at 130 (discussing defendants’ inability “to establish sufficient consensus to regulate” nursing homes and defendants’ failure “during the last 10 years” to issue new nursing home regulations).

243. H.R. Rpt. No.100-391(I), 100th Cong., 1st Sess., P452, reprinted in 1987 U.S.C.C.A.N. 2313-1 at 2313-272 (IOM Report is the “starting point” for OBRA ’87); Ex. 14, IOM 2001 at 142.

244. *Smith v. Bowen*, No. CIV-75-M-539, 1988 WL 235574, at *1 (D. Colo. Feb. 18, 1988).

Court, “[t]he passage of OBRA [’87] in no way modifies or preempts the Tenth Circuit’s decision.”²⁴⁵

On June 17, 1988, almost one year after OBRA ’87 was adopted and four years after first being ordered to do so, defendants promulgated regulations containing nursing home survey forms, procedures and guidelines that had been in the development process prior to OBRA ’87.²⁴⁶ After fighting the need for prescriptive regulations, in the preamble, CMS explained that OBRA ’87 had “ratif[ied] items included in [the] proposed regulations” so that any OBRA ’87 related change in the survey process would be in the nature of “necessary refinements to the survey process forms and guidelines.”²⁴⁷

Notwithstanding the statement that future developments would be “in the nature of necessary refinements,” CMS adopted new and different survey forms, protocols, procedures, and methodologies to implement OBRA ’87. Despite the *Smith* courts’ holdings, the OBRA ’87 survey forms, procedures, and protocols are not contained in the regulations, but, for the most part, are found in CMS’ State Operations Manual.

The survey procedures, forms and methodologies adopted by defendants in June 1988 never have been repealed. To the contrary, the 1988 survey rules still are codified in the Code of Federal Regulations and are the only survey procedures, forms, and protocol promulgated through formal notice and comment rulemaking.²⁴⁸ Those regulations are not followed, however. Despite taking the position that OBRA ’87 had ratified those regulations and that the regulations conformed so closely to OBRA ’87 that only “refinements” would be necessary, CMS subsequently began to assert that the regulations were “obsoleted” by the passage of OBRA ’87. Moreover, CMS makes the assertion despite the fact that the Administrative Procedure Act expressly provides that the amendment, repeal, or modification of an existing rule requires formal notice and comment rulemaking by the agency proposing the change.²⁴⁹

B. The Consequences of the Failure to “Test and Validate.”

Almost twenty-five years after OBRA ’87 was adopted, the statutory mandate that quality be assessed: (1) based on resident outcomes; and (2) using a protocol that had been tested and validated prior to implementation remains to be realized. The government’s own reports and studies recognize that survey results remain inaccurate, invalid and inconsistent on the whole.

The story, however, is not entirely one of failure. It is fair to say, that most people believe the substantive quality standards promulgated to implement

245. *Id.*

246. *See* 53 Fed. Reg. 22,850 (June 17, 1988) (codified at 42 C.F.R. Part 488, Subpart C).

247. *Id.* at 22,853.

248. *See* 42 C.F.R. Subpart C.

249. 5 U.S.C. § 551(5) (2012).

OBRA '87 are both appropriate and, indeed, largely well done. Similarly, there appears to be widespread belief that the structure of the remedial process is likewise generally appropriate. Ultimately, the problem is that the detection methodology—the third leg of the stool of effective enforcement—is so problematic that it compromises the integrity, and undermines the utility, of the survey and enforcement process.

The OIG, GAO, and other government studies and reports address the problems of the survey process using an antiseptic, theoretical, and philosophic public policy approach that ignores the real-world implications of CMS' failures for nursing facility residents, providers, and caregivers. Instead, as illustrated by the government reports' recent, near exclusive focus on understatement of deficiencies to the exclusion of possible overstatement, CMS seems driven by a political goal to appear tough on violators, rather than the desire to fix the systemic problems with the survey process and achieve across-the-board accuracy and consistency. That is not a value-neutral choice; to the contrary, it has consequences that undermine effective enforcement and facility-based performance improvement efforts.

Individual residents and families seeking safe and effective nursing facility care for themselves or for loved ones have virtually nothing other than anecdotes and the survey results to guide their choice of a nursing facility. Yet, notwithstanding the uniform results of government studies responsible government officials continue to represent that those results are trustworthy and reliable because they are “data from onsite inspections conducted by trained, objective surveyors from state public health departments and CMS.”²⁵⁰ Such statements may be politically expedient, but they are grossly misleading and a betrayal of the public trust.

The deception foisted upon prospective nursing facility residents and their family members is exacerbated because the government—despite the findings of its own studies—publicly endorses the utility of such findings and repeatedly encourages consumers to rely on survey results when choosing among nursing facilities. Indeed, by incorporating and endorsing survey results as “the core of the Nursing Home Compare 5 Star rating system [*sic*],”²⁵¹ the government affirmatively represents that such results are valid, accurate, and an appropriate basis on which individuals should decide which nursing facilities are most likely to provide safe and effective care.

The plethora of government and government-sponsored studies demonstrating the unreliability and inconsistency of survey results suggests that, in a commercial context, similar representations would be an actionable,

250. OFFICE OF THE PRESS SECRETARY, THE WHITE HOUSE, FACT SHEET: ADMINISTRATION ANNOUNCES NEW EXECUTIVE ACTIONS TO IMPROVE QUALITY OF CARE FOR MEDICARE BENEFICIARIES (2014).

251. *Id.*

misleading, unfair, or deceptive trade practice. In addition, once admitted to a nursing facility, residents and their families justifiably view the survey process as providing assurance that quality of care will be maintained. That, too, is an illusion. The government's studies and reports effectively acknowledge that the survey can provide no such assurance.

The flaws in the survey process are equally problematic for care providers. To be effective, as IOM and many others recognize, a regulatory system must be based on valid, accurate assessment of quality. Indeed, IOM anticipated the current situation: a regulatory system that fails to prioritize those values lacks integrity, and ultimately will be disrespected and ignored, if not the subject of derision.²⁵² Likewise, a regulatory system characterized by arbitrary and capricious results will be resented, as is the nursing facility survey in many quarters, and viewed to be something to be tolerated or avoided rather than a means to improve, or guarantee, performance.²⁵³ Moreover, such a regulatory system is a strong disincentive discouraging the professional administrators and caregivers needed to provide quality nursing facility services from entering or remaining in the field.

Indeed, rather than encouraging pursuit of aspirational performance goals such as those embodied in the substantive performance standards imposed by OBRA '87,²⁵⁴ the nursing facility survey system and regulatory system have

252. Abt's Final Report made a similar observation. See FINAL REPORT, *supra* note 66.

253. See J. Scott-Cawiezell, et al., *Moving from a Culture of Blame to a Culture of Safety in the Nursing Home Setting*, 41(3) NURSING FORUM 133 (2006).

254. The aspirational nature of OBRA's performance standards are evident from both the statutory language establishing such criteria and from the regulations and manuals implementing the statute. The regulation that implements OBRA's quality of care criteria for nursing homes participating in the Medicare and Medicaid programs provides: "Each resident must receive and the facility must provide the necessary care and services to attain or maintain the highest practicable physical, mental, and psychosocial well-being" 42 C.F.R. § 483.25 (2014). See also 42 U.S.C. § 1395i-3(b)(2)(2012). CMS's State Operations Manual states that: "'Highest practicable physical, mental, and psychosocial well-being' is defined as the highest possible level of functioning and well-being, limited by the individual's recognized pathology and the normal aging process." CTRS. FOR MEDICARE & MEDICAID SERVS., STATE OPERATIONS MANUAL, APPENDIX PP, GUIDANCE TO SURVEYORS 159 (Rev. 133, 2015), available at https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/som107ap_pp_guidelines_tpcf.pdf. These requirements obligate a nursing facility to "ensure that the resident obtains optimal improvement or does not deteriorate within the limits of a resident's right to refuse treatment, and within the limits of recognized pathology and the normal aging process." *Id.*

That OBRA imposes an aspirational standard of care rather than defining the minimally acceptable level of performance is evident. First, the federal standards do not employ the familiar, comparative or community-based assessment that asks whether a facility provided the care and services that comparable providers in the same or similar community would have provided in the same or similar circumstances. To the contrary, the performance of similarly situated facilities is irrelevant. Instead, compliance with the OBRA standard is determined by focusing exclusively on whether the care delivered met the needs of each individual nursing facility resident. Third, the

institutionalized a regulatory culture of blame that impedes development a culture of safety in nursing facilities.²⁵⁵ When a regulated entity believes that accurate assessment is not the primary goal of the regulatory system, or, at least, consistently likely, the regulatory system becomes an impediment to quality improvement.

No regulatory system can be effective where as here, government regulators operate on a dual track system that effectively excuses the government's failure to comply with statutory mandates but holds nursing facilities to a zero tolerance performance standard.²⁵⁶ The survey inspection process also has come to be held in disrespect because of the appearance, if not the reality, that federal regulators respond—as was the case with the Nursing Home Compare and QIS—to the opportunity to score political points even in the face of evidence that the survey process is seriously flawed.²⁵⁷

The state regulators and inspectors who actually perform the bulk of nursing facility inspections are likewise placed in an impossible position. The disrespect in which the survey process is held and the perceived lack of integrity of the process create tension and adversity between the providers and the regulators who are the face of the process. That daily tension—along with the fact that the survey and enforcement system is, at bottom, punitive, rather than improvement oriented—makes it difficult, if not impossible, for providers

determination whether the regulatory standard was met is not based on the reasonableness of the provider's actions, but, instead, is based on whether each resident achieved optimal physical, mental, and psychosocial status.

255. See Jill Scott-Cawiezell et al., *supra* note 254, at 133, 138. See also Alice F. Bonner et al., *Patient Safety Culture: A Review of the Nursing Home Literature and Recommendations for Practice*, 16 ANNALS OF LONG-TERM CARE, March 2008, available at <http://www.annalsoflongtermcare.com/article/8425>. (“Nursing homes are driven in part by a punitive regulatory environment, governed by regulations from the Centers for Medicare & Medicaid Services (CMS), and each state’s Department of Public Health . . . Certain principles of [Patient Safety Culture] that apply to other industries or to hospitals, such as a fair and just culture versus a culture of blame, may not resonate with nursing homes as they are currently structured.” Further noting that “an open environment around error reporting does not exist in nursing homes. Given the punitive nature of the state survey process, this is not surprising. Efforts to reconcile enforcement of regulation with encouragement of open communication and error reporting in nursing homes is warranted.”).

256. *Compare* 42 C.F.R. § 442.30(a) (2014) (provider agreement is not valid evidence of compliance with certification requirements if the survey agency “failed to use the Federal standards, and the forms, methods, and procedures prescribed by CMS as required under § 431.610 (f)(1) or 488.318(b) of this chapter, for determining the qualifications of providers”), *with* 42 C.F.R. § 488.318 (2014) (Failure to [u]se Federal standards, protocols, and the forms, methods, and procedures specified by CMS . . . does not . . . [i]nvalidate adequately documented deficiencies.”), *and* 42 C.F.R. § 488.305 (1) (2014) (“The State survey agency’s failure to follow the procedures set forth in this section [regarding the conduct of surveys] will not invalidate otherwise legitimate determinations that a facility’s deficiencies exist.”).

257. CMS MILESTONES, *supra* note 214, at 16.

and regulators to cooperate in what should be a joint effort to pursue a shared goal of improving quality. When survey results are modified or citations eliminated at the rates previously discussed, state inspectors and supervisors, just as the providers, lose respect for the survey system, lack confidence in its integrity, and ultimately come to doubt whether what they do actually improves nursing facility performance.

Conducting a nursing home or nursing facility survey is a difficult task in any circumstance. It is made even more difficult when the known defects in the protocol undermine the validity of the results and encourage disrespect for the system as well as for those who apply and enforce the regulations. In such circumstances, the effectiveness of the process is undermined by regulatory fatigue that causes inspectors increasingly to see themselves as performing an ultimately futile and ineffective task.²⁵⁸ It should surprise no one that there is a chronic shortage of surveyors.

In an article published in 1988, a senior federal official responsible for oversight of the survey process stated his opinion that, as a result of OBRA '87, "the nursing home problem has now been fixed through legislation."²⁵⁹ With the benefit of the "retrospectoscope" it can confidently be said that the prediction was both premature and unjustifiably optimistic.

The almost thirty-year old promise of OBRA '87 has been broken because responsible federal officials, rather than implement the mandate of the statute, continued, as before adoption of OBRA '87, to respond to the demands of a dysfunctional political system and to pursue their own agendas. Similarly, nursing home industry and resident advocates, as well as well-regarded social scientists, and public health policy commentators have accepted the failure to implement the statute's mandates, rather than holding public officials accountable. Indeed, some well-regarded notable researchers have asserted, as have highly placed public officials, that the survey data are in fact accurate and sufficiently reliable to be used by consumers, in enforcement actions, public health research, analysis, and policy formulation.²⁶⁰ Likewise, industry and

258. See, e.g., Patricia A Butler, *Assuring the Quality of Care and Life in Nursing Homes: The Dilemma of Enforcement*, 57 N. CAR. L. REV. 1317, 1327-30 (1979); See also Lee et al., *supra* note 213, at 772 n. 6. ("Dissatisfaction with the survey process is wide-spread. Resident advocacy groups stress that state survey teams often miss important problems with care and fail to respond to complaints quickly Surveyors question the integrity of the inspection, political pressures to water down inspection findings, and the effectiveness of the enforcement process Industry representatives agree that the current survey and enforcement system "is an entirely subjective, process oriented snapshot inspection system that focuses on punishment – not quality").

259. Morford, *supra* note 8, at 129, 132.

260. See, e.g., OFFICE OF THE PRESS SECRETARY, THE WHITE HOUSE, FACT SHEET: ADMINISTRATION ANNOUNCES NEW EXECUTIVE ACTIONS TO IMPROVE QUALITY OF CARE FOR MEDICARE BENEFICIARIES (Oct. 6, 2014) ("[O]n-site inspections form the core of the [5-Star]

resident advocates use survey data to support their own perspectives. So long as that situation continues, the interests of nursing facility residents will be, at least secondary, and it is unlikely that the goals set forth in IOM's 1986 study and incorporated into OBRA '87 ever will be achieved.²⁶¹

In 1984, the Tenth Circuit, in a lengthy opinion, had to parse numerous provisions of the Medicaid Act to find CMS's duty to adopt and utilize a survey protocol that accurately assessed the quality of nursing facility care. In OBRA '87, Congress made such painstaking piecework unnecessary. Congress wrote the agency's duty into the statute in clear and unequivocal language and, lest there be any residual doubt, Congress summarized that obligation in a single, explicit provision that makes clear that the needs of nursing home residents, not administrative convenience or political expediency, ought to be the touchstone:

It is the duty and the responsibility of the Secretary to assure that the requirements which govern the provision of care in skilled nursing facilities under this subchapter, and the enforcement of such requirements are adequate to protect the health, safety, welfare, and rights of the residents and to promote the effective and efficient use of public moneys.²⁶²

Just as was the case with the *Smith* court's orders, CMS seems determined to ignore the obligations created by the statute and to continue the pattern begun in the 1980s. Almost thirty years after OBRA '87 was adopted, it is long past time for the Secretary to fulfill her duty.

rating system"; quality of care information is "based on data from onsite inspections conducted by trained, objective surveyors from state public health departments and CMS."); Christiana Williams et al., *Measuring Nursing Home Quality – The Five-Star Rating System*, SAS GLOBAL FORUM (2010) at 4, available at <http://support.sas.com/resources/papers/proceedings10/161-2010.pdf>. ("CMS wanted the health inspection domain to play a predominant role in determining the overall 5-star rating. The main reason for this is that, unlike the staffing and quality measures domains which are based on data self-reported by nursing homes, independent surveyors carry out the health inspection; thus, the data should be more objective and unbiased."); C. Harrington et al., *The Regulation and Enforcement of Federal Nursing Home Standards, 1991-1997* 56 MED. CARE RES. REV., Dec. 1999, at 474. ("The deficiency data . . . are recorded by state surveyors after they determine whether the facility has met, or has not met, each standard. These data are considered accurate because they are reviewed carefully by [CMS] and subject to challenges by nursing facilities.").

261. See, e.g. Lee et al, *supra* note 213, at 779. ("Yahoo! Reports that "Nursing Home Compare" is the nation's second most popular nursing home care sites and in some of the most frequently visited sections of the Medicare Web site [Thus], the reliability of nursing home surveys becomes an even more visible public policy issue. Survey results will have the greatest impact on nursing home quality if consumers and the industry believe that deficiencies are valid, reliable measures of quality. This belief will be undercut by variations in the number, scope, and severity of deficiencies when the facts are held constant. The appropriate policy response is to acknowledge these variations and address them by clarifying definitions and interpretations, by improving training, and by providing feedback to surveyors.").

262. 42 U.S.C. § 1395i-3(f)(1) (2012).