

Beneath the Veneer of Healthcare Quality Measurement

Adeline J. Wilcox*

Abstract

Under contract to the Centers for Medicare & Medicaid Services, The Joint Commission specifies sample sizes for healthcare quality measurement. Their sample size specifications do not follow established methodology for sample size estimation. Sample sizes for estimates are specified without explanation or calculations for precision. A recent manifesto by the National Academy of Medicine (NAM) makes clear that the analytical objectives of healthcare quality measurement remain to be determined [35]. The NAM authors use the term reliability without addressing either precision or sample size. Healthcare quality measurement is a form of self-regulation by the medical profession. Physicians control healthcare quality measurement. Organized statisticians have no apparent role. Until January 2017, the Office of Management and Budget (OMB) required the principal statistical agencies to plan the precision of their estimates. However, OMB did not apply their requirements to healthcare quality measurement. Physicians expect to practice free from outside scrutiny. With its "consumer-oriented website", Hospital Compare, reporting healthcare quality measures, CMS bestows an illusion of quality on the medical profession.

Key Words: sample size, estimation, healthcare, quality measurement

1. Twenty Years of Aimless Measurement

Healthcare quality measures are also known as performance measures. While healthcare quality measurement may have begun as early as 1754 [26], healthcare quality measurement as we know it began about 20 years ago [35].

In a recent discussion paper, leaders in healthcare performance measurement describe a dilemma they see when reporting scores[35]. They cannot decide whether to report statistically significant or clinically significant differences. From the patient's perspective, clinically significant differences is a no-brainer. The paper is one of a series of papers published by the National Academy of Medicine, intended to give expert guidance on health policy. Ironically, the National Academy of Medicine is under the umbrella of the National Academies of Sciences, Engineering, and Medicine along with the Committee on National Statistics (CNSTAT). Evidently, the authors did not consult the CNSTAT statisticians.

With analytical objectives ill-defined, no design for sample size estimation exists.

2. Discovery of Naive Sample Size Specifications

From 1998, hospitals seeking accreditation by The Joint Commission (TJC) have been required to report performance measurement data [4]. Later, the Centers for Medicare & Medicaid Services (CMS) also began requiring performance measurement data reporting.

*Retired, 10901 Fleetwood Drive, Beltsville, MD 20705

To meet both requirements, the U.S. Department of Veterans Affairs (VA), Veterans Health Administration (VHA), reports measure data to TJC and CMS, respectively. Except VHA, hospitals need accreditation to bill CMS [10]. VHA cannot bill CMS [32]. Apart from granting deeming authority to TJC, CMS contracts with TJC for work including sample size specifications [20]. A number of federal government and non-governmental agencies are involved in healthcare performance measurement. Of the four agencies listed in Table 1, two of the four are headed by a physician [17, 15, 33, 12]. Both CMS and the Agency for Healthcare Research and Quality (AHRQ) are Divisions of the U.S. Department of Health & Human Services. At CMS, the Deputy Administrator, a physician, directs healthcare quality measurement. It appears that organized statisticians or data scientists have no role in healthcare performance measurement.

Agency	Physician	Title
Centers for Medicare & Medicaid Services	Patrick Conway, M.D.	Deputy Administrator for Innovation and Quality
Agency for Healthcare Research and Quality	Gopal Khanna, M.B.A.	Director
U.S. Department of Veterans Affairs	David J. Shulkin, M.D.	Secretary
The Joint Commission	Mark R. Chassin, M.D.	President and Chief Executive Officer

Table 1: Top Administrators with Oversight of Healthcare Performance Measurement

At VHA, data for computing performance measures were abstracted from VHA medical records by WVMI & Quality Insights [23], VHA's External Peer Review Program (EPRP) contractor. However, WVMI & Quality Insights is not an ORYX® vendor. Only ORYX vendors may report performance measure data to TJC [9].

**Quarterly Sample Size
Based on Hospital's Initial Patient Population for the ABC Measure Set**

Average Quarterly Initial Patient Population "N"	Minimum Required Sample Size "n"
≥ 1551	311
391 - 1550	20% of the Initial Patient Population
78 - 390	78
6 - 77	No sampling; 100% of the Initial Patient Population is required
0 - 5	Submission of patient level data is encouraged but not required: <ul style="list-style-type: none"> • CMS: if submission occurs, 1 – 5 cases of the Initial Patient Population may be submitted • The Joint Commission: if submission occurs, 100% Initial Patient Population required

Figure 1: A Naive Sampling Specification Published by The Joint Commission

VHA is a unique ORYX vendor. First, it appears to be one of only two vendors that also serve as their own ORYX vendor [13]. Besides VHA, only McLean Hospital in Belmont MA, is a hospital as well as an ORYX vendor. Second, VHA is the only ORYX vendor that does not accept business from other providers. Except for VHA, raising questions about sample size is probably not in the interest of ORYX vendors. Examination of TJC's list of vendors reveals that few are also providers. Many ORYX vendors are data processing corporations for whom greater sample size means greater profit.

From February 2010 through April 2016, I was employed by VHA as a Health System Specialist. VHA hired me to serve as a backup, should the employee then solely responsible for writing and executing SAS[®] programs for sampling, be unable to carry out his duties. This employee gave me only limited information about the mechanics of modifying and executing his undocumented SAS programs. To better understand the work I was expected to do, I visited TJC's website where I located their sampling specifications [7]. I found their sample size specifications do not follow established methods for sample size estimation. Figure 1 shows an example. TJC's specifications do not explain how TJC developed their sample size algorithm. TJC specifications say nothing about sampling error, precision or confidence intervals. The box holds most of the body of the letter of complaint I sent to Andy Slavitt, Acting Administrator, Centers for Medicare & Medicaid Services on 02Feb2016 [1]. The Enclosure I referred to in my letter was a copy of the specifications displayed in Figure 1.

As shown in the Enclosure [7], as population size increases, TJC's specified sample size for unstratified measure sets begins at 100 percent of the population, decreases to an arbitrary 20 percent of the population, then, as population size increases above 1550, plateaus at a fixed sample size of 311. How TJC developed their sample size recipe is not explained. TJC specifications say nothing about sampling error, precision, or confidence intervals.

Analytical objectives are not given in TJC specifications. I found them in two documents on "Target Analysis" citetarget1,target2. Neither document illuminates TJC's sample size requirements.

I find TJC's documentation unconvincing. I doubt the sample size specifications given by TJC could win a grant award from the National Institutes of Health.

More statistical science underlies estimates of unemployment [30] and polls of persons identifying as voters than estimates dictated by TJC's quality measurement program. Even online merchandising benefits from more sophisticated application of statistical methods than healthcare quality measurement specified by TJC.

TJC does not recognize the important role of statisticians in improving healthcare safety and quality. No Chief Analytics Officer is listed on their web site [5].

Whether the sample sizes specified by TJC are too small or too large cannot be determined from their documentation. Work on deliverables to TJC consumes Veterans Health Administration resources [3] that could be directed at evidence-based efforts to improve healthcare quality.

TJC's statistical methodology needs examination. Please act to assure all patients benefit from statistically sound healthcare quality measurement.

About one month later, a CMS Nurse Consultant named Megan R. Hayden spoke with me by telephone [20], followed by a conference call with TJC [19] several days later. Two of the slides TJC prepared for the conference call are reproduced in Figure 3 and Figure 4.

While preparing for the conference call with TJC arranged by CMS, I discovered the paper by Harmon S. Jordan titled *Maximizing Sampling Efficiency* [24]. He recommended specifying the precision of sample estimates made for healthcare performance measurement.

A few days before the conference call, Megan Hayden sent me a more recent version of a document from TJC titled *Target Analysis (core measures only)*. In late 2015, I had already found an earlier version of this document on TJC website [2]. This document and a related one [8] are no longer available on the public facing TJC website.

Three days after the conference call with TJC, I wrote Megan Hayden an email [39] in which I reminded her that I had stated that page 12 of the document titled *Target Analysis (core measures only)* looked "badly faked". In that email I advised her to compare TJC's "mixed model" with the Wikipedia entry for mixed model. Finally, I wrote "I recommend CMS have a committee made up of academic statisticians review The Joint Commission's sampling methodology for performance measurement."

The letter of reply to my complaint came from a CMS physician [27]. Dr. Kate Goodrich gave no indication that CMS consulted any statisticians besides Stephen Schmaltz at TJC. No one at CMS ever responded to my allegations that TJC faked their documentation.

3. Healthcare Quality Measures

Among TJC's measure sets is one named Immunization. Hospitalized patients with documented pneumococcal vaccination are "In Numerator Population". The "Measure Population" comprises the denominator [7]. Vaccinated patients are assigned a score of one while patients who, by protocol, should have been vaccinated but were not, are assigned a score of zero. IMM-1a, the example measure described in Figure 2, is a process measure, as opposed to an outcome measure. Like IMM-1a, most healthcare quality measures are dichotomous. Measures are reported to the public on a CMS website named Hospital Compare [18].

IMM-1: Pneumococcal Immunization

Numerator: Inpatient discharges who were screened for pneumococcal vaccine status and received pneumococcal vaccine prior to discharge, if indicated.

Denominator: Inpatient discharges 65 years of age and older, and 5 through 64 years of age who have a high risk condition.

Variable Key: Patient Age

Stratification Table:

Set Measure ID#	Stratified Measure Name
IMM-1a	Pneumococcal Immunization – Overall Rate
IMM-1b	Pneumococcal Immunization – Age 65 and Older
IMM-1c	Pneumococcal Immunization – High Risk Populations (Age 5 through 64 years)

Figure 2: Pneumococcal Immunization Measure Descriptions

4. Healthcare Quality Measures Compared with Survey Estimates

From the Centers for Disease Control (CDC) Behavioral Risk Factor Surveillance System (BRFSS), the estimated maximum rate of pneumococcal vaccination coverage among adults ≥ 65 years of age was 76.1 percent in 2014 [14]. For persons age 18-64 years at increased risk, the maximum rate of pneumococcal vaccination coverage in 2014 was 37.9 percent. Unlike the CDC, AHRQ gives neither a sample size nor a confidence interval for their estimate, 92.2 percent of "Hospital patients who received pneumococcal immunization" [16].

More disheartening than the inattention to sample size estimation are the poorly defined analytical objectives. Indeed, as if analysis was an afterthought, AHRQ arbitrarily compares measure estimates to benchmarks "derived from the top-performing States."

Without citing a source, TJC stated pneumococcal vaccine has been administered to too few members of the population. TJC sees hospitalization as an opportunity for vaccination. I see IMM-1a measuring, to an extent, how well a hospital corrects omissions in outpatient care. Of course, hospitalized patients may be less likely to refuse vaccination than they were in an outpatient setting. How this measure could be useful to patients choosing a hospital, the TJC documentation does not explain. For patients who were vaccinated against pneumococcus before hospitalization, documentation may not be available. If the patient's record shows the patient's vaccinations are "up to date", TJC accepted this as evidence of pneumococcal vaccination.

It should come as no surprise that a notice dated 17Jun2016 on TJC web site tells us that the hospital Immunization measures are no longer required by TJC or CMS [11]. In 2010, the CDC published updated pneumococcal vaccination recommendations [29].

Sample Size – Margin of Error



- ▶ The sample size was determined so that a hospital's annual measure rate would have a margin of error of 10% (assuming a national rate of 50% to be conservative).



Figure 3: Sampling Methodology The Joint Commission March 2016 Discussion. Slide 4.

5. Healthcare Quality Measurement Harm

Pronovost reported a case in which a couple decided against having the husband's esophageal surgery performed at a hospital with greater procedure volume only because of the hospital's safety scores. Following complications at a hospital with higher safety scores but lower surgical volume for the procedure, the patient died [34].

On occasion, measurement has taken precedence over patient care. Based on a news report [22], VA likely met a colon cancer screening quality measure but neglected to actually care for a patient who screened positive. This regular patient at a VA clinic had a fecal occult blood test performed three years in a row. Although the test result was positive each time, the patient's VA healthcare providers never informed him. Only when the patient switched doctors, did he learn he had metastatic cancer. Following the overdue diagnosis, the patient lived less than two years.

In a case reported by a staff nurse, a surgical intensive care unit patient's family wanted palliative care [36]. The advance directive in the patient's chart supported palliative care over intensive care. Yet, citing a 30 day mortality rate quality measure, the physician would not allow the patient to be removed from respiratory support until 30 days after the surgery from which the patient had not recovered.

6. Neoliberalism Manifested in Healthcare Quality Measurement

Healthcare deregulation manifests as healthcare quality measurement, shifting responsibility for quality from government to the consumer [25]. Implicit in reporting scores for what appears to be the universe of providers, is the decision made by CMS not to identify and investigate providers whose scores indicate they provide substandard care. In leaving patients to choose the best providers, CMS follows neoliberalism [41].

Neoliberalism is a theory and practice of political economy that demands a reduction in government provision of services and advocates for the primacy of the private market. At the center of neoliberal ideology is the belief that open, competitive, and unregulated markets, liberated from all forms of state interference, are the optimal mechanism for economic development.

Under a government which puts citizens' interest before those of the free market, the death associated with use of performance measurement scores Pronovost reported [34], might not have occurred. Under such a government, the objectives and implementation of healthcare quality measurement would be different. Hospitals might be regulated in a way that requires them to perform at least a minimum number of each surgical procedure. Pronovost is so committed to neoliberal principles that he does not recommend regulation even to prevent deaths like this one. By explicitly embracing neoliberalism, the authors of the National Academy of Medicine discussion paper [35] put politics ahead of patient safety. They consider performance measurement critical to an efficiently functioning free healthcare market.

20% Sampling Band - History



- ▶ The 20% sampling band is a holdover from the old CMS sampling requirements before alignment with The Joint Commission on the measures in common with CMS.
- ▶ These original CMS requirements were not statistically-based.
- ▶ The requirement was driven by practical considerations and what CMS initially thought that hospitals would be able to handle as far as data collection burden.



Figure 4: Sampling Methodology The Joint Commission March 2016 Discussion. Slide 9.

7. Discussion

To make inferences about the superpopulation, sampling 100 percent of cases from hospitals with eligible populations of size 78 or fewer, as shown in Figure 1, may not be valid [21]. In his slide in Figure 4, for hospitals with a population size of 391 to 1550, Stephen Schmaltz noted that CMS based the 20 percent sampling rate on their guess at the sample size hospitals would tolerate without protest [37]. This is not science. For hospitals with a population sizes of 78 to 390 and 1551 or greater, I find no calculation of precision for the respective sample sizes of 78 and 311 in either TJC specifications or their slide deck. In 2013, increases in sample size specified by TJC were estimated to increase the VHA's EPRP costs by 3 million dollars [40]. As noted above, except possibly VHA, it is not in the interest of ORYX vendors to raise questions about sample sizes.

Malcolm K. Sparrow observed [38]

Society has traditionally paid physicians the compliment, as a profession, of subjecting medical judgement to scrutiny only by another physician.

Unlike the principal federal statistical agencies which were required to plan precision of estimates, CMS escaped regulation by the Office of Management and Budget [31]. TJC slide [37] in Figure 3 gives a margin of error of 10. While this could be close to the true margin of error for healthcare quality measures, a 10% margin of error is not precision.

8. Conclusion

CMS and TJC have not yet improved their specifications for sample size estimation [6].

The slide deck TJC [37] prepared for the conference call [19] among Megan R. Hayden, TJC's biostatistician Stephen Schmaltz, several other members of TJC's staff and myself did not satisfactorily explain their sample size estimation methodology. One of the slides, Figure 4, titled 20 Sampling Band–History, includes the statement

The 20 sampling band is a holdover from the old CMS sampling requirements before alignment with The Joint Commission on the measures in common with CMS.

and admits the 20 sampling band is "not statistically-based". I recommend CMS cease contracting for healthcare quality measurement sample size estimation. CMS should hand this work over to a statistical office to be created in the U.S. Department of Health & Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Michael Millenson noted that physicians do not police themselves effectively[28].

As I would later discover when transitioning from journalist to policy wonk, the idea that doctors ever effectively disciplined the miscreants among them is just one of the "Golden Age" myths to which physicians cling.

Fundamentally, the neoliberal objective of healthcare quality measurement is not health and safety but economic growth. Healthcare quality measurement needs to be reconsidered.

Credit Notices

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. ORYX® is The Joint Commission's performance measurement and improvement initiative, first implemented in 1997.

REFERENCES

- [1] U.S. Department of Veterans Affairs. Veterans Health Administration Adeline J. Wilcox, 02 February 2016. Letter to Andy Slavitt, Acting Administrator, Centers for Medicare & Medicaid Services.

- [2] Anonymous. Target analysis (core measures only). Web, 16 January 2009.
- [3] Mark R. Chassin, Jerod M. Loeb, Stephen P. Schmaltz, and Robert M. Wachter. Accountability Measures – Using Measurement to Promote Quality Improvement. *N Engl J Med* 363:7, 12 August 2010.
- [4] Mark R. Chassin, Jerod M. Loeb, Stephen P. Schmaltz, and Robert M. Wachter. Accountability Measures–Using Measurement to Promote Quality Improvement. *The New England Journal of Medicine*, 36;7:683–688, 12 August 2010.
- [5] The Joint Commission. Joint Commission Officers Quick List. Web. Accessed 27Jan2016.
- [6] The Joint Commission. Specifications manual for national hospital inpatient quality measures discharges 01-01-17 (1q17) through 12-31-17 (4q17) version 5.2a. Accessed 30 September 2017.
- [7] The Joint Commission. Specifications Manual for National Hospital Inpatient Quality Measures. Discharges 10-01-15 (4Q15) through 06-30-16 (2Q16) Version 5.0b. Web. <http://www.jointcommission.org/> Click on Measurement. Accessed 28 January 2016.
- [8] The Joint Commission. Target analysis methodology for assessing hospital performance on the aligned cms/joint commission national hospital quality measures (core measures). publisher=The Joint Commission, note=Accessed 05May15. Web.
- [9] The Joint Commission. General ORYX requirements for hospitals. Web, 12 January 2015. Accessed 24 April 2016.
- [10] The Joint Commission. Facts about federal deemed status and state recognition. Web, Accessed 09 September 2016.
- [11] The Joint Commission. Immunization. Web, 17 June 2016. Accessed 27Aug2016.
- [12] The Joint Commission. Joint commission officers. Web, Accessed 29 September 2017.
- [13] The Joint Commission. Listed oryx vendors and candidates supporting chart-based measure sets for accreditation and certification. Web, 29 August 2017. Accessed 29 September 2017.
- [14] Centers for Disease Control and Prevention. Adultvaxview. 2014 Adult Vaccination Coverage General Population Report. Web. Accessed 05Sep2016. <http://www.cdc.gov/vaccines/imz-managers/coverage/adultvaxview/data-reports/general-population/reports/2014.html>.
- [15] Agency for Healthcare Research and Quality. Office of the director. Web, May 2017. Accessed 29 September.
- [16] Agency for Healthcare Research and Quality. U.S. Department of Health & Human Services. National Quality Measures Compared to Achievable Benchmarks. National Healthcare Quality and Disparities Reports. Web. Accessed 03 September 2016.
- [17] Centers for Medicare & Medicaid Services. Web, Accessed 03 September 2016.
- [18] Centers for Medicare & Medicaid Services. Hospital Compare. Web, 04 May 2016. Accessed 08 September 2016.
- [19] Megan R. Hayden. Conference Call 22 March 2016. Megan R. Hayden, Centers for Medicare & Medicaid Services, Adeline Wilcox, U.S. Department of Veterans Affairs, Veterans Health Administration, Stephen Schmaltz and David Baker MD, The Joint Commission and several more employees of TJC.
- [20] Megan R. Hayden. Telephone call 15 March 2016. Megan R. Hayden, Centers for Medicare & Medicaid Services, Adeline Wilcox, U.S. Department of Veterans Affairs, Veterans Health Administration.

- [21] R. Houchens. *Inferences with HCUP State Databases Final Report*. HCUP Methods Series Report # 2010-05. U.S. Agency for Healthcare Research and Quality, Online October 12 2010. www.hcup-us.ahrq.gov/reports/methods/2010_05.pdf Accessed 29 April 2015.
- [22] Creela Belle Howard and Reuven Blau. Legal settlements at Veterans Affairs more than triple since 2011, many due to medical malpractices. *New York Daily News*, 30 May 2016. Accessed 6 June 2016.
- [23] WVMI Quality Insights. Our major projects. Web. Accessed 07 September 2016.
- [24] Harmon S. Jordan. Maximizing Sampling Efficiency. *Applied Mathematics*, 4:1547–1557, November 2013. doi: 10.4236/am.2013.411209.
- [25] Sue McGregor. Neoliberalism and health care. *International Journal of Consumer Studies*, 25(2):82—89, June 2001.
- [26] Dennis McIntyre, Lisa Rogers, and Ellen Jo Heier. Overview, history, and objectives of performance measurement. *Health Care Financing Review*, 22(3), Spring 2001.
- [27] Kate Goodrich MD, 15 April 2016. Centers for Medicare & Medicaid Services, Center for Clinical Standards and Quality. Letter to Adeline Wilcox, Office of Analytics and Business Intelligence, Veterans Health Administration.
- [28] Michael Millenson. When The American Medical Association Cheered Hillary. Web, 30 July 2016. Forbes Pharma & Healthcare. Accessed 28Aug2016.
- [29] JP Nuorti and CG Whitney. Updated Recommendations for Prevention of Invasive Pneumococcal Disease Among Adults Using the 23-Valent Pneumococcal Polysaccharide Vaccine (PPSV23). *Morbidity and Mortality Weekly Report (MMWR)*, 59(34):1102–1106, 03 September 2010. Div of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, CDC, for the ACIP Pneumococcal Vaccines Working Group.
- [30] Bureau of Labor Statistics and U.S. Census Bureau. *Current Population Survey. Technical Paper 63RV. Design and Methodology*. Web. Accessed 22Jan2016., March 2002.
- [31] Office of Management and Budget. Office of Management and Budget Standards and Guidelines for Statistical Surveys. Web, September 2006. Accessed 07 September 2016.
- [32] U.S. Department of Veterans Affairs. Why doesn't VA bill Medicare? Web, Accessed 09 September 2016.
- [33] U.S. Department of Veterans Affairs. Web, Accessed 29 September 2017.
- [34] Peter Pronovost. Why surgical volumes should be public. *U.S. News & World Report Health Care Second Opinion*, 18 November 2016. Web. Accessed 7 February 2017.
- [35] Peter J. Pronovost, J. Matthew Austin, Christine K. Cassel and Suzanne F. Delbanco, Ashish K. Jha, Bob Kocher, Elizabeth A. McGlenn, Lewis G. Sandy, and John Santa. *Fostering Transparency in Outcomes, Quality, Safety, and Costs*. Vital Directions for Health and Health Care Initiative. National Academy of Medicine, Washington DC, 19 September 2016.
- [36] Betty Rambur, Carol Vallett, Judith A. Cohen, and Jill Mattuck Tarule. Metric-driven harm: An exploration of unintended consequences of performance measurement. *Applied Nursing Research*, 26:269–272, 2013.
- [37] Stephen Schmaltz and Dave Morton. Sampling Methodology. The Joint Commission March 2016 Discussion, 20 March 2016. Slide program. Unpublished attachment to email from Megan R. Hayden to Adeline Wilcox. 21 March 2016.
- [38] Malcolm K. Sparrow. *License to Steal: How Fraud Bleeds America's Health Care System*. Westview Press, Boulder, 2000.

- [39] Adeline Wilcox. Subject: Re: Sampling methodology. E-mail, 25 March 2016. To: Megan Hayden.
- [40] Steven M. Wright. Performance Measurement Team. Conference Call, 31 July 2014.
- [41] Yue Zhang. Historic Preservation in a Neoliberal Context: From the Medinah Temple to Bloomingdale's. In Larry Bennett, Roberta Garner, and Euan Hague, editors, *Neoliberal Chicago*. University of Illinois Press, Urbana, Chicago, and Springfield, 2017.