Sample Size Estimation with PROC FREQ and PROC POWER
Adeline J. Wilcox, Retired

Abstract

Under contract to the Centers for Medicare & Medicaid Services, The Joint Commission specifies sample sizes for healthcare quality measurement. Their sample size specifications pay no heed to established methods for sample size estimation. SAS/STAT® 14.1 PROC POWER can be used to compute sample size estimates with precision.

From healthcare measurement data I used as pilot samples, I computed upper and lower confidence limits. To do this, I used the BINOMIAL and CL options on the PROC FREQ TABLE statement. After examining these results, I chose input values for the PROC POWER HALFWIDTH and PROBWIDTH options. More statistics textbooks cover sample size estimation for hypothesis testing than for estimation.

Introduction

From 1998, hospitals seeking accreditation by The Joint Commission (TJC) have been required to report performance measurement data [2]. Later, the Centers for Medicare & Medicaid Services (CMS) also began requiring performance measurement data reporting. 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 [8]. VHA cannot bill CMS [33]. Apart from granting deeming authority to TJC, CMS contracts with TJC for work including sample size specifications [18].

At VHA, data for computing performance measures are 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 [5].

VHA is a unique ORYX vendor. First, it appears to be one of only two vendors that also serve as their own ORYX vendor [7]. 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 hospitals.

A number of federal government and non-governmental agencies are involved in healthcare performance measurement. Of the four agencies listed in Table 1, all but CMS are headed by a physician [15, 13, 32, 6]. 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.

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 [4]. 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 on the next page holds an excerpt from the letter of complaint I sent to Andy Slavitt, Acting Administrator, Centers for Medicare & Medicaid Services on 02Feb2016 [1]. About one month later, a CMS Nurse Consultant named Megan R. Hayden spoke with me by telephone [18], followed by a conference call with TJC [17] several days later. Two of the slides TJC prepared for the conference call are reproduced in Figure 3 and Figure 7.
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. The letter of reply to my complaint came from a CMS physician [27]. It gave no indication that CMS consulted any statisticians.

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.

### Healthcare Quality Measures

Healthcare quality measures are synonymous with performance 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 [4]. 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 [16].

### Healthcare Quality Measurement 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 [12]. See Figure 4. 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" [14].

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."

<table>
<thead>
<tr>
<th>Agency</th>
<th>Physician</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centers for Medicare &amp; Medicaid Services</td>
<td>Patrick Conway, M.D.</td>
<td>Deputy Administrator for Innovation and Quality and CMS Chief Medical Officer</td>
</tr>
<tr>
<td>Agency for Healthcare Research and Quality</td>
<td>Andrew B. Bindman, M.D.</td>
<td>Director</td>
</tr>
<tr>
<td>U.S. Department of Veterans Affairs</td>
<td>David J. Shulkin, M.D.</td>
<td>Under Secretary for Health</td>
</tr>
<tr>
<td>The Joint Commission</td>
<td>Mark R. Chassin, M.D.</td>
<td>President and Chief Executive Officer</td>
</tr>
</tbody>
</table>

Table 1: Top Physicians with Oversight of Healthcare Performance Measurement
Quarterly Sample Size
Based on Initial Patient Population for the ABC Measure Set

<table>
<thead>
<tr>
<th>Hospital’s Measure</th>
<th>Average Quarterly Initial Patient Population</th>
<th>Minimum Required Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“N”</td>
<td>“n”</td>
</tr>
<tr>
<td>≥ 1551</td>
<td>311</td>
<td></td>
</tr>
<tr>
<td>391 - 1550</td>
<td>20% of the Initial Patient Population</td>
<td></td>
</tr>
<tr>
<td>78 - 390</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>6 - 77</td>
<td>No sampling; 100% of the Initial Patient Population is required</td>
<td></td>
</tr>
<tr>
<td>0 - 5</td>
<td>Submission of patient level data is encouraged but not required:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• CMS: if submission occurs, 1 – 5 cases of the Initial Patient Population may be submitted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The Joint Commission: if submission occurs, 100% Initial Patient Population required</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1: A Naive Sampling Specification Published by The Joint Commission](image)

**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:**

<table>
<thead>
<tr>
<th>Set Measure ID#</th>
<th>Stratified Measure Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMM-1a</td>
<td>Pneumococcal Immunization – Overall Rate</td>
</tr>
<tr>
<td>IMM-1b</td>
<td>Pneumococcal Immunization – Age 65 and Older</td>
</tr>
<tr>
<td>IMM-1c</td>
<td>Pneumococcal Immunization – High Risk Populations (Age 5 through 64 years)</td>
</tr>
</tbody>
</table>

![Figure 2: Pneumococcal Immunization Measure Descriptions](image)

Without citing a source, TJC states 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 accepts 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 [9]. This development was not known to me when I computed sample estimates of the binomial proportion by VA Medical Center (VAMC). In 2010, the CDC published updated pneumococcal vaccination recommendations [29].
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 3: Sampling Methodology The Joint Commission March 2016 Discussion. Slide 9.

**Practice Analyses with SAS**

The PROC FREQ and PROC POWER analyses are not intended to show how TJC should have specified their sample sizes; they are intended only to demonstrate that SAS can be used for computing an estimated sample size and the probability that a desired precision can be achieved. No special consideration has been given for sampling from a non-finite population [24]. My computations are only a beginner’s effort, merely illustrating that software for sample size estimation and analyzing precision for planning is available.

Figure 4: Pneumococcal vaccination coverage among adults ≥ 65 years
Table 2: Sample Estimates of the Binomial Proportion for Measure IMM-1a by VA Medical Center

<table>
<thead>
<tr>
<th>VA Medical Center</th>
<th>Lower Confidence Limit</th>
<th>Upper Confidence Limit</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>0.93939</td>
<td>1.00000</td>
<td>33</td>
</tr>
<tr>
<td>514</td>
<td>0.96970</td>
<td>1.00000</td>
<td>33</td>
</tr>
<tr>
<td>543</td>
<td>0.75862</td>
<td>0.88933</td>
<td>58</td>
</tr>
<tr>
<td>546</td>
<td>0.86154</td>
<td>0.96120</td>
<td>65</td>
</tr>
<tr>
<td>555</td>
<td>0.75676</td>
<td>0.92084</td>
<td>37</td>
</tr>
<tr>
<td>584</td>
<td>0.88636</td>
<td>0.99767</td>
<td>44</td>
</tr>
<tr>
<td>609</td>
<td>0.94872</td>
<td>1.00000</td>
<td>39</td>
</tr>
<tr>
<td>647</td>
<td>0.91667</td>
<td>1.00000</td>
<td>36</td>
</tr>
<tr>
<td>670</td>
<td>0.96000</td>
<td>1.00000</td>
<td>50</td>
</tr>
<tr>
<td>677</td>
<td>0.92593</td>
<td>1.00000</td>
<td>27</td>
</tr>
</tbody>
</table>

Analysis could be aimed at answering the question: Is the hospital where my surgeon/other specialist/primary care provider has admitting privileges good enough? Patients do not choose a hospital from a large number of hospitals. Their provider usually chooses a hospital for them. If the patient is not satisfied with the hospital their provider wishes to admit them to, the patient may need to find another provider with admitting privileges at the desired hospital. Travel costs also limit the number of hospitals patients consider.

For another analysis, suppose we know a population health expert who tells us 94 percent of patients in the Measure Population should be immunized. Suppose this rate has been adjusted downwards for patients who refuse care, are too ill to be immunized, or are in comfort care, yet still provides herd immunity. We could design a sample aimed at giving a narrow confidence interval inside which we can expect our estimate to fall with known probability.

Computing Confidence Limits With PROC FREQ

Listing 1 gives the SAS code I wrote to compute upper and lower confidence limits for the binomial proportion. I chose measure IMM-1a because I could not run PROC FREQ as listed on missing data. To request confidence limits for the binomial proportion, I added the CL option to the TABLES statement.

To compute the binomial proportion of cases meeting the measure and other binomial statistics from sampled hospital stays, I used the BINOMIAL option on the TABLES statement [21]. Also on my TABLES statement, I set the confidence level for the confidence limits with the ALPHA option. On the OUTPUT statement, the BINOMIAL option is required to output binomial statistics.

Table 2 displays 10 of the 116 observations in the output data set named freqout. Although I reported the use of invalid station identifiers to VA management in July 2015 [35], the data set I read into SAS in April 2016 still contained these invalid VHA hospital identifiers.

Sample Size Estimation With PROC POWER

The code shown in Listing 2 somewhat follows Example 89.7 in the SAS/STAT 14.1 documentation [22]. I did not compute the standard deviation from the IMM-1a measure data. To use PROC POWER with the ONESAMPLEMEANS statement, I estimated the standard deviation as 0.21. For illustrative purposes, I also used 0.11 as a standard deviation value. Viewing the proportion of immunized patients as a special case of the mean [26], we can estimate sample size using PROC POWER with the ONESAMPLEMEANS statement.
Listing 1: Compute Upper and Lower Confidence Limits with PROC FREQ

libname rebuilt /OABI_PM/eprpsamp/;
%let binstats=BIN_ L_BIN U_BIN N;
/* deletes 11 VAMCs with no zero values of imm1a*/
data nzeroimm1; input vamc;
cards;
;
proc sort data=nzeroimm1; by vamc;
run;
proc sort data=rebuilt.fy16_ind_eprp(keep=imm1a vamc where=(imm1a ne .))
   out=sorted2;
by vamc;
run;
data imm1a; merge nzeroimm1(in=cleanit) sorted2 by vamc;
if not cleanit;
run;
proc sort data=imm1a; by vamc;
run;
proc freq data=imm1a noprint;
table imm1a/missing binomial(level=2) cl alpha=0.02 nocum;
by vamc;
output out=freqout binomial;
run;
proc print data=freqout;
var vamc &binstats;
run;

Listing 2: PROC POWER with ONESAMPLEMEANS statement

proc power;
onesamplemeans ci=t
   alpha = 0.02
   halfwidth = 0.01
   probwidth = 0.98
   stddev = 0.11 0.21
   ntotal = .;
plot x = effect min = 0.01 max = 0.10;
run;

The HALFWIDTH statement is used to specify the desired margin of error [25]. Results from Listing 2 are listed in Figure 5 and the Output from the PLOT statement is shown in Figure 6. With X=EFFECT, the parameter plotted on the X axis is the one SAS chose as the best representative of effect size.

In the SAS/STAT 14.1 documentation titled Introduction to Power and Sample Size Analysis [20], we are told:

An analysis of confidence interval precision is analogous to a traditional power analysis, with CI Half-Width taking the place of effect size and Prob(Width) taking the place of power. The CI Half-Width is the margin of error associated with the confidence interval, the distance between the point estimate and an endpoint. The Prob(Width) is the probability of obtaining a confidence interval with at most a target half-width. The POWER procedure performs confidence interval precision analyses for t-based confidence intervals for one-sample, paired, and two-sample designs, and for several varieties of confidence intervals for a binomial proportion.

The confidence interval half-width is more commonly called the margin of error.
On the ONESAMPLEMEANS statement, the CI= option "Specifies an analysis of precision of the confidence interval for the mean". Also on the ONESAMPLEMEANS statement, when the result parameter CI is assigned a value of T (CI=T), SAS solves for either Prob(width) or sample size. For the former, the syntax is PROBWIDTH= . while NTOTAL= . gives sample size.

Analysis of Confidence Interval Precision for a Single Binomial Proportion

To compute confidence interval precision for a single binomial proportion with PROC POWER, I used the analysis statement ONESAMPLEFREQ. The code is shown in Listing 3. The binomial proportion value of 0.90671 came from the overall proportion of all 5,767 observations analyzed in Listing 1.

Listing 3: PROC POWER with ONESAMPLEFREQ statement

```plaintext
proc power;
onesamplefreq ci=wilson
alpha=0.02
halfwidth=0.01
probwidth=.
proportion=0.90671
ntotal=731;
run;
```

Figure 8 gives the results obtained from executing PROC POWER with the ONESAMPLEFREQ statement. The resulting Prob(Width) of < 0.001 indicates the estimate is highly unlikely to fall within the extremely narrow specified margin of error, 0.01.
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 [19]. In his slide in Figure 3, 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 [34]. This is not science. For hospitals with a population sizes of 78 to 390 and 1551 or greater, we 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 [36]. Except possibly VHA, it is not in the interest of ORYX vendors to raise questions about sample sizes.

SAS can be used to develop better sample size estimates for healthcare quality measurement than those improvised by TJC. Unlike the principal federal statistical agencies which must plan precision of estimates, CMS escapes regulation by the Office of Management and Budget [31]. TJC slide [34] in Figure 7 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.
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).

The POWER Procedure
Wilson Score Confidence Interval for Binomial Proportion

<table>
<thead>
<tr>
<th>Fixed Scenario Elements</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Exact</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.02</td>
</tr>
<tr>
<td>Binomial Proportion</td>
<td>0.90671</td>
</tr>
<tr>
<td>CI Half-Width</td>
<td>0.01</td>
</tr>
<tr>
<td>Total Sample Size</td>
<td>731</td>
</tr>
<tr>
<td>Number of Sides</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computed Prob(Width)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob(Width)</td>
</tr>
<tr>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Figure 7: Sampling Methodology The Joint Commission March 2016 Discussion. Slide 4.

Figure 8: ONESAMPLEFREQ Results Listing
Conclusion

CMS and TJC have not yet improved their specifications for sample size estimation [3].

More skepticism should be directed at analyses of VHA performance measurement data comparing the quality of VHA healthcare with private sector healthcare [10]. Not surprisingly, RAND found variable performance among VHA healthcare facilities.

The slide deck TJC [34] prepared for the conference call [17] 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 3, 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 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.

In his textbook on study planning, Geoff Cumming gives a method for planning research that allows computation of confidence interval precision without specifying the Type I error rate, \( \alpha \) [11]. When PROC POWER is used with either the ONESAMPLEMEANS or the ONESAMPLEFREQ statement, SAS uses a default value of ALPHA even if the SAS programmer does not specify a value for ALPHA [22]. Perhaps SAS Institute, Inc. will develop a PROC PRECISION following Cumming’s method.

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.

References


[27] Kate Goodrich MD, 15 April 2016. Centers for Medicare & and Medicaid Services, Center for Clinical Standards and Quality. Letter to Adeline Wilcox, Office of Analytics and Business Intelligence, Veterans Health Administration.


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.

Contact Information

Your comments and questions are valued and encouraged.

Contact the author: Adeline J. Wilcox
10901 Fleetwood Drive
Beltsville, MD 20705-2502
301.537.0931
awilcox1 at mindspring dot com