**A comparison of methods for identifying hospital performance outliers in cardiac surgery**

Short title: Identifying hospital performance outliers

Xiaoting Wu, Ph.D.1; Min Zhang, Ph.D.2; Richard L. Prager, M.D.1,3; Donald S. Likosky, Ph.D.1,3; for the Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative

(1) Department of Cardiac Surgery, University of Michigan, Ann Arbor, Michigan, (2) Department of Biostatistics, University of Michigan; (3) Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative

Presented at the QCOR 2017 Scientific Sessions, Quality of Care and Outcomes Research, Arlington, VA, April 2 – 3, 2017

**Key Words:** Quality improvement; Statistical methodologies; Cardiac surgery

**Direct Correspondence to:**

Donald S. Likosky, Ph.D., Associate Professor, Section Head, Section of Health Services Research and Quality, Department of Cardiac Surgery (5346 CVC), University of Michigan Medical School, Ann Arbor, MI 48109-5864, Email: [likosky@umich.edu](mailto:likosky@umich.edu), 734-232-4216 (P), 734-763-7353 (F)

**Total word count:** 3576

**Subject Terms:** Quality and Outcomes; Cardiovascular Surgery

**Abstract**

**Background:**

A number of statistical approaches have been advocated and implemented to estimate adjusted hospital outcomes for public reporting or reimbursement. The ability of these methods to identify hospital performance outliers in support of quality improvement has not been fully investigated.

**Methods and Results:**

We leveraged data from patients undergoing coronary artery bypass grafting surgery between 2012-2015 at 33 hospitals participating in a statewide quality collaborative. We applied 5 different statistical approaches (1: indirect standardization with standard logistic regression models, 2: indirect standardization with fixed effect models, 3: indirect standardization with random effect models, 4: direct standardization with fixed effect models, 5: direct standardization with random effect models) to estimate hospital post-operative pneumonia rates, adjusting for patient risk. Unlike the standard logistic regression models, both fixed effect and random effect models accounted for hospital effect. We applied each method to each year, and subsequently compared methods in their ability to identify hospital performance outliers.

Pneumonia rates ranged from 0% to 26.2%. The standard logistic regression models for 2013-2015 had c-statistics of 0.73-0.75, fixed effect models had c-statistics of 0.81-0.83, and random effect models had c-statistics of 0.80-0.83. Each method differed in its ability to identify performance outliers. In direct standardization, random effect models stabilized the hospital rates by moving the estimated rates toward the average, fixed effect models produced larger standard errors of hospital effect (e.g., low case volume hospitals). In indirect standardization, the three models showed high agreement on their derived observed/ expected ratio. Indirect standardization with fixed or random effect models identified similar hospital performance outliers in each year.

**Conclusions:**

The surveyed approaches varied in their ability to identify performance outliers. Given its higher sensitivity to outlier hospitals and more stable estimates of hospital effects, indirect standardization methods with random effect models may best support quality improvement activities.

**Introduction**

There has been an increasing movement towards the development of hospital performance standards. National Quality Forum (NQF) is a leader convening both public and private-sector input on developing and endorsing measures to assess healthcare performance. Indeed, the NQF has endorsed several risk-adjusted standards for performance assessment in the fields of cardiac surgery.1 Increasingly these measures are now used for public reporting2-5 and hospital reimbursement.6

Current hospital performance measures utilize a variety of statistical approaches to address confounding at the patient and hospital level. Indirect standardization, which often is expressed as observed/expected (O/E) ratio, standardizes hospital performance with a different hospital case-mix; while direct standardization method standardizes the hospital performance using a reference population.7, 8 These two standardization methods can be combined with different statistical models, such as standard logistic model, fixed effect and random effect models to adjust for patient and/or hospital level confounders.9 Each of these approaches has their strengths and weaknesses, especially as it relates to deriving estimates for public reporting and/or reimbursement. For instance, the random effect models reduced the variation across hospitals by shrinking the hospital effects toward the average hospital effect, particularly for hospitals with small volume.10-13 Nonetheless, reducing the variation across hospitals through shrinkage may not be considered by some as a useful approach for supporting benchmarking quality improvement.

While most of the work to date has focused on the application of these statistical methods for public reporting and reimbursement, few studies have evaluated the application of these methods for assessing performance outliers to support quality improvement. We compared a number of commonly used statistical methods in their ability to identify hospital performance outliers in support of quality improvement. Given recent efforts through our statewide quality collaborative in reducing hospital-acquired infections, we implemented each of these models to compare rates of post-operative pneumonia following isolated coronary artery bypass grafting (CABG) surgery.

**Methods**

This study was approved by the Institutional Review Board at the University of Michigan.

### Patient population

The Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative (MSTCVS-QC) is a surgeon-led multidisciplinary group encompassing all 33 non-federal hospitals performing adult cardiac surgical procedures in the state of Michigan. All programs use the Society of Thoracic Surgeons’ (STS) data collection form and submit data on a quarterly basis to both the STS database and the MSTCVS-QC data warehouse. Routine audits conducted by trained quality collaborative nurses demonstrate excellent data integrity (>98% data accuracy).

We included 20,923 consecutive patients undergoing isolated CABG at any of the 33 centers in the state of Michigan from 2012 - 2015.

**Measures**

The primary outcome for this study was the development of post-operative pneumonia during the index admission. Our group has previously published a risk prediction model for post-operative pneumonia that leverages risk factors known at the time of the operation.5 This model includes the following variables: age, race; pre-operative factors such as hematocrit levels, white blood cell count, bilirubin levels, ejection fraction, dyslipidemia, chronic lung disease, peripheral arterial disease (PVD), cerebrovascular disease (CVD), liver disease, diabetes, home oxygen, history of pneumonia, history of arrhythmia, smoking status within 2 weeks prior to surgery, operative status. This model performs well, as assessed by discrimination and validation.

To assess the effect of hospital volume on the adjusted outcomes, we categorized the 33 hospitals into terciles based on their total case volume in 2013-2015. Hospitals with less than 332 cases within 3 years were considered “small volume” hospitals, hospitals with cases larger than 451 cases were considered “large volume” hospitals, and hospitals with cases number between 332 and 451 were considered “medium volume” hospitals.

**Statistical analysis**

*Risk-adjustment models*

We implemented three different types of models towards risk adjustment (Table 1A) to estimate a patient’s adjusted post-operative pneumonia risk. We considered two levels of effect in our models, patient level effect and hospital level effect. Model type-1, standard logistic regression models, adjusted for patient covariates, and considered no hospital effect. Model type-2, fixed effect models, adjusted for patient covariates and accounted for hospital fixed effects. Model type-3, random effects models, adjusted for patient covariates and accounted for hospital random effects with empirical Bayes estimates.

In order to consider changes in the effect of covariates over time, each of the 3 types of models were run separately in 2013, 2014 and 2015 (Supplemental Table 1A, 1B and 1C). To address potential instability of point estimates, we used two years of data to obtain patient level effects and subsequently estimated hospital level effects for each of the three years (Table 1B). The model performance was assessed by model discrimination (i.e., C-statistic).

*Calculating Rates Using* Direct *and Indirect Standardization*

Direct standardization assumes the same reference case-mix for every hospital, which was the combined patient sample among the 33 hospitals. Bootstrapping 95% CIs14 were constructed to quantify the uncertainty in the direct standardized rates.

The observed to expected (O/E) ratio was calculated through indirect standardization. The observed number of events was the sum of patients identified as having pneumonia within each hospital. The expected number of events for each hospital was the sum of the adjusted patients’ risk within each hospital combined with the median hospital effect calculated from the fixed effect models or the mean hospital effect from the random effect models. The expected number of events for each hospital from standard logistic models was the sum of the adjusted patients’ risk without any hospital effect. Indirect standardized pneumonia rates for each hospital then were calculated by multiplying the hospital specific O/E ratio with the overall pneumonia rate of 33 hospitals. To quantify the consistency among different types of models, the intra-class correlation coefficient (ICC) of O/E ratios derived from different models was calculated. Bootstrapping 95% CIs of O/E ratios were then constructed to identify significant performance outliers (hospitals whose 95% CI did not encompass 1 were considered performance outliers, either negative or positive).

**Results:**

Post-operative pneumonia rates varied across hospitals (min: 0.0%, max: 26.2%) and year: 2013 (overall: 3.2 %, average: 4.3%, min: 0.0%, max: 26.2%), 2014 (overall: 2.4 %, average: 3.3%, min: 0.0%, max: 21.0%), 2015 (overall: 2.2%, average: 2.5%, min: 0.0%, max: 9.8%) (Figure 1A). Pneumonia rate decreased more over time among small hospitals. (Figure 1B)

The models performed well as assessed through the C-statistic. The C-statistics of the standard logistic regression models were 0.73 in the 2013 model, 0.75 in 2014, and 0.73 in 2015; the C-statistics of the fixed effect models were 0.81 in the 2013 model, 0.83 in 2014, and 0.82 in 2015; and the C-statistics of the random effect models were 0.81 in the 2013 model, 0.83 in 2014, and 0.80 in 2015.

*Agreement from three models with indirect standardized rates:*

The intra-class correlation coefficients revealed similarity among the three models based on their derived O/E ratio: 2013 (ICC = 0.98), 2014 (ICC = 0.99), 2015 (ICC=0.97). The O/E ratio derived from fixed effect model and random effect model showed a minor deviation from O/E ratio derived from the standard logistic regression models in 2015, while showing high agreements in 2013 and 2014 (Figure 2).

*Identifying Performance Outliers*

We assessed the ability of the different methods to identify performance outliers. Figure 3 displays the distribution of standardized rates by various methods in 2015. Each triangle signifies a hospital, with the position of each triangle representing the hospital’s ranking. Of all of the surveyed models, the direct standardized method with random effect model displayed the narrowest range of standardized pneumonia rates. We considered hospitals with standardized rates below 25th percentiles or above 75th percentiles as potential performance outliers. The direct standardized rates from the random effect model had 5 hospitals above the 75th percentile, while indirect standardized rates with various methods had 8 hospitals above the 75th percentile. In addition, unlike other surveyed models, the direct standardized method with random effect model moved higher performing hospitals (i.e., those with an observed pneumonia rate of 0.0%) towards the average hospital rate. Hospitals with smaller case volumes moved quantitatively closer to the average compared to hospitals with higher case volumes. As an example, a small volume hospital (e.g., hospital id S-23) moved quantitatively more toward the average compared to a larger hospital (e.g., hospital id L-29), Figure 3. The distribution of indirect standardized rates in 2013 and 2014 represent similar patterns (Supplemental Figure 1A and 1B, respectively).

The indirect standardized rates derived from the fixed effect and random effect models had larger variance, which facilitated differentiation of hospital performance. Since the standard logistic model did not capture the hospital effect and the hospital effect in the fixed effect model was highly unstable with larger standard errors (Supplemental Table 1A, 1B and 1C), the random effect model with indirect standardized method was used to further identify significant hospital performance outliers. Eight hospitals were identified as potential performance outliers above the 75th percentile in 2015 when using the random effect model with indirect standardization (Figure 3). Two out of these 8 potential hospital performance outliers performed significantly worse than the average hospital (Figure 4), given their confidence intervals did not overlap 1. Seven hospitals were identified as potential performance outliers below the 25th percentile; all of these hospitals performed significantly better than the average hospital (Figure 4).

**Discussion:**

We leveraged a statewide clinical database of 33 hospitals participating in a surgeon-led quality collaborative to evaluate five statistical methods in their ability to identify hospital performance outliers for benchmarking quality improvement. We found that direct standardized rates from the fixed effect models were sensitive to low hospital sample size, while direct standardized rates from the random effect models provided stable hospital rates at the cost of reducing variability in hospital performance. Our findings suggest that indirect standardization, especially the random effect model, facilitates identification of hospital performance outliers for supporting benchmarking quality improvement, given it retains sufficient variance in hospital standardized rates. Last, our approach facilitates benchmarking quality improvement activities by identifying hospitals that consistently performed better or worse than the average over time. Our results provide valuable information of sites with consistent performance to support quality improvement activities.

We acknowledge some limitations inherent in our study. First, as with any observational cohort study, our findings are subject to unmeasured confounding. In the case of indirect standardization, rates are only comparable between hospitals if the case-mix is comparable between hospitals. We recognize that some important patient-level factors may not have been accounted for and thus some hospitals may be falsely identified as being performance outliers. Second, we recognize that our findings may only be generalizable to our participating centers; nonetheless, our study represents data from all 33 programs performing cardiac surgery in the state of Michigan. We would expect a greater number of performance outliers when applied to a larger sample of hospitals, including those with smaller case volumes. Last, we applied our models to only one single outcome measure (i.e., pneumonia). Further work is warranted to survey these methods across other outcomes with varying baseline rates.

Our present study builds upon the existing literature that has focused much of its work on utilizing risk adjustment for hospital public reporting and reimbursement. The National Quality Forum1, 15 (NQF) has served as the national clearing house for reviewing and endorsing performance standards for Medicare and other payers. The Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database is the largest clinical database for cardiac surgery, and to date has partnered with biostatisticians to develop hospital performance measures; the STS currently has the most NQF-endorsed measures of any professional society.1, 15, 16 Dimick et al., utilizing the American College of Surgeons' National Surgical Quality Improvement Program, demonstrated the value of utilizing reliability adjustment to remove statistical noise in adjusted hospital outcomes reporting.10 Glance et al. used data from the New York State Cardiac Surgery Database to evaluate different risk adjustment models for identifying performance outliers, including standard logistic regression, fixed-effects or random-effects modeling.9 Glance found that the random-effects models identified fewer performance outliers than the fixed effects or standard logistic regression approaches, in part due to the use of shrinkage estimators. Our study demonstrated a reduction of variability of direct standardized hospital pneumonia rates when using the shrinkage estimates from random effects models. Such shrinkage effect underestimated the difference among hospital performance, in particularly for smaller hospitals. Noticeably, we only used the mean hospital effect instead of each hospital’s effect estimated from the random effect models in the indirect standardization. Thus, the indirect standardization method combined with the random effect models did no shrink the rates toward average. Statistical methods aimed at supporting benchmarking quality improvement benefit from balancing both hospital-level variability in performance and development of accurate estimates of hospital performance.

Among the three types of statistical models we explored, the fixed effect and random effect models appeared to perform better, as assessed through the C-statistic, than standard logistic regression models. The fixed effect and random effect models additionally suggest that some variation in pneumonia rates may be explained by hospital level factors, given that patients are clustered within hospitals. Random effect models correctly estimate patient-level factors only when patient risk factors are not correlated hospital-level effects. Fixed hospitals effects allow unbiased estimates of hospital effects in the case when hospital effects are associated with patient risk factors. Unlike prior reports,17 we obtained consistent results of patient-level effects from both the fixed effect vs. random effect models. For example, the O/E ratios derived from random effects models were quantitatively similar to the O/E ratios derived from fixed effects models, which indicates that the hospital-level effects were unlikely correlated with the patient-level factors. The results from the fixed effect models thus support the results from the random effect models.

Each of the surveyed methods offer advantages (and disadvantages) related to risk adjustment. It should be emphasized that the goal of this study was to identify performance outliers. The five statistical approaches varied in their ability to identify performance outliers. The intended purpose of our study was to review which of the commonly used statistical approaches may best meet the needs of benchmarking quality improvement, namely choosing hospitals that consistently perform better than average for site visiting.18 In the MSTCVS-QC, our collaborative identifies hospitals that would benefit from being visited (due to lower or higher than expected performance). We anticipate utilizing the findings from this current study to more reliably estimate which hospitals are true performance outliers. Given its higher sensitivity to outlier hospitals and more stable estimates of hospital effect, indirect standardization methods with random effect models may be best suited to support these quality improvement activities.

**Acknowledgments**: The authors acknowledge the contribution of Mr. QIxing Liang, BS.

**Sources of Funding:** Dr. Likosky is supported in part by grant number R01HS022535 from the Agency for Healthcare Research and Quality (AHRQ). The opinions expressed in this document are those of the authors and do not reflect the official position of the AHRQ or the U.S. Department of Health and Human Services.

Support for the MSTCVS Quality Collaborative is provided by the Blue Cross and Blue Shield of Michigan and Blue Care Network as part of the BCBSM Value Partnerships program.

**Disclosures**

Although Blue Cross Blue Shield of Michigan and MSTCSV-QC work collaboratively, the opinions, beliefs and viewpoints expressed by the author do not necessarily reflect the opinions, beliefs and viewpoints of BCBSM or any of its employees.

**Reference**

1. National Quality Forum. National Voluntary Consensus Standards for Cardiac Surgery.

2. Leonardi MJ, McGory ML and Ko CY. Publicly available hospital comparison web sites: determination of useful, valid, and appropriate information for comparing surgical quality. *Archives of surgery*. 2007;142:863-8; discussion 868-9.

3. Rothberg MB, Morsi E, Benjamin EM, Pekow PS and Lindenauer PK. Choosing the best hospital: the limitations of public quality reporting. *Health affairs*. 2008;27:1680-7.

4. HANYS Quality Institute. HANYS' Report on Report Cards: Understanding Publicly Reported Hospital Quality Measures. 2013.

5. Strobel RJ, Liang Q, Zhang M, Wu X, Rogers MA, Theurer PF, Fishstrom AB, Harrington SD, DeLucia A, 3rd, Paone G, Patel HJ, Prager RL, Likosky DS, Michigan Society of T and Cardiovascular Surgeons Quality C. A Preoperative Risk Model for Postoperative Pneumonia After Coronary Artery Bypass Grafting. *The Annals of thoracic surgery*. 2016;102:1213-9.

6. Centers for Medicare & Medicaid Services. Hospital Inpatient Quality Reporting Program

7. Nicholl J, Jacques RM and Campbell MJ. Direct risk standardisation: a new method for comparing casemix adjusted event rates using complex models. *BMC medical research methodology*. 2013;13:133.

8. Pouw ME, Peelen LM, Lingsma HF, Pieter D, Steyerberg E, Kalkman CJ and Moons KG. Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization. *PloS one*. 2013;8:e59160.

9. Glance LG, Dick A, Osler TM, Li Y and Mukamel DB. Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. *Medical care*. 2006;44:311-9.

10. Dimick JB, Ghaferi AA, Osborne NH, Ko CY and Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Annals of surgery*. 2012;255:703-7.

11. Dimick JB, Staiger DO and Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health services research*. 2010;45:1614-29.

12. Thomas N, Longford NT and Rolph JE. Empirical Bayes methods for estimating hospital-specific mortality rates. *Statistics in medicine*. 1994;13:889-903.

13. Silber JH, Rosenbaum PR, Brachet TJ, Ross RN, Bressler LJ, Even-Shoshan O, Lorch SA and Volpp KG. The Hospital Compare mortality model and the volume-outcome relationship. *Health services research*. 2010;45:1148-67.

14. Carpenter J and Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine*. 2000;19:1141-64.

15. Shahian DM, Edwards FH, Ferraris VA, Haan CK, Rich JB, Normand SL, DeLong ER, O'Brien SM, Shewan CM, Dokholyan RS, Peterson ED and Society of Thoracic Surgeons Quality Measurement Task F. Quality measurement in adult cardiac surgery: part 1--Conceptual framework and measure selection. *The Annals of thoracic surgery*. 2007;83:S3-12.

16. O'Brien SM, Shahian DM, DeLong ER, Normand SL, Edwards FH, Ferraris VA, Haan CK, Rich JB, Shewan CM, Dokholyan RS, Anderson RP and Peterson ED. Quality measurement in adult cardiac surgery: part 2--Statistical considerations in composite measure scoring and provider rating. *The Annals of thoracic surgery*. 2007;83:S13-26.

17. He K, Kalbfleisch JD, Li Y and Li Y. Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime data analysis*. 2013;19:490-512.

18. Prager RL, Armenti FR, Bassett JS, Bell GF, Drake D, Hanson EC, Heiser JC, Johnson SH, Plasman FB, Shannon FL, Share D, Theurer P, Williams J, Michigan Society of T and Cardiovascular S. Cardiac surgeons and the quality movement: the Michigan experience. *Seminars in thoracic and cardiovascular surgery*. 2009;21:20-7.

**Figure Legend:**

Figure 1. Observed pneumonia rates isolated coronary artery bypass graft. The line in the middle of each box is the median. The box represents the middle 50% of the data. The box edges are the 25th and 75th percentiles. (A) Overall observed pneumonia rates among 2013-2015. (B) Observed pneumonia rates among 2013-2015 by terciles of hospital case volume.

Figure 2: Scatter plots of observed-to-expected (O/E) ratios as a function of surveyed statistical models. The 45° line represents the line of identity.

Figure 3: Distribution of estimated pneumonia rates across statistical approaches. The data represent procedures performed during 2015. The line in the middle of each box is the median. The box represents the middle 50% of the data. The box edges are the 25th and 75th percentiles. Each triangle represents a hospital. Hospitals outside of the box are considered potential performance outliers. S-23 indicates a small case volume hospital (hospital ID 23); L-29 indicates a larger case volume hospital (hospital ID 29).

Figure 4: Forest plot for the hospitals at the upper and lower 25 percentiles of indirect standardized rates from random effect models. The vertical line is the reference line for O/E =1. Black dots: O/E ratios; Grey lines: bootstrapping 95 % confidence intervals (Cl).

Figure 1A

Observed Hospital Pneumonia Rates

2013

2014

2015

0.00

0.05

0.10

0.15

0.20

0.25

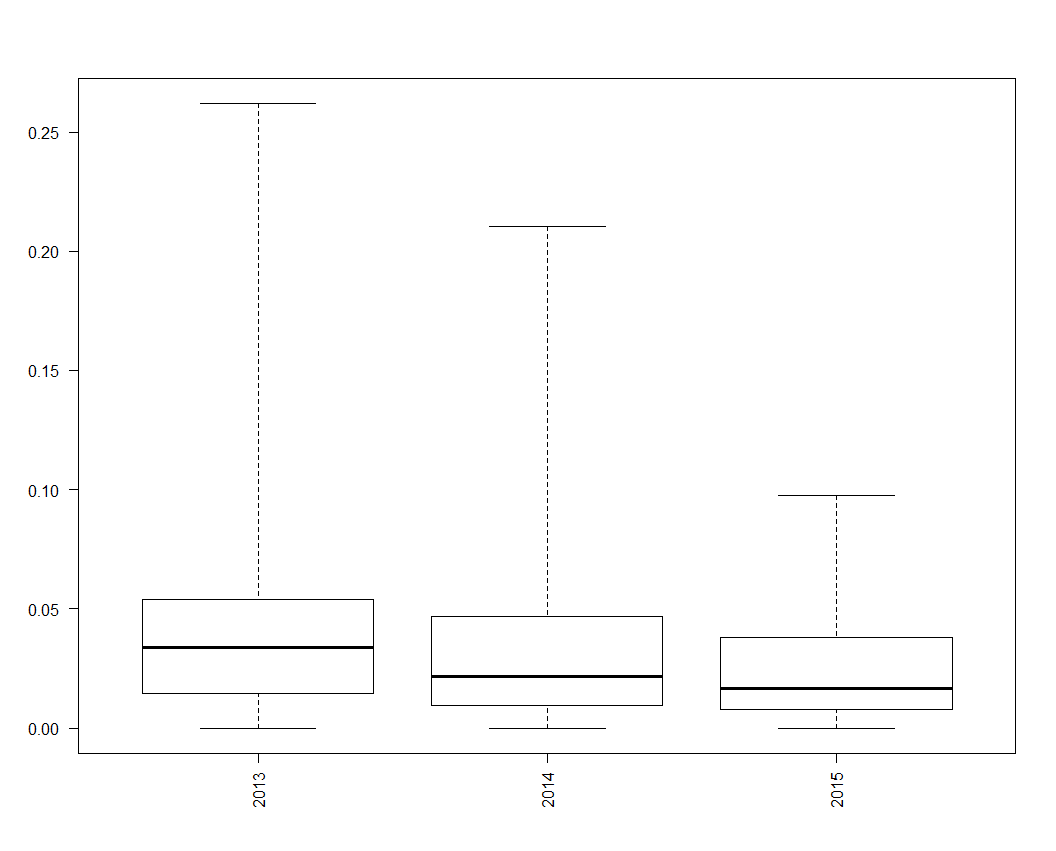


Figure 1B

Small hospital

Medium hospital

Large hospital

Observed Hospital Pneumonia Rates

2013

2014

2015

2013

2014

2015

2013

2014

2015

0.00

0.05

0.10

0.15

0.20

0.25

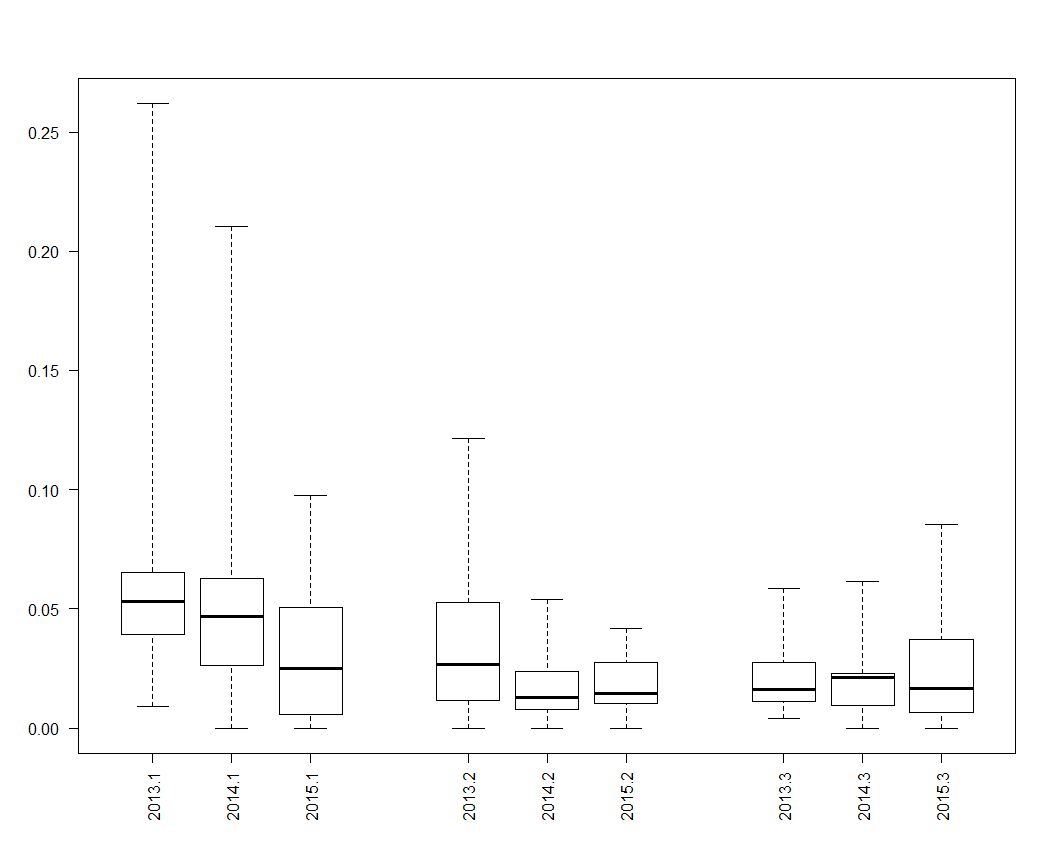
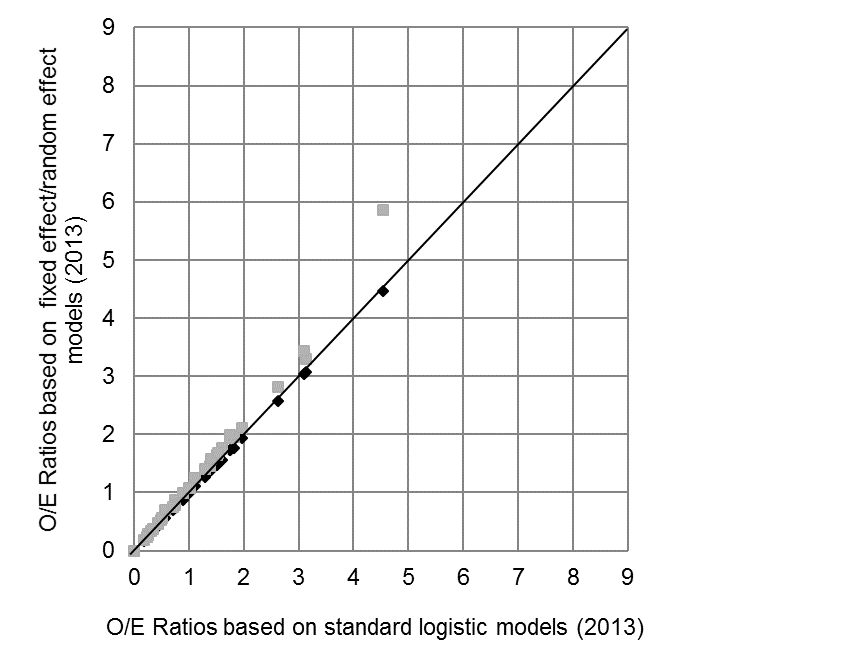


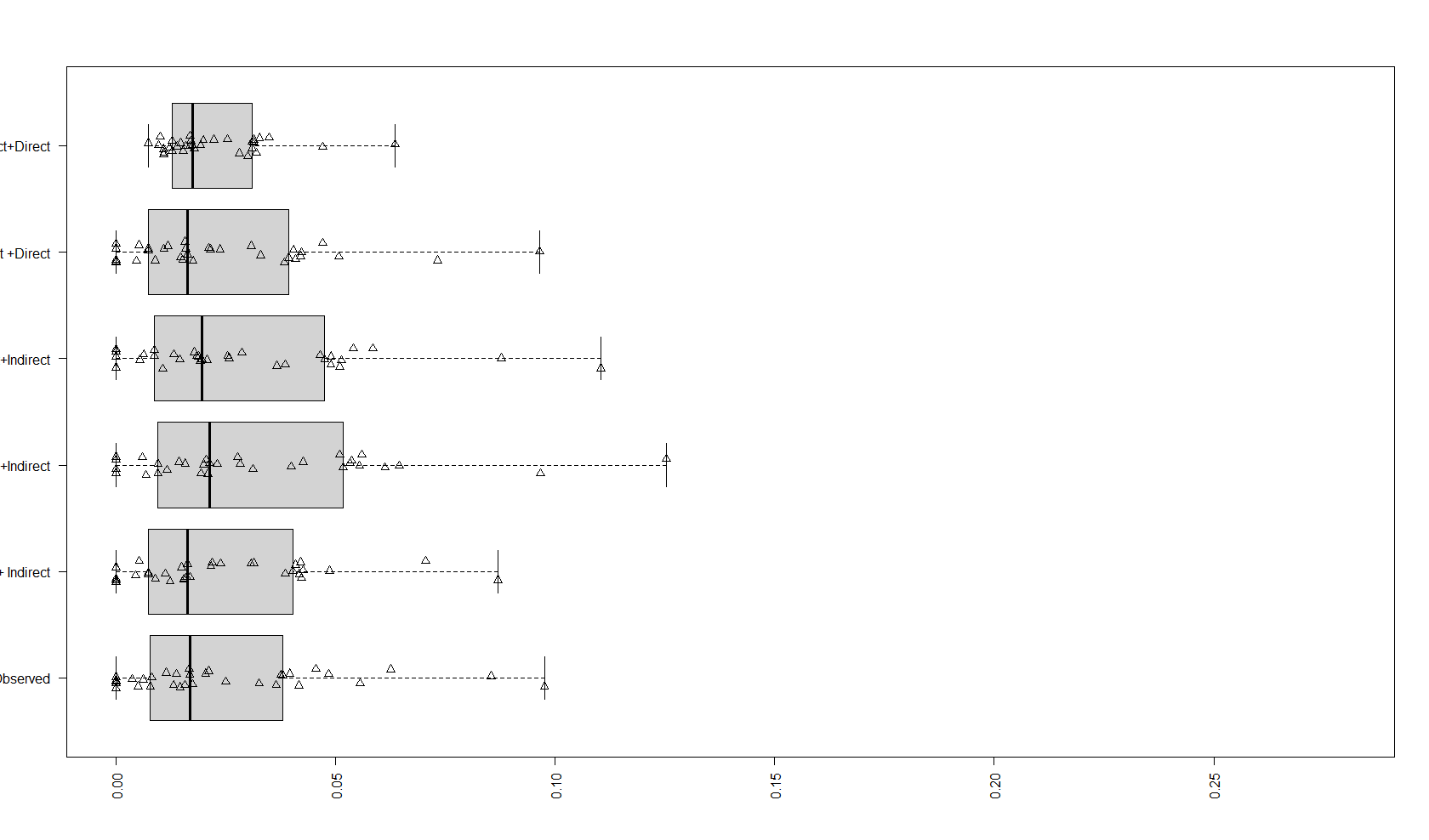
Figure 2



Fixed Effect vs. Standard Logistic

Random Effect vs. Standard Logistic

Figure 3.



Observed

Indirect

Direct

Random effect

Fixed effect

Random effect

Fixed effect

Logistic

Pneumonia rate (2015)

S-23

L-29

S-23

L-29

S-23

L-29

S-23

L-29

S-23

L-29

S-23

L-29

0.00

0.05

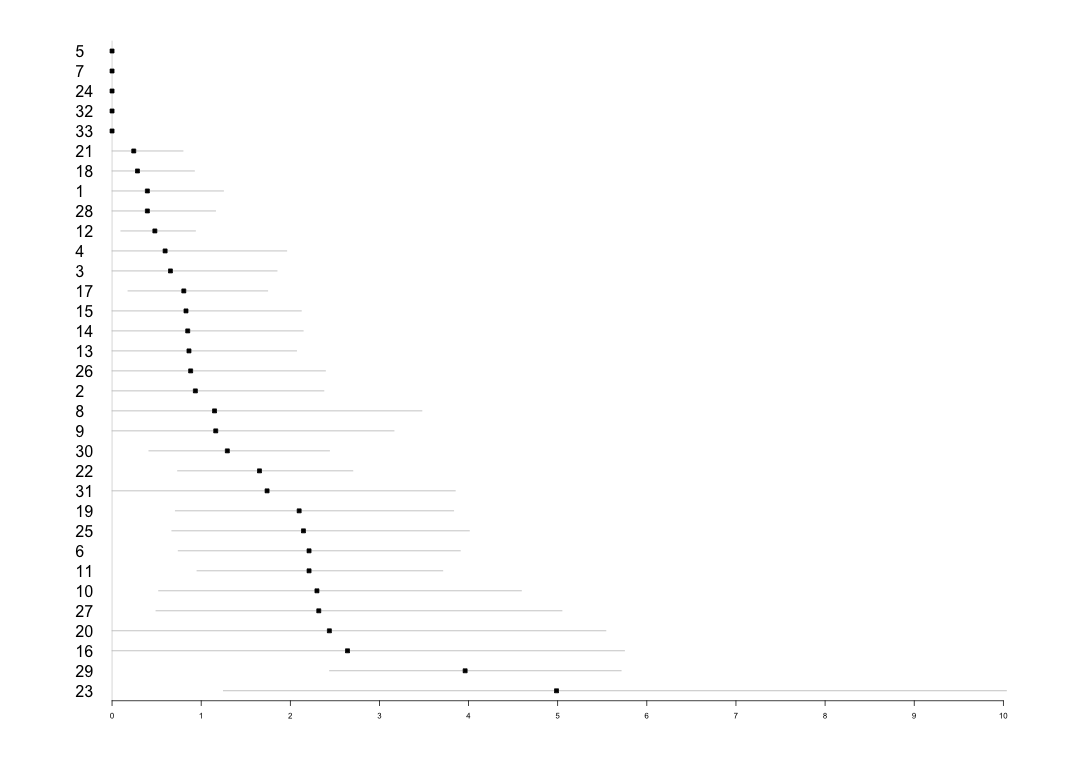
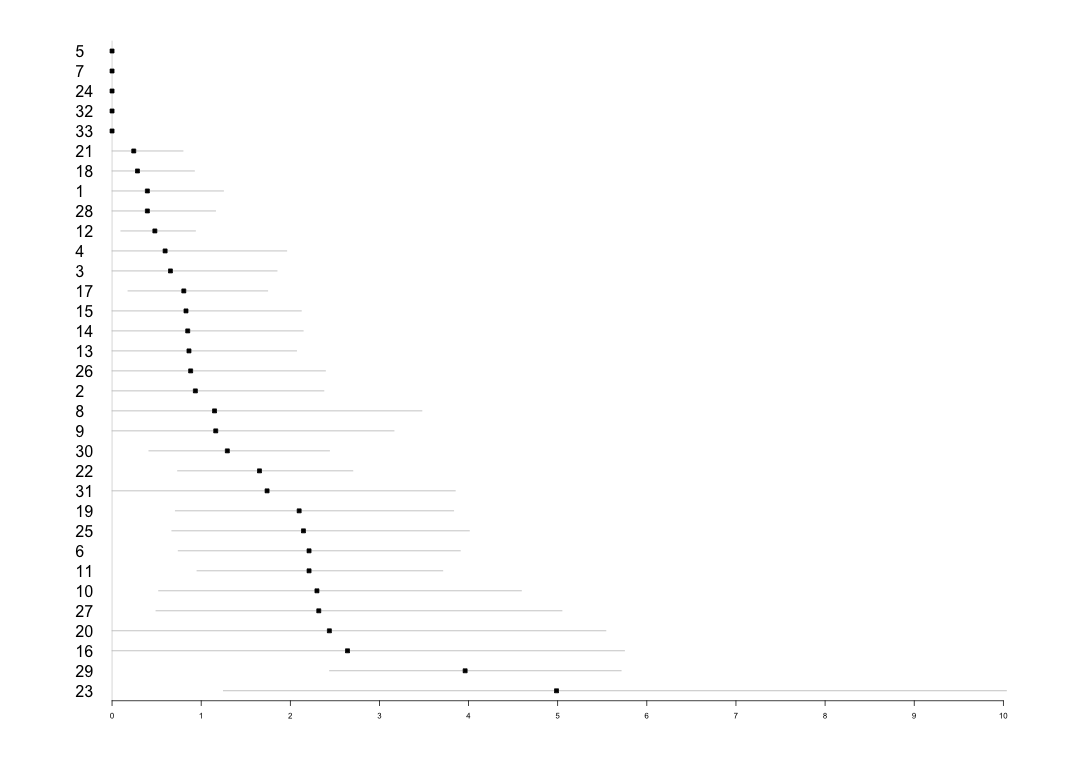
0.10

0.15

0.20

0.25

Figure 4:



0

1

2

3

4

5

6

7

8

9

10

...

Upper 25 percentiles

Lower 25 percentiles

ID

O/E ratios with 95 % Cl (2015)

5

7

24

32

33

21

18

6

11

27

20

16

29

10

23

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1A: Methods for identifying hospital performance outliers** | | | |
| **Model Type** | **Standardized method** | | **Measures** |
| 1- Standard logistic regression | Indirect | | O/E ratios, rates and Bootstrapping CI\* |
| 2- Fixed effects model | Indirect | | O/E ratios, rates and Bootstrapping CI |
| Direct | | Rates |
| 3- Random effects model | Indirect | | O/E ratios, rates and Bootstrapping CI |
| Direct | | Rates |
|  | |  |  |
| **Table 1B: Data sources for models** | | | |
| **Models** | **Patient level effect** | | **Hospital level effect** |
| Model -2013 | Data from 2012-2013 | | Data from 2013 |
| Model -2014 | Data from 2013-2014 | | Data from 2014 |
| Model -2015 | Data from 2014-2015 | | Data from 2015 |

\* CI indicates confidence interval; O/E, observed-to-expected.

**SUPPLEMENTAL MATERIAL**

Method: Statistical Approaches Used in this Paper

Figure 1A: Distribution of estimated pneumonia rates across statistical approaches for procedures performed during 2013.

Figure 1B: Distribution of estimated pneumonia rates across statistical approaches for procedures performed during 2014.

Table 1A: Model estimates for 2013 models

Table 1B: Model estimates for 2014 models

Table 1C: Model estimates for 2015 models

**Statistical Approaches Used in this Paper**

Standard logistic regression

The standard logistic regression model is the simplest model used to calculate risk-adjusted rate. For the *j*th patient in *i*th hospital, the logistic model is

where xijk is the value of *k*th covariate for the *j*th patient in *i*th hospital. This model assumes the effects of covariates and also the baselines are the same across hospitals. To calculate the risk-adjusted rate for each hospital, first, the predicted probabilities of event for patients in each hospital were summed to obtain the expected number of events. Second, the risk-adjusted rate for each hospital is calculated by taking the ratio of the observed number of events and expected number of events (O/E) in each hospital and then multiplying by the overall event rate. This is known as the indirect standardization.

Fixed effect model

For the *j*th patient in *i*th hospital, the logistic model is

where I (hospitali) is an indicator variable for *i*th hospital. In this model, we consider the hospital effects as fixed effects. Different from Method 1, we do not assume the baselines across hospitals are the same, although we still assume the effects of covariates are the same across hospitals.

The indirect standardization method from the fixed effect models was performed as following. The expected rates for specific hospital were calculated assuming the patient in this hospital was treated by the hospital with a “median” performance. We then calculated the indirect standardized rates by multiplying the O/E ratio by the overall rate.

Direct standardization method was performed by the following steps: the estimated model was applied to the same reference hospital population. We assumed the reference population, which was the combination of all patients in the 33 hospitals, was served by one specific hospital at each time. Then the expected hospital rate for this hospital was obtained by summing the expected rates for all patients and then dividing by the total number of the patients in the reference hospital.

Random hospital effect model

where γi represents the random intercept effect for the *i*th hospital. In this mixed effect model, we assume the hospital effects to be random effects which follow a normal distribution. Empirical Bayes method was used to estimate this model. The risk-adjusted rates were calculated using either direct or indirect standardization based on this model. In indirect standardization, the expected number of events for each hospital was the sum of the adjusted patients’ risk within each hospital combined with the grand mean hospital effect from the random effect models.

2. Direct and Indirect standardization

Let N be total number of patients from 33 hospitals, ni be the number of

patients in hospital i, R be overall observed pneumonia rate in 33 hospitals, pj

be the number of observed pneumonia cases in hospital i:

2.1 Indirect standardization:

Indirect standardized rate can be calculated from Model 1, 2, 3.

Indirect standardized pneumonia rate for hospital i =

Oi : observed number of pneumonia cases in hospital i

Ei : expected number of pneumonia cases in hospital i , Ei =

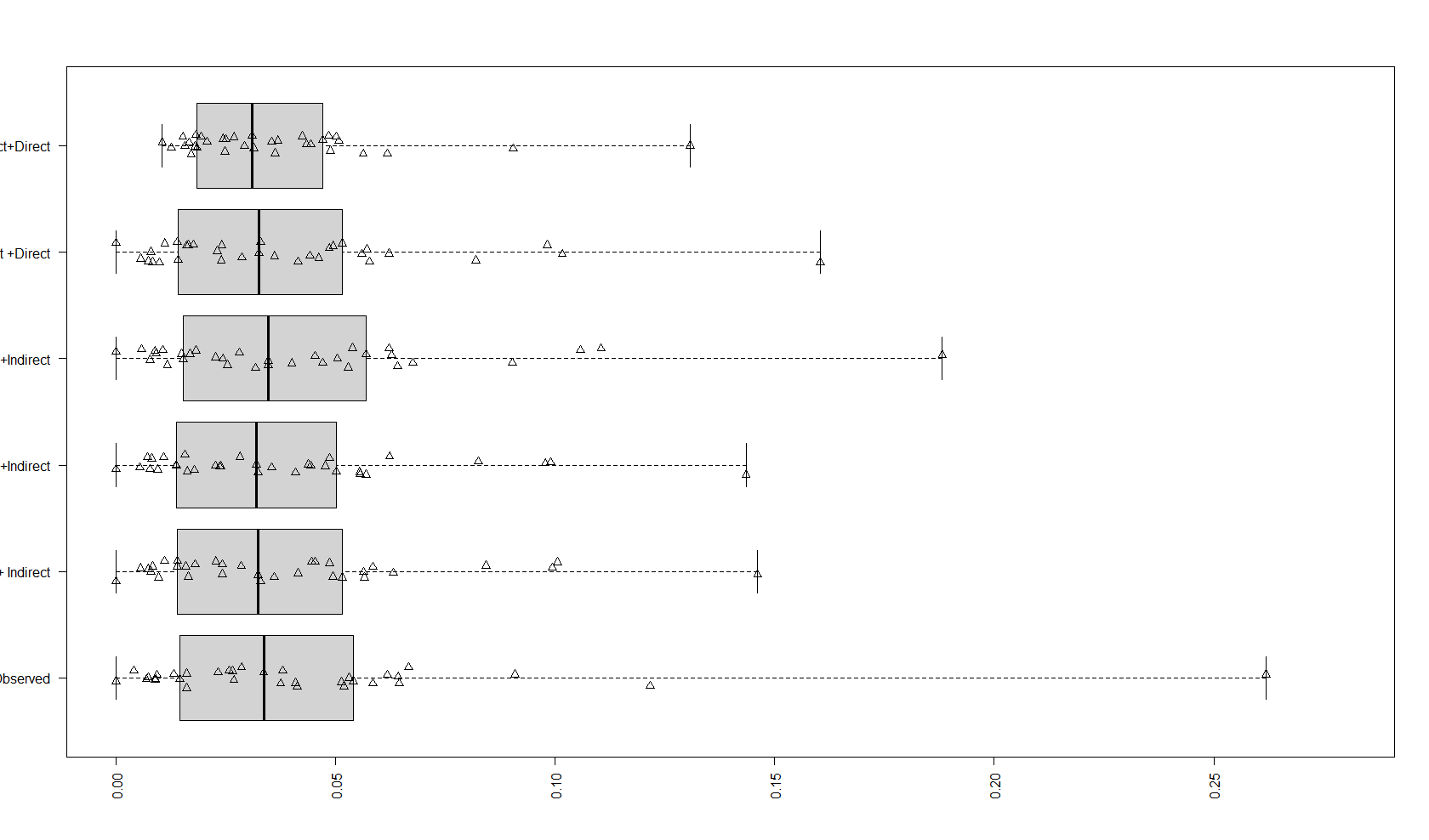
2.2 Direct standardization:

Direct standardization assume all patients go to a same hospital. Direct

standardized rate can be calculated from Model 2 and Model 3.

Direct standardized pneumonia rate for hospital i =

Supplemental Figure 1A: Distribution of estimated pneumonia rates across statistical approaches for procedures performed during 2013. The line in the middle of each box is the median. The box edges are the 25th and 75th percentiles. Each triangle represents a hospital. Hospitals outside of the box are considered potential performance outliers.



Observed

Indirect

Direct

Random effect

Fixed effect

Random effect

Fixed effect

Logistic

Pneumonia rate (2013)

0.00

0.05

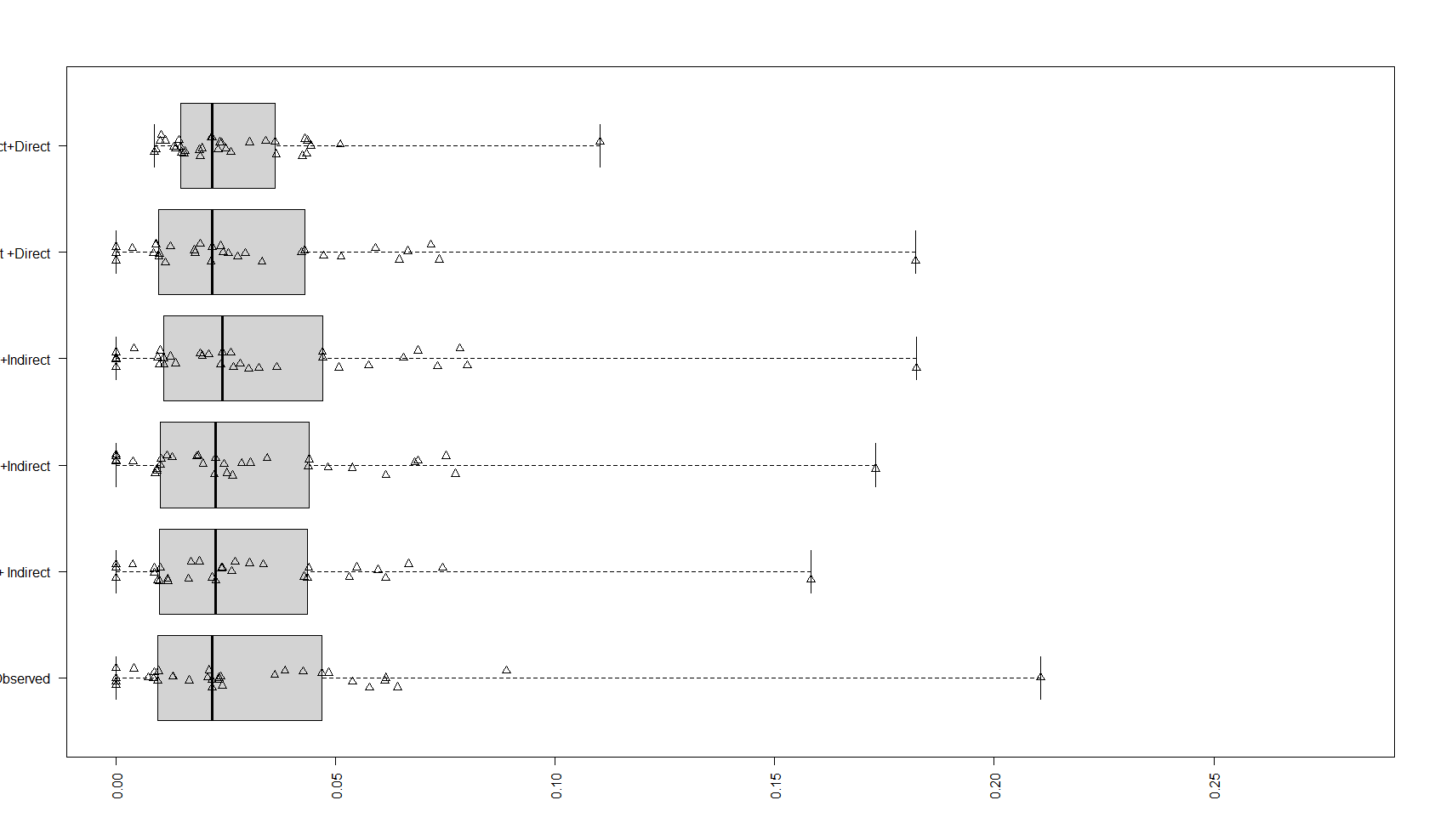
0.10

0.15

0.20

0.25

Supplemental Figure 1B: Distribution of estimated pneumonia rates across statistical approaches for procedures performed during 2014. The line in the middle of each box is the median. The box edges are the 25th and 75th percentiles. Each triangle represents a hospital. Hospitals outside of the box are considered potential performance outliers.



Observed

Indirect

Direct

Random effect

Fixed effect

Random effect

Fixed effect

Logistic

Pneumonia rate (2014)

0.00

0.05

0.10

0.15

0.20

0.25

Supplemental Table 1A: Model estimates for 2013 models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | 2013 | | | | | |
| Model Types | Logistic Model Rate Calculation | | Fixed Effect Model Rate Calculation | | Random Effect Model Rate Calculation | |
| C-statistics | 0.727 | | 0.808 | | 0.805 | |
| Model Parameters | Estimate | Std. | Estimate | Std. | Estimate | Std. |
| (Intercept) | -1.384 | 0.079 | -2.521 | 1.007 | -2.770 | 0.630 |
| Age | -0.057 | 0.049 | -0.053 | 0.049 | -0.562 | 0.504 |
| Age Squared | 0.001 | 0.000 | 0.001 | 0.000 | 0.751 | 0.500 |
| Hematocrit | -0.016 | 0.010 | -0.018 | 0.011 | -0.018 | 0.011 |
| white blood cell (4,500 cells/µL-10,000 cells/µL vs. Less than 4,500 cells/µL) | 0.186 | 0.333 | 0.132 | 0.337 | 0.141 | 0.336 |
| white blood cell (greater than 10,000 cells/µL vs less than 4,500 cells/µL) | 0.592 | 0.346 | 0.570 | 0.351 | 0.572 | 0.350 |
| Pre-operative intra-aortic balloon pump | 0.472 | 0.179 | 0.460 | 0.186 | 0.455 | 0.185 |
| Ejection Fraction | -0.012 | 0.004 | -0.011 | 0.004 | -0.012 | 0.004 |
| White vs. Non-white | -0.599 | 0.148 | -0.300 | 0.176 | -0.366 | 0.169 |
| Non-elective vs. elective | 0.168 | 0.129 | 0.214 | 0.136 | 0.207 | 0.135 |
| Current cigarette smoking | 0.586 | 0.129 | 0.586 | 0.132 | 0.584 | 0.132 |
| Dyslipidemia | -0.359 | 0.179 | -0.211 | 0.187 | -0.235 | 0.186 |
| Peripheral vascular disease | 0.414 | 0.129 | 0.345 | 0.134 | 0.360 | 0.133 |
| Cerebrovascular disease | 0.353 | 0.131 | 0.345 | 0.133 | 0.347 | 0.132 |
| Diabetes | 0.343 | 0.114 | 0.312 | 0.116 | 0.317 | 0.116 |
| Chronic lung disease (Mild vs. None) | 0.229 | 0.141 | 0.137 | 0.146 | 0.147 | 0.145 |
| Chronic lung disease (Moderate/Severe vs. None) | 0.621 | 0.142 | 0.550 | 0.149 | 0.559 | 0.148 |
| Liver disease | 0.903 | 0.230 | 0.795 | 0.242 | 0.823 | 0.240 |
| Home oxygen therapy | 0.331 | 0.319 | 0.263 | 0.330 | 0.276 | 0.328 |
| History of pneumonia | 0.389 | 0.164 | 0.283 | 0.171 | 0.294 | 0.170 |
| History of Arrhythmia (yes vs. no) | 0.165 | 0.141 | 0.213 | 0.145 | 0.204 | 0.144 |
| Hospital-1 |  |  | ref | ref | -0.488 | 0.525 |
| Hospital-2 |  |  | 0.927 | 1.133 | -0.099 | 0.419 |
| Hospital-3 |  |  | 0.503 | 1.236 | -0.372 | 0.478 |
| Hospital-4 |  |  | 1.675 | 1.082 | 0.411 | 0.363 |
| Hospital-5 |  |  | 1.573 | 1.168 | 0.237 | 0.484 |
| Hospital-6 |  |  | 1.257 | 1.130 | 0.054 | 0.425 |
| Hospital-7 |  |  | 1.621 | 1.087 | 0.386 | 0.369 |
| Hospital-8 |  |  | 2.262 | 1.137 | 0.694 | 0.479 |
| Hospital-9 |  |  | -0.208 | 1.423 | -0.567 | 0.515 |
| Hospital-10 |  |  | 1.740 | 1.113 | 0.433 | 0.421 |
| Hospital-11 |  |  | 0.920 | 1.131 | -0.171 | 0.413 |
| Hospital-12 |  |  | -0.567 | 1.232 | -1.064 | 0.420 |
| Hospital-13 |  |  | 0.355 | 1.165 | -0.503 | 0.424 |
| Hospital-14 |  |  | -0.172 | 1.425 | -0.592 | 0.513 |
| Hospital-15 |  |  | 1.865 | 1.094 | 0.528 | 0.393 |
| Hospital-16 |  |  | 2.473 | 1.141 | 0.796 | 0.491 |
| Hospital-17 |  |  | -0.292 | 1.233 | -0.879 | 0.434 |
| Hospital-18 |  |  | 0.373 | 1.164 | -0.517 | 0.422 |
| Hospital-19 |  |  | 1.856 | 1.075 | 0.567 | 0.351 |
| Hospital-20 |  |  | 3.079 | 1.071 | 1.669 | 0.355 |
| Hospital-21 |  |  | 0.534 | 1.105 | -0.440 | 0.367 |
| Hospital-22 |  |  | 1.696 | 1.035 | 0.500 | 0.233 |
| Hospital-23 |  |  | 0.602 | 1.246 | -0.203 | 0.502 |
| Hospital-24 |  |  | 0.127 | 1.164 | -0.688 | 0.410 |
| Hospital-25 |  |  | 2.514 | 1.040 | 1.230 | 0.257 |
| Hospital-26 |  |  | 1.111 | 1.131 | -0.008 | 0.422 |
| Hospital-27 |  |  | 1.949 | 1.093 | 0.580 | 0.393 |
| Hospital-28 |  |  | 0.880 | 1.167 | -0.183 | 0.448 |
| Hospital-29 |  |  | 1.355 | 1.059 | 0.195 | 0.304 |
| Hospital-30 |  |  | 1.243 | 1.071 | 0.064 | 0.328 |
| Hospital-31 |  |  | 1.831 | 1.093 | 0.540 | 0.392 |
| Hospital-32 |  |  | 1.503 | 1.170 | 0.216 | 0.484 |
| Hospital-33 |  |  | -12.522 | 374.637 | -0.659 | 0.581 |

Supplemental Table 1B: Model estimates for 2014 models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | 2014 | | | | | |
| Model Types | Logistic Model Rate Calculation | | Fixed Effect Model Rate Calculation | | Random Effect Model Rate Calculation | |
| C-statistics | 0.747 | | 0.834 | | 0.827 | |
| Model Parameters | Estimate | Std. | Estimate | Std. | Estimate | Std. |
| (Intercept) | -0.114 | 0.093 | -16.484 | 785.158 | -2.761 | 0.702 |
| Age | -0.114 | 0.050 | -0.092 | 0.051 | -0.984 | 0.522 |
| Age Squared | 0.001 | 0.000 | 0.001 | 0.000 | 1.271 | 0.516 |
| Hematocrit | -0.024 | 0.011 | -0.030 | 0.012 | -0.029 | 0.012 |
| white blood cell (4,500 cells/µL-10,000 cells/µL vs. Less than 4,500 cells/µL) | 0.301 | 0.396 | 0.356 | 0.401 | 0.346 | 0.400 |
| white blood cell (greater than 10,000 cells/µL vs less than 4,500 cells/µL) | 0.720 | 0.408 | 0.782 | 0.414 | 0.769 | 0.413 |
| Pre-operative intra-aortic balloon pump | 0.808 | 0.176 | 0.853 | 0.186 | 0.832 | 0.184 |
| Ejection Fraction | -0.016 | 0.005 | -0.016 | 0.005 | -0.016 | 0.005 |
| White vs. Non-white | -0.305 | 0.182 | 0.052 | 0.212 | -0.035 | 0.205 |
| Non-elective vs. elective | 0.138 | 0.144 | 0.097 | 0.149 | 0.107 | 0.149 |
| Current cigarette smoking | 0.808 | 0.144 | 0.828 | 0.147 | 0.825 | 0.146 |
| Dyslipidemia | -0.261 | 0.194 | -0.193 | 0.203 | -0.204 | 0.201 |
| Peripheral vascular disease | 0.386 | 0.145 | 0.363 | 0.149 | 0.370 | 0.149 |
| Cerebrovascular disease | 0.182 | 0.145 | 0.188 | 0.149 | 0.183 | 0.148 |
| Diabetes | 0.209 | 0.125 | 0.208 | 0.128 | 0.208 | 0.128 |
| Chronic lung disease (Mild vs. None) | 0.275 | 0.151 | 0.139 | 0.156 | 0.154 | 0.155 |
| Chronic lung disease (Moderate/Severe vs. None) | 0.425 | 0.163 | 0.331 | 0.172 | 0.339 | 0.170 |
| Liver disease | 0.757 | 0.243 | 0.625 | 0.251 | 0.655 | 0.249 |
| Home oxygen therapy | 0.226 | 0.368 | 0.266 | 0.377 | 0.264 | 0.376 |
| History of pneumonia | 0.404 | 0.180 | 0.329 | 0.189 | 0.333 | 0.187 |
| History of Arrhythmia (yes vs. no) | 0.196 | 0.166 | 0.222 | 0.170 | 0.215 | 0.169 |
| Hospital-1 |  |  | ref | ref | -0.783 | 0.588 |
| Hospital-2 |  |  | 15.193 | 785.159 | -0.126 | 0.464 |
| Hospital-3 |  |  | 15.711 | 785.159 | 0.208 | 0.488 |
| Hospital-4 |  |  | -0.149 | 1251.237 | -0.668 | 0.608 |
| Hospital-5 |  |  | 16.099 | 785.158 | 0.491 | 0.465 |
| Hospital-6 |  |  | 15.561 | 785.159 | 0.117 | 0.479 |
| Hospital-7 |  |  | 15.646 | 785.159 | 0.164 | 0.488 |
| Hospital-8 |  |  | 16.717 | 785.159 | 0.750 | 0.555 |
| Hospital-9 |  |  | 14.484 | 785.159 | -0.396 | 0.557 |
| Hospital-10 |  |  | 16.469 | 785.158 | 0.773 | 0.444 |
| Hospital-11 |  |  | 15.841 | 785.158 | 0.367 | 0.362 |
| Hospital-12 |  |  | 14.714 | 785.158 | -0.503 | 0.369 |
| Hospital-13 |  |  | 14.566 | 785.159 | -0.473 | 0.481 |
| Hospital-14 |  |  | 14.423 | 785.159 | -0.415 | 0.555 |
| Hospital-15 |  |  | 14.584 | 785.159 | -0.339 | 0.564 |
| Hospital-16 |  |  | 17.832 | 785.158 | 1.813 | 0.425 |
| Hospital-17 |  |  | 13.598 | 785.159 | -0.878 | 0.499 |
| Hospital-18 |  |  | -0.187 | 1074.300 | -0.929 | 0.566 |
| Hospital-19 |  |  | 14.814 | 785.159 | -0.330 | 0.494 |
| Hospital-20 |  |  | 16.688 | 785.159 | 0.725 | 0.546 |
| Hospital-21 |  |  | 15.512 | 785.158 | 0.100 | 0.372 |
| Hospital-22 |  |  | 15.260 | 785.158 | -0.093 | 0.364 |
| Hospital-23 |  |  | 16.601 | 785.158 | 0.755 | 0.492 |
| Hospital-24 |  |  | 15.398 | 785.158 | 0.023 | 0.351 |
| Hospital-25 |  |  | 16.116 | 785.158 | 0.563 | 0.399 |
| Hospital-26 |  |  | 14.492 | 785.159 | -0.383 | 0.558 |
| Hospital-27 |  |  | 16.566 | 785.158 | 0.928 | 0.384 |
| Hospital-28 |  |  | 15.387 | 785.159 | 0.007 | 0.471 |
| Hospital-29 |  |  | 16.223 | 785.158 | 0.739 | 0.267 |
| Hospital-30 |  |  | 15.486 | 785.158 | 0.085 | 0.373 |
| Hospital-31 |  |  | 15.182 | 785.159 | -0.121 | 0.517 |
| Hospital-32 |  |  | 16.311 | 785.159 | 0.555 | 0.526 |
| Hospital-33 |  |  | -0.181 | 1184.181 | -0.753 | 0.593 |

Supplemental Table 1C: Model estimates for 2015 models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | 2015 | | | | |  |
| Model Types | Logistic Model Rate Calculation | | Fixed Effect Model Rate Calculation | | Random Effect Model Rate Calculation | |
| C-statistics | 0.725 | | 0.818 | | 0.802 | |
| Model Parameters | Estimate | Std. | Estimate | Std. | Estimate | Std. |
| (Intercept) | 0.942 | 0.096 | -0.695 | 1.006 | -3.635 | 0.957 |
| Age | -0.156 | 0.050 | -0.146 | 0.052 | -1.512 | 0.521 |
| Age Squared | 0.001 | 0.000 | 0.001 | 0.000 | 1.658 | 0.524 |
| Hematocrit | -0.044 | 0.012 | -0.043 | 0.013 | -0.044 | 0.013 |
| white blood cell (4,500 cells/µL-10,000 cells/µL vs. Less than 4,500 cells/µL) | 1.253 | 0.717 | 1.305 | 0.721 | 1.291 | 0.720 |
| white blood cell (greater than 10,000 cells/µL vs less than 4,500 cells/µL) | 2.013 | 0.724 | 2.049 | 0.728 | 2.041 | 0.727 |
| Pre-operative intra-aortic balloon pump | 0.764 | 0.199 | 0.840 | 0.206 | 0.824 | 0.204 |
| Ejection Fraction | -0.011 | 0.005 | -0.011 | 0.005 | -0.011 | 0.005 |
| White vs. Non-white | 0.092 | 0.229 | 0.282 | 0.245 | 0.210 | 0.239 |
| Non-elective vs. elective | -0.150 | 0.157 | -0.202 | 0.162 | -0.189 | 0.161 |
| Current cigarette smoking | 0.837 | 0.157 | 0.843 | 0.160 | 0.840 | 0.159 |
| Dyslipidemia | -0.309 | 0.203 | -0.279 | 0.209 | -0.287 | 0.207 |
| Peripheral vascular disease | 0.431 | 0.163 | 0.361 | 0.167 | 0.374 | 0.166 |
| Cerebrovascular disease | 0.200 | 0.157 | 0.219 | 0.161 | 0.211 | 0.160 |
| Diabetes | 0.152 | 0.140 | 0.174 | 0.142 | 0.169 | 0.141 |
| Chronic lung disease (Mild vs. None) | 0.374 | 0.160 | 0.345 | 0.164 | 0.346 | 0.163 |
| Chronic lung disease (Moderate/Severe vs. None) | 0.097 | 0.203 | 0.046 | 0.209 | 0.053 | 0.207 |
| Liver disease | 0.476 | 0.276 | 0.364 | 0.282 | 0.387 | 0.281 |
| Home oxygen therapy | 0.620 | 0.338 | 0.669 | 0.350 | 0.647 | 0.346 |
| History of pneumonia | 0.041 | 0.223 | 0.014 | 0.231 | 0.019 | 0.228 |
| History of Arrhythmia (yes vs. no) | -0.155 | 0.211 | -0.192 | 0.213 | -0.181 | 0.213 |
| Hospital-1 |  |  | ref | ref | -0.379 | 0.522 |
| Hospital-2 |  |  | 0.895 | 1.236 | -0.035 | 0.502 |
| Hospital-3 |  |  | 0.487 | 1.244 | -0.236 | 0.486 |
| Hospital-4 |  |  | 0.404 | 1.427 | -0.198 | 0.550 |
| Hospital-5 |  |  | -14.367 | 1127.284 | -0.425 | 0.599 |
| Hospital-6 |  |  | 1.783 | 1.092 | 0.604 | 0.393 |
| Hospital-7 |  |  | -14.738 | 1093.842 | -0.558 | 0.576 |
| Hospital-8 |  |  | 1.087 | 1.432 | 0.046 | 0.593 |
| Hospital-9 |  |  | 1.109 | 1.238 | 0.076 | 0.516 |
| Hospital-10 |  |  | 1.832 | 1.134 | 0.548 | 0.461 |
| Hospital-11 |  |  | 1.797 | 1.065 | 0.674 | 0.332 |
| Hospital-12 |  |  | 0.190 | 1.103 | -0.554 | 0.356 |
| Hospital-13 |  |  | 0.793 | 1.167 | -0.093 | 0.448 |
| Hospital-14 |  |  | 0.775 | 1.239 | -0.087 | 0.498 |
| Hospital-15 |  |  | 0.746 | 1.239 | -0.099 | 0.496 |
| Hospital-16 |  |  | 2.033 | 1.183 | 0.559 | 0.521 |
| Hospital-17 |  |  | 0.714 | 1.129 | -0.153 | 0.407 |
| Hospital-18 |  |  | -0.336 | 1.421 | -0.543 | 0.500 |
| Hospital-19 |  |  | 1.727 | 1.091 | 0.562 | 0.389 |
| Hospital-20 |  |  | 1.949 | 1.177 | 0.513 | 0.512 |
| Hospital-21 |  |  | -0.478 | 1.421 | -0.625 | 0.490 |
| Hospital-22 |  |  | 1.491 | 1.059 | 0.444 | 0.311 |
| Hospital-23 |  |  | 2.773 | 1.146 | 1.001 | 0.513 |
| Hospital-24 |  |  | -14.340 | 576.936 | -0.945 | 0.513 |
| Hospital-25 |  |  | 1.752 | 1.092 | 0.582 | 0.392 |
| Hospital-26 |  |  | 0.816 | 1.236 | -0.066 | 0.498 |
| Hospital-27 |  |  | 1.826 | 1.131 | 0.549 | 0.459 |
| Hospital-28 |  |  | -0.005 | 1.425 | -0.381 | 0.523 |
| Hospital-29 |  |  | 2.448 | 1.033 | 1.339 | 0.235 |
| Hospital-30 |  |  | 1.212 | 1.089 | 0.197 | 0.368 |
| Hospital-31 |  |  | 1.563 | 1.175 | 0.332 | 0.492 |
| Hospital-32 |  |  | -14.392 | 1461.515 | -0.293 | 0.629 |
| Hospital-33 |  |  | -14.541 | 863.193 | -0.668 | 0.555 |