Chapter 2: Risk Adjustment Summary Report: Hospital Harm – Postoperative Respiratory Failure

Patient Safety Measure Development and Maintenance

American Institutes for Research
University of California Davis
Clinician-Driven Quality Solutions

February 2024

American Institutes for Research
10420 Little Patuxent Parkway, Suite 300
Columbia, MD 21044
+1.443.259.5500 | Fax: +1.443.367.0477
AIR.ORG

Table of Contents

Tables

Exhibits

Submitted To

Centers for Medicare & Medicaid Services (CMS)
Center for Clinical Standards and Quality (CCSQ)

Attention

Donta Henson
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244-1850

Submitted By

American Institutes for Research
10420 Little Patuxent Parkway
Suite 300
Columbia, MD 21044
(443)256-5500
www.air.org

Project

Patient Safety Measure Development and Maintenance
Contract Number: 75FCMC18D0027

Task & Deliverable

Chapter 6 Measure Testing
Deliverable 6-4 Risk Adjustment Methodology Report

Authors

Monika Ray, UC Davis
Patrick Romano, UC Davis
Irina Tokareva, UC Davis
Anna Michie, AIR
Mia Nievera, AIR
Chris Evensen, AIR
Chana West, CDQS

1. Introduction

This project, titled Patient Safety Measure Development and Maintenance, is performed under the Measure & Instrument Development and Support (MIDS) contract for the Centers for Medicare & Medicaid Services (CMS).

The goal of the project is to develop, maintain, re-evaluate, and implement patient safety measures for CMS's hospital-level quality reporting programs. Broadly, these programs range from the Hospital Inpatient Quality Reporting (IQR) Program to Hospital-Acquired Condition (HAC) Reduction Program, and the Promoting Interoperability (PI) Program.

This report is an independent chapter of the overall measure testing report and describes the rationale and methodology for a risk adjustment model for the Hospital Harm – Postoperative Respiratory Failure (HH-PRF) electronic clinical quality measure (eCQM). In this report, we explain risk factors selection, provide the model specification, summarize empirical findings, and present coefficient estimates.

We note that the baseline risk adjustment model presented in this chapter welcomes future update and refinement, as the number of hospitals (or test sites) participating in testing and model development is small. We aim to improve the model by augmenting hospitals and hence enlarging sample size to improve the model's generalizability to the full population in the future (during measure implementation).

2. Measure Description

HH-PRF assesses the proportion of elective inpatient hospitalizations for patients aged 18 years and older without an obstetrical condition who have a procedure resulting in postoperative respiratory failure. The initial population (IP) includes all elective inpatient hospitalizations that end during the measurement period for patients aged 18 and older without an obstetrical condition and at least one surgical procedure was performed within the first 3 days of the encounter, and all payers. The measure denominator population is a subset of the IP, where encounters meeting denominator exclusion criteria are excluded. The measure numerator population is, in turn, a subset of denominator population where the patient has a PRF. To qualify for the measure numerator, one harm event suffices.

Table 1 summarizes the measure's core components without delving into the technical details. For an in-depth review of the measure specification, please refer to **Chapter 1: Measure Testing Summary Report.**

Table 1. Measure Initial Population, Denominator Exclusion, Denominator, and Numerator

Hospital Harm - PRF	Description
Measure IP	All elective inpatient hospitalizations that end during the measurement period for patients aged 18 and older without an obstetrical condition and at least one surgical procedure was performed within the first 3 days of the encounter, and all

Hospital Harm - PRF	Description
	payers. Elective inpatient hospitalizations include time in observation or outpatient surgery when the transition between these encounters (if they exist) and the inpatient encounter are within an hour or less of each other.
Denominator Exclusion	Inpatient hospitalizations for patients: • Who have mechanical ventilation that starts more than one hour prior to the start of the first operating (OR) procedure • With arterial partial pressure of oxygen (PaO2)<50 mmHg within 48 hours or less prior to the start of the first OR procedure • With arterial partial pressure of carbon dioxide (PaCO2)>50 mmHg combined with an arterial pH<7.30 within 48 hours or less prior to the start of the first OR procedure • With a principal diagnosis for acute respiratory failure • With a secondary diagnosis for acute respiratory failure present on admission • With any diagnosis present on admission for the existence of a tracheostomy • Where a tracheostomy is performed before or on the same day as the first OR procedure • With any diagnosis for neuromuscular disorder or degenerative neurological disorder • With any procedure for selected pharyngeal, nasal, oral, facial, or tracheal surgery involving significant risk of airway compromise likely to require prophylactic retention of the endotracheal tube for at least 48 hours
Denominator	Elective inpatient hospitalizations that end during the measurement period for patients aged 18 and older without an obstetrical condition and at least one surgical procedure was performed within the first 3 days of the encounter.
Numerator	Elective inpatient hospitalizations for patients with postoperative respiratory failure (PRF) as evidenced by: **Criterion A:** Mechanical Ventilation (MV) initiated within 30 days after first operating room (OR) procedure, as evidenced by: • A.1. Intubation that occurs outside of a procedural area and within 30 days after the end of the first OR procedure of the encounter; or • A.2. MV that occurs outside of a procedural area within 30 days after the end of the first OR procedure of the encounter and is preceded by a period of non-invasive oxygen therapy between the end of the OR procedure and the MV occurrence, and without a subsequent OR procedure between the non-invasive oxygen therapy and the MV occurrence **or** **Criterion B:** MV with a duration of more than 48 hours after the first OR procedure, as evidenced by: • B.1. Extubation that occurs outside of a procedural area more than 48 hours after the end of an OR procedure and within 30 days after the end of the first OR procedure, and is not preceded by a period of non-invasive oxygen therapy or a subsequent OR procedure between the end of the OR procedure and the extubation occurrence; or • B.2 Mechanical ventilation that occurs outside of a procedural area between 48 and 72 hours after the end of an OR procedure and within 30 days after the end of the first OR procedure, and is not preceded by a non-invasive oxygen therapy or a subsequent OR procedure between the end of the OR procedure and the MV occurrence.

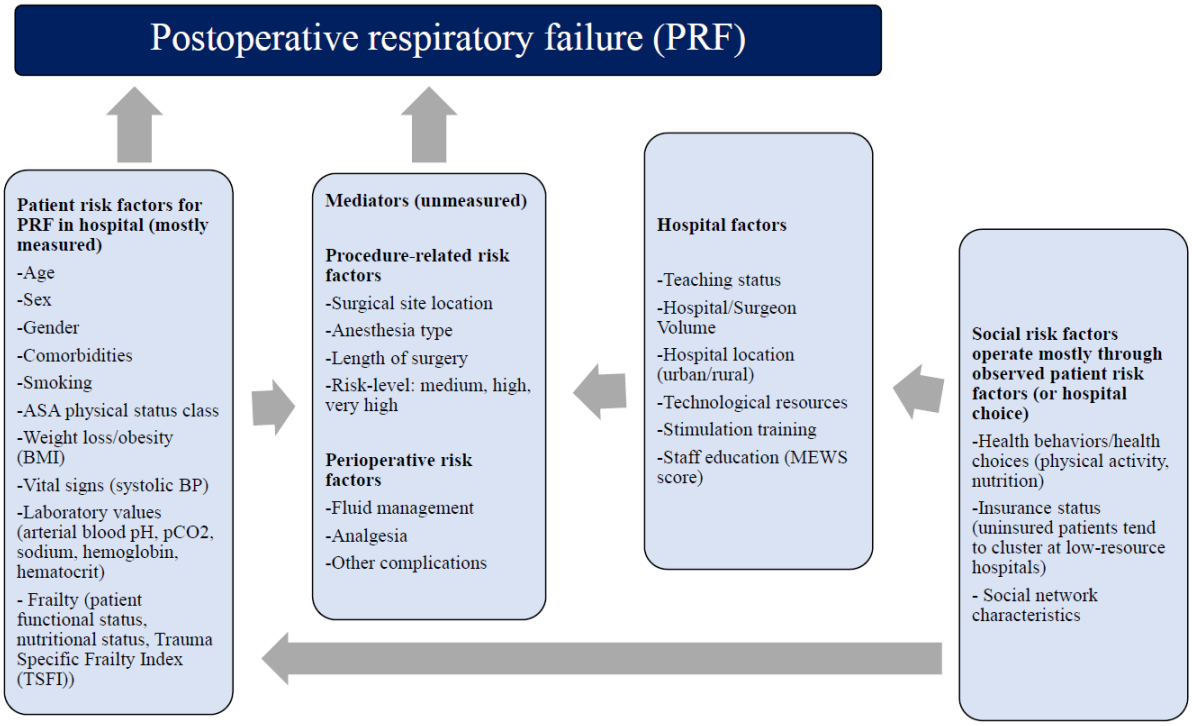
3. Rationale For Risk Adjustment

Extensive research and clinical practice have confirmed that it is not possible to reduce PRF rates to zero; some patients are inherently at higher risk than others, even with the best possible care. As Canet and colleagues have described, "the pathogenesis of PRF depends on factors related to patient status as well as anaesthetic and surgical procedure."[1] In order to permit fair comparisons among hospitals that serve very different patient populations, risk-adjustment is used in the HH-PRF eCQM.

4. Streamlined Conceptual Model

Exhibit 1 is a simplified conceptual framework that guided our risk model development. The conceptual model shows how patient risk factors, intraoperative factors, and perioperative management are thought to interact in contributing to the development of PRF. In the sections below we discuss how each of these components relate to the outcome of interest and whether they belong to the risk model, and if so in what forms.

Exhibit 1. Simplified Conceptual Model That Guided the Risk Adjustment Model Development



[1] Canet, Jaume; Sabaté, Sergi; Mazo, Valentín; Gallart, Lluís; de Abreu, Marcelo Gama; Belda, Javier; Langeron, Olivier; Hoeft, Andreas; Pelosi, Paolo For the PERISCOPE group. Development and validation of a score to predict postoperative respiratory failure in a multicentre European cohort: A prospective, observational study. European Journal of Anaesthesiology 32(7):p 458-470, July 2015. | DOI: 10.1097/EJA.0000000000000223

4.1 Patient Characteristics

Patient specific characteristics that increase the risk of PRF include age, gender, BMI, smoking status, comorbidities such as chronic obstructive pulmonary disease (COPD), American Society of Anesthesiologists (ASA) class,[2] preoperative physiologic characteristics (systolic blood pressure), laboratory values (arterial blood pH, pCO2, sodium, hemoglobin, hematocrit), and complications present on admission. Some of these prior studies are cited in detail below.

Based on data from the American College of Surgeons National Surgical Quality Improvement Program on all elderly vascular and general surgery patients undergoing operations from 2005 to 2008, univariate predictors of unplanned postoperative intubation (UPI) were older age, chronic obstructive pulmonary disease, low pre-operative functional status as well as emergency operation.[3]

Svensson and colleagues (1991) analyzed data from June 1960 to September 1990 on 1414 patients who underwent repair of thoracoabdominal aortic aneurysms.[4] The independent predictors of respiratory failure were chronic pulmonary disease, smoking history, cardiac and renal complications. In patients with chronic pulmonary disease, the only independent predictor was FEF25 ($p = 0.030$).

In a cohort study of 44 VA medical centers (Arozullah, 2000), PRF developed in 2,746 patients (3.4%).[5] The respiratory failure risk index was developed from a simplified logistic regression model and included abdominal aortic aneurysm repair, thoracic surgery, neurosurgery, upper abdominal surgery, peripheral vascular surgery, neck surgery, emergency surgery, albumin level less than 30 g/L, blood urea nitrogen level more than 30 mg/dL, dependent functional status, chronic obstructive pulmonary disease, and age.

Canet and colleagues (2015) reported a prospective observational study of a multicenter cohort and described a predictive score for PRF that includes seven independent risk factors: low preoperative SpO2; at least one preoperative respiratory symptom; preoperative chronic liver

[2] ASA categories are defined here: https://www.asahq.org/standards-and-practice-parameters/statement-on-asa-physical-status-classification-system

[3] Nafiu, O. O., Ramachandran, S. K., Ackwerh, R., Tremper, K. K., Campbell, D. A., Jr, & Stanley, J. C. (2011). Factors associated with and consequences of unplanned post-operative intubation in elderly vascular and general surgery patients. European journal of anaesthesiology, 28(3), 220–224. https://doi.org/10.1097/EJA.0b013e328342659c

[4] Svensson, L. G., Hess, K. R., Coselli, J. S., Safi, H. J., & Crawford, E. S. (1991). A prospective study of respiratory failure after high-risk surgery on the thoracoabdominal aorta. Journal of vascular surgery, 14(3), 271–282.

[5] Arozullah, A. M., Daley, J., Henderson, W. G., & Khuri, S. F. (2000). Multifactorial risk index for predicting postoperative respiratory failure in men after major noncardiac surgery. The National Veterans Administration Surgical Quality Improvement Program. Annals of surgery, 232(2), 242–253. https://doi.org/10.1097/00000658-200008000-00015

disease; history of congestive heart failure; open intrathoracic or upper abdominal surgery; surgical procedure lasting at least 2 hours; and emergency surgery.[6]

Ramachandran and colleagues (2011) analyzed data from 222,094 adult patients who underwent nonemergent, noncardiac surgery in the American College of Surgeons-National Surgical Quality Improvement Program database.[7] Independent predictors of unanticipated early postoperative intubation included current ethanol use, current smoking, dyspnea, chronic obstructive pulmonary disease, diabetes mellitus needing insulin, active heart failure, hypertension requiring medication, abnormal liver function, cancer, prolonged hospitalization, recent weight loss, body mass index less than 18.5 or ≥ 40 kg/m, medium-risk surgery, high-risk surgery, very-high-risk surgery, and sepsis.

Johnson and colleagues (2007) analyzed data from 14 academic and 128 Veterans Affairs Medical Centers from October 2001 through September 2004, and developed a predictive model for PRF using logistic regression.[8] Independent risk factors for PRF included Current Procedural Terminology group, American Society of Anesthesiologists classification, emergency operations, complex operations (work relative value units), preoperative sepsis, and elevated creatinine. Older patients, male patients, smokers, and those with a history of heart failure or COPD were also predisposed. The model's discrimination (c-statistic) was excellent, with no decrement from development (0.856) to validation (0.863) samples.

Burton and colleagues (2018) used data from the Nationwide Inpatient Sample from 2010 to 2014 to identify adult patients who underwent sinus surgery.[9] In this population, the rate of PRF was 3.35% and independent risk factors included pneumonia, bleeding disorder, alcohol dependence, nutritional deficiency, heart failure, paranasal fungal infections, and chronic kidney disease.

4.2 Association with Hospital and Health System Characteristics

Several studies have examined the association between PRF and hospital or health system characteristics. In a multivariable analysis of Nationwide Inpatient Sample (NIS) data from the Healthcare Cost and Utilization Project (HCUP), Rahman and colleagues (2013) found that PRF was less likely in patients admitted to nonteaching hospitals than those admitted to

[6] Canet J, Sabaté S, Mazo V, et al. Development and validation of a score to predict postoperative respiratory failure in a multicentre European cohort: A prospective, observational study. Eur J Anaesthesiol. 2015;32(7):458-470.

[7] Ramachandran, S. K., Nafiu, O. O., Ghaferi, A., Tremper, K. K., Shanks, A., & Kheterpal, S. (2011). Independent predictors and outcomes of unanticipated early postoperative tracheal intubation after nonemergent, noncardiac surgery. *Anesthesiology*, *115*(1), 44–53. https://doi.org/10.1097/ALN.0b013e31821cf6de

[8] Johnson, R. G., Arozullah, A. M., Neumayer, L., Henderson, W. G., Hosokawa, P., & Khuri, S. F. (2007). Multivariable predictors of postoperative respiratory failure after general and vascular surgery: results from the patient safety in surgery study. Journal of the American College of Surgeons, 204(6), 1188–1198. https://doi.org/10.1016/j.jamcollsurg.2007.02.070

[9] Burton BN, Gilani S, Swisher MW, Urman RD, Schmidt UH, Gabriel RA. Factors Predictive of Postoperative Acute Respiratory Failure Following Inpatient Sinus Surgery. Annals of Otology, Rhinology & Laryngology. 2018;127(7):429-438. doi:10.1177/0003489418775129

teaching hospitals (OR 0.89, 95% CI 0.85 to 0.93).[10] The odds of developing PRF increased by 6% for each level increase in hospital size from small to large (OR 1.06, 95% CI 1.03 to 1.09). Using data from 116 VA hospitals and NIS data from 992 community hospitals, Rivard and colleagues (2010) reported lower risk-adjusted rates of PRF in VA hospitals (3.86 per 1,000, 95% CI 2.83 to 4.88) than in the NIS (4.87 per 1,000, 95% CI 3.92 to 5.81).[11]

4.3 Mediating Factors

Several care processes and intermediate factors (or mediators) may contribute to the occurrence of PRF. Some of these factors are within the hospital's/surgeon's control, while others may reflect patient's specific needs, and are therefore not considered as risk factors. These factors include procedure related risk factors such as surgical site, anesthesia type, fluid management, and duration of surgery (which reflects both the complexity of the operation and the skill of the surgical team).

Analyzing data on 50,367 patient admissions for common adult surgical procedures using an anesthesia information system between 2004 and 2009, Blum et al. (2013) identified intraoperative risk factors associated with respiratory failure among patients with similar preoperative risk: ventilator drive pressure (OR=1.17), fraction inspired oxygen (OR=1.02), erythrocyte transfusion (OR=5.36), and crystalloid administration in liters (OR=1.37).[12] The number of different anesthetics administered during the admission was associated with higher risk of ARDS (OR=1.37). Fair evidence also suggests that short-acting neuromuscular blocking agents result in lower rates of residual neuromuscular blockade and may reduce risk for pulmonary complications.[13]

In a multivariable analysis of the National Surgical Quality Improvement Program (NSQIP) database of adult inpatients who underwent neurosurgery under general anesthesia (2005-2010), Shalev and co-authors found that operative time exceeding 3 hours was associated with increased risk of reintubation (OR 2.9; 95%CI 1.8–4.8).[14] In a retrospective time-matched cohort study, Attaallah (2019) found that operative-specific risk factors including ASA status, elective case type, and surgical duration were significantly associated with PRF.[15]

[10] Rahman M, Neal D, Fargen KM, Hoh BL. Establishing standard performance measures for adult brain tumor patients: a Nationwide Inpatient Sample database study. Neuro Oncol. 2013;15(11):1580-1588.
[11] Rivard, P. E., Elixhauser, A., Christiansen, C. L., Shibei Zhao, & Rosen, A. K. (2010). Testing the association between patient safety indicators and hospital structural characteristics in VA and nonfederal hospitals. Medical care research and review : MCRR, 67(3), 321–341.
[12] Blum JM, Stentz MJ, Dechert R, et al. Preoperative and intraoperative predictors of postoperative acute respiratory distress syndrome in a general surgical population. Anesthesiology. 2013;118(1):19-29.
[13] Kor DJ, Lingineni RK, Gajic O, et al. Predicting risk of postoperative lung injury in high-risk surgical patients: a multicenter cohort study. Anesthesiology. 2014;120(5):1168-1181.
[14] Shalev D., Kamel H. (2014). Risk of Reintubation in Neurosurgical Patients. Neurocritical care.
[15] Attaallah, A. F., Vallejo, M. C., Elzamzamy, O. M., Mueller, M. G., & Eller, W. S. (2019). Perioperative risk factors for postoperative respiratory failure. Journal of perioperative practice, 29(3), 49–53. https://doi.org/10.1177/1750458918788978

Lukannek and colleagues (2019) analyzed data from a registry of adult patients undergoing non-cardiac surgery between 2005 and 2017 at two independent healthcare networks.[16] Intraoperative predictors of early postoperative tracheal re-intubation included early post-tracheal intubation desaturation; prolonged duration of surgery; high fraction of inspired oxygen; high vasopressor dose; blood transfusion; the absence of volatile anesthetic use; and the absence of lung-protective ventilation.

4.4 Social Risk Factors

Social factors or social determinants of health (SDOH), have been studied for surgical patients by the American College of Surgeons (ACS) and others. For example, the Strong for Surgery initiative uses checklists to screen patients for risk factors that "can lead to surgical complications, and to provide appropriate interventions to ensure better surgical outcomes."[17] Strong for Surgery targets several topics that have been shown to be associated with surgical outcomes such as nutrition, smoking, and glycemic control, and encourages surgical teams to mitigate associated risks through preoperative interventions. The residual impact of these social factors is captured through measured patient characteristics such as smoking, ASA classification, weight loss, obesity, and laboratory test results such as serum albumin. Some social risk factors, such as social network characteristics, access to transportation, etc., are likely to have effects mediated through hospital choice. For these reasons, there is little conceptual rationale for adjusting for social risk factors in the risk-adjustment model for PRF.

Notwithstanding the above points, we tested the marginal impact of selected social risk factors available in EHR data, after adjusting for patient characteristics. These social risk factors included

5. Methodology

5.1 Data Sources

Lowess (locally weighted scatterplot smoother) plots were created for all the continuous variables to determine the variable type in the analytical models. All risk factors were dichotomous (0/1) except for lab values, which were multi-categorical using Apache categories[18] and then dichotomized for model building purposes, and age, which was created as a piecewise variable. Data sources included:

- ICD-10-CM diagnosis codes for comorbidities present on admission, including acquired immune deficiency syndrome (AIDS), alcohol abuse, deficiency anemia, autoimmune conditions, chronic blood loss anemia, leukemia, lymphoma, metastatic cancer, solid

[16] Lukannek, C., Shaefi, S., Platzbecker, K., Raub, D., Santer, P., Nabel, S., Lecamwasam, H. S., Houle, T. T., & Eikermann, M. (2019). The development and validation of the Score for the Prediction of Postoperative Respiratory Complications (SPORC-2) to predict the requirement for early postoperative tracheal re-intubation: a hospital registry study. Anaesthesia, 74(9), 1165–1174. https://doi.org/10.1111/anae.14742

[17] Strong for Surgery. American College of Surgeons. Available at: www.facs.org/quality-programs/strong-for-surgery. Accessed March 15, 2021

[18] https://www.mdcalc.com/calc/1868/apache-ii-score , accessed September 2023

tumor without metastasis, cerebrovascular disease, coagulopathy, dementia, depression, diabetes (with and without chronic complications, drug abuse, congestive heart failure, hypertension (complicated and uncomplicated), liver disease (mild and moderate to severe), chronic pulmonary disease, neurological disorders, seizures and epilepsy, obesity, paralysis, peripheral vascular disease, psychoses, pulmonary circulation disease, renal failure (moderate and severe), hypothyroidism, other thyroid disorders, peptic ulcer with bleeding, valvular disease, and weight loss;

- Anesthesia, mechanical ventilation, intubation and extubation record for surgery;
- Electronic health record (EHR) lab values for white blood cells (leukocytes), albumin, bilirubin, BUN, creatinine, hematocrit, temperature, heart rate, pH arterial blood gas (ABG), partial pressure of oxygen in the arterial blood (PaO2), and sodium; and
- EHR demographic fields for age, sex, race, ethnicity, and primary payer.

5.1.1 Physiologic Characteristics

Physiologic characteristics included: oxygen, leukocyte count, creatinine, albumin, blood urea nitrogen (BUN), bilirubin, body temperature, heart rate, pH of arterial blood, sodium, and hematocrit. We used a combination of the well-accepted APACHE II[19] and III[20] (severity of disease classification systems) to categorize 11 of these physiologic characteristics into groups without being subject to the influence of data idiosyncrasies. Some physiologic characteristics were coded using APACHE II, including body temperature, heart rate, pH of arterial blood, sodium, and hematocrit. The remaining physiologic characteristics were coded using APACHE III, including oxygen (partial pressure), leukocyte count, creatinine, albumin, blood urea nitrogen (BUN), bilirubin. APACHE II transforms every vital sign into a five-valued score (0, 1, 2, 3, or 4), while APACHE III transforms physiologic characteristics into a multi-valued, weighted score that reflected the degree of distance from 'normal' in either direction (i.e., below normal or above normal). Multiple validations of these groupings have been published.[21] [22]

We binarized each vital sign based on the APACHE II or III coding by combining normal values (collapsed into category zero) with all below normal and/or all above normal values according to the following logic:

- If a below normal value indicated a prognostically negative vital sign (e.g., albumin < 2.5 g/dL), these values were coded as one, and then all other values (i.e., normal or above normal) were coded as zero.

[19] Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical care medicine*, *13*(10), 818-829.
[20] Knaus, W. A., Wagner, D. P., Draper, E. A., & Zimmerman, J. E. (1991). APACHE III Prognostic System: Risk Prediction of Hospital Mortality for Critically III Hospitalized Adults, *Chest*, Volume 100 (6), 1619-1636.
[21] Keegan MT, Gajic O, Afessa B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance. *Chest*. 2012 Oct;142(4):851-858.
[22] Vasilevskis EE, Kuzniewicz MW, Cason BA, Lane RK, Dean ML, Clay T, Rennie DJ, Vittinghoff E, Dudley RA. Mortality probability model III and simplified acute physiology score II: assessing their value in predicting length of stay and comparison to APACHE IV. *Chest*. 2009 Jul;136(1):89-101.

- If an above normal value indicated a prognostically negative vital sign (e.g., creatinine >= 1.5 mg/dL), these values were coded as one, and then all other values (i.e., normal or below normal) were coded as zero.
- For some physiologic characteristics (e.g., serum sodium and pH of arterial blood), both above and below normal values were evaluated as potential features, and then combined into a single category with a value of one (i.e., either above or below normal), if appropriate. Normal values were coded as zero.
- Laboratory data values containing inequality signs were recoded as the numeric value immediately above or below that threshold. For example, a BUN value of ">100" was recoded as 100.1. Values of TNP ("test not performed") and NULL were recoded as normal.

Using temperature[23] as an example, APACHE II assigns values in the following manner:

- <30° = 4
- 30° to <32° = 3
- 32° to <34° = 2
- 34° to <36° = 1
- 36° to < 38.5° = 0
- 38.5° to <39° = 1
- 39° to <41° = 3
- ≥41° = 4

Following the logic described above, we collapsed temperature into three categories (below normal, normal, above normal), and then further collapsed it into two categories after feature selection demonstrated that high temperatures lacked prognostic value, thereby converting temperature to a binary indicator variable:[24]

- Category 0: 36° – 40.9°
- Category 1: < 36° (no values over 40.9 were observed)

Analogously, we categorized each remaining vital sign using the same logic.

Oxygen saturation (PO2):

- Category 1: < 80 mmHg
- Category 0: ≥ 80 mmHg

Leukocyte count:

[23] Reported in centigrade.

[24] We also experimented the idea of expanding the number of categories to fully acknowledge the value range used in APACHE II. However, due to the sample size constraints, this approach can render zero numerator frequency for certain categories. Thus, we used two categories for most features.

- Category 0: < 20,000 µL
- Category 1: ≥ 20,000 µL

Creatinine:

- Category 0: < 1.5 mg/dL
- Category 1: ≥ 1.5 mg/dL

Albumin:

- Category 0: ≥ 2.5 g/dL
- Category 1: < 2.5 g/dL

BUN:

- Category 0: < 14.4 mg/dL
- Category 1: ≥ 14.4 mg/dL

Bilirubin:

- Category 0: < 2.0 mg/dL
- Category 1: ≥ 2.0 mg/dL

Heart rate:

- Category 0: < 110 (bpm)
- Category 1: >= 110 (bpm)

pH of arterial blood:

- Category 0: 7.25 to 7.49
- Category 1 : < 7.25 or > 7.49

Sodium:

- Category 0: 130 to 149 mmol/L
- Category 1: < 130 or > 149 mmol/L

Hematocrit:

- Category 0: ≥ 30%
- Category 1: < 30%

Our decision for more parsimonious grouping of these variables was partly driven by clinical knowledge and partly by the small sample size. As mentioned above, a more granular classification can yield zero numerator prevalence in certain categories.

5.1.2 Patient Comorbidities

We used the AHRQ Elixhauser Comorbidity Software Refined for ICD-10-CM (version 2022.1) to capture patients' comorbidities using the ICD-10-CM diagnoses. The AHRQ Elixhauser software

is coded using publicly available Healthcare Cost and Utilization Project (HCUP) software, annually updated, and extensively validated.[25] The comorbidities tested included:

- AIDS
- Alcohol abuse
- Deficiency anemias
- Arthropathies
- Chronic blood loss anemia
- Leukemia
- Lymphoma
- Metastatic cancer
- Solid tumor without metastasis, in situ
- Solid tumor without metastasis, malignant
- Cerebrovascular disease
- Congestive heart failure
- Coagulopathy
- Dementia
- Depression
- Diabetes with chronic complications
- Diabetes without chronic complications
- Drug abuse
- Hypertension, complicated
- Hypertension, uncomplicated
- Liver disease, mild
- Liver disease, moderate to severe
- Chronic pulmonary disease
- Neurological disorders affecting movement
- Other neurological disorders
- Seizures and epilepsy
- Obesity
- Paralysis

[25] https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp

- Peripheral vascular disease
- Psychoses
- Pulmonary circulation disease
- Renal failure, moderate
- Renal failure, severe
- Hypothyroidism
- Other thyroid disorders
- Peptic ulcer with bleeding
- Valvular disease
- Weight loss

5.2 Model Development

Guided by the conceptual model, we developed the baseline risk adjustment model for HH-PRF using a 2-step sequential process (A) feature Selection followed by (B) risk adjustment (RA) model development as explained below.

1. Created contingency tables (see Table 2) for all the categorical features to identify any that had zero cells for either the positive or negative outcome. These features were not considered for feature selection due to anticipated model convergence problems (i.e., quasi-complete separation) in the RA model. For continuous variables, such as age, we ran locally weighted bivariate regressions (i.e., locally weighted scatterplot smoothing, or LOWESS) to understand the functional form of the relationship. This analysis confirmed that the risk of fall with injury was linearly (piecewise linear) related to age (see **Exhibit 2**).
2. Obtained summary statistics such as event rate by facility, overall event rate, overall event rate based on encounter days, and unadjusted observed event rates by facility.
3. Randomly partitioned the full denominator data into a 80% training set and a 20% hold-out test set. The hold out test set was used to evaluate the generalisability of the features chosen. The feature selection algorithm was applied to the training set with 100-fold cross-validation (CV).

Table 2. Contingency Table of Categorical Features

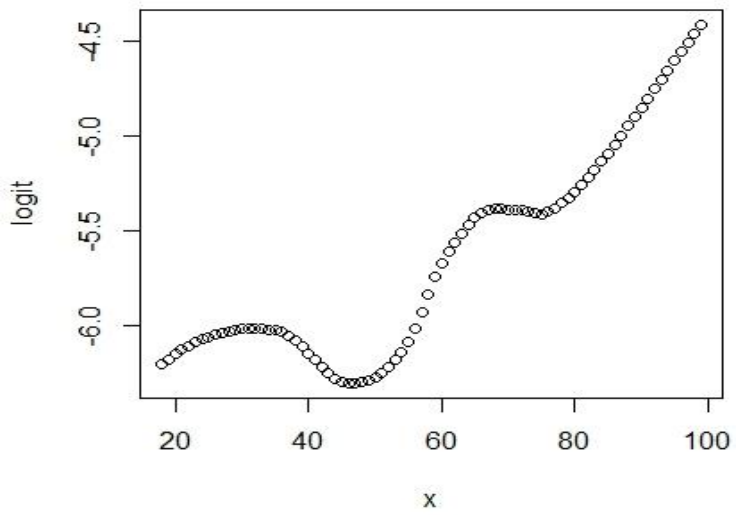
Categorical Feature	# Negative events	# Positive Events	Event Rate (%)
ASA_Category 0 Normal to Mild	13801	10	0.03%
ASA_Category 1 Severe, Not life threatening	15026	53	0.17%
ASA_Category 2 Severe, Life threatening	1532	28	0.09%
Oxygen saturation (PO2): Category 0: ≥ 80 mmHg	30479	81	0.26%
Oxygen saturation (PO2): Category 1: < 80 mmHg	250	14	0.05%
Creatinine Category 0: < 1.5 mg/dL	29353	79	0.26%
Creatinine Category 1: ≥ 1.5 mg/dL	1376	16	0.05%
Albumin Category 0: ≥ 2.5 g/dL	30549	81	0.26%

Categorical Feature	# Negative events	# Positive Events	Event Rate (%)
Albumin Category 1: < 2.5 g/dL	180	14	0.05%
BUN Category 0: < 14.4 mg/dL	21140	31	0.10%
BUN Category 1: ≥ 14.4 mg/dL	9589	64	0.21%
Bilirubin Category 0: < 2.0 mg/dL	30545	82	0.27%
Bilirubin Category 1: ≥ 2.0 mg/dL	184	13	0.04%
Temperature Category 0: 36° – 41°+	30512	92	0.30%
Temperature Category 1: < 36°	217	3	0.01%
Leukocyte count Category 0: < 20,000 µL	30271	85	0.28%
Leukocyte count Category 1: ≥ 20,000 µL	458	10	0.03%
Heart rate Category 0: < 110 (bpm)	30232	92	0.30%
Heart rate Category 1: ≥ 110 (bpm)	497	3	0.01%
pH of arterial blood Category 0: 7.25 to 7.49	30581	80	0.26%
pH of arterial blood Category 1 : < 7.25 or > 7.49	148	15	0.05%
Sodium Category 0: 130 to 149 mmol/L	30359	88	0.29%
Sodium Category 1: < 130 or > 149 mmol/L	370	7	0.02%
Hematocrit Category 0: ≥ 30%	28240	68	0.22%
Hematocrit Category 1: < 30%	2489	27	0.09%
Age Category 0: < 50	8794	16	0.05%
Age Category 1: 50 to 64	8833	25	0.08%
Age Category 2: 65 to 79	10758	41	0.13%
Age Category 3: >= 80	2344	13	0.04%
Gender: Male	13457	57	0.18%
Gender: Female	17272	38	0.12%
Surgical Procedures (no)	26517	69	0.22%
Surgical Procedures (yes)	4212	26	0.08%
AIDS POA (no)	30487	94	0.30%
AIDS POA (yes)	179	1	0.00%
Alcohol abuse POA (no)	30475	92	0.30%
Alcohol abuse POA (yes)	191	3	0.01%
Deficiency anemias POA (no)	28734	71	0.23%
Deficiency anemias POA (yes)	1932	24	0.08%
Arthropathies POA (no)	29743	90	0.29%
Arthropathies POA (yes)	923	5	0.02%
Chronic blood loss anemia POA (no)	30553	93	0.30%
Chronic blood loss anemia POA (yes)	113	2	0.01%
Leukemia POA (no)	30581	94	0.30%
Leukemia POA (yes)	85	1	0.00%
Lymphoma POA (no)	30600	92	0.30%
Lymphoma POA (yes)	66	3	0.01%
Metastatic cancer POA (no)	29873	89	0.29%
Metastatic cancer POA (yes)	793	6	0.02%
Solid tumor without metastasis, in situ POA (no)	30611	95	0.31%
Solid tumor without metastasis, in situ POA (yes)	55	0	0.00%
Solid tumor without metastasis, malignant POA (no)	30065	88	0.29%
Solid tumor without metastasis, malignant POA (yes)	601	7	0.02%
Cerebrovascular disease POA (no)	30250	92	0.30%
Cerebrovascular disease POA (yes)	416	3	0.01%
Congestive heart failure POA (no)	29217	60	0.19%
Congestive heart failure POA (yes)	1449	35	0.11%

Categorical Feature	# Negative events	# Positive Events	Event Rate (%)
Coagulopathy POA (no)	30129	83	0.27%
Coagulopathy POA (yes)	537	12	0.04%
Dementia POA (no)	30641	95	0.31%
Dementia POA (yes)	25	0	0.00%
Depression POA (no)	27929	84	0.27%
Depression POA (yes)	2737	11	0.04%
Diabetes with chronic complications POA (no)	28675	68	0.22%
Diabetes with chronic complications POA (yes)	1991	27	0.09%
Diabetes without chronic complications POA (no)	27065	88	0.29%
Diabetes without chronic complications POA (yes)	3601	7	0.02%
Drug abuse POA (no)	30469	94	0.30%
Drug abuse POA (yes)	197	1	0.00%
Hypertension, complicated POA (no)	27540	51	0.17%
Hypertension, complicated POA (yes)	3126	44	0.14%
Hypertension, uncomplicated POA (no)	18938	64	0.21%
Hypertension, uncomplicated POA (yes)	11728	31	0.10%
Liver disease, mild POA (no)	29184	85	0.28%
Liver disease, mild POA (yes)	1482	10	0.03%
Liver disease, moderate to severe POA (no)	30543	88	0.29%
Liver disease, moderate to severe POA (yes)	123	7	0.02%
Chronic pulmonary disease POA (no)	25894	71	0.23%
Chronic pulmonary disease POA (yes)	4772	24	0.08%
Neurological disorders affecting movement (no)	30423	94	0.30%
Neurological disorders affecting movement (yes)	243	1	0.00%
Other neurological disorders (no)	30097	92	0.30%
Other neurological disorders (yes)	569	3	0.01%
Seizures and epilepsy (no)	30262	92	0.30%
Seizures and epilepsy (yes)	404	3	0.01%
Obesity (no)	23065	61	0.20%
Obesity (yes)	7601	34	0.11%
Paralysis (no)	30387	91	0.30%
Paralysis (yes)	279	4	0.01%
Peripheral vascular disease (no)	29359	73	0.24%
Peripheral vascular disease (yes)	1307	22	0.07%
Psychoses (no)	30286	94	0.30%
Psychoses (yes)	380	1	0.00%
Pulmonary circulation disease (no)	30115	80	0.26%
Pulmonary circulation disease (yes)	551	15	0.05%
Renal failure, moderate (no)	29539	82	0.27%
Renal failure, moderate (yes)	1127	13	0.04%
Renal failure, severe (no)	29913	81	0.26%
Renal failure, severe (yes)	753	14	0.05%
Hypothyroidism (no)	27645	85	0.28%
Hypothyroidism (yes)	3021	10	0.03%
Other thyroid disorders (no)	30030	94	0.30%
Other thyroid disorders (yes)	636	1	0.00%
Peptic ulcer with bleeding (no)	30596	92	0.30%
Peptic ulcer with bleeding (yes)	70	3	0.01%
Valvular disease (no)	29059	67	0.22%

Categorical Feature	# Negative events	# Positive Events	Event Rate (%)
Valvular disease (yes)	1607	28	0.09%
Weight loss (no)	29923	74	0.24%
Weight loss (yes)	743	21	0.07%
Hispanic	4321	10	0.03%
Not Hispanic	22839	71	0.23%
Race: White	16135	48	0.16%
Race: Black	4026	16	0.05%
Race: Other	7679	19	0.06%
Race: Unknown	2889	12	0.04%
Has Medicaid	6698	24	0.08%
No Medicaid	24031	71	0.23%

Exhibit 2: LOWESS Smoothing; Patient Age (x-axis) and PRF



4. We performed feature selection using the least absolute shrinkage and selection operator (LASSO) on the training set using 100-fold cross-validation (CV). We ran LASSO using all the clinically justifiable features on the training set using 100-fold cross-validation (CV) (see **Exhibit 3**). This step helped understand how many features get selected at different values of the regularization parameter (lambda) and to assess model fit on the training set. We extracted the final set of features chosen by the model where the regularization parameter (lambda) was set to lambda1se, i.e., "one-standard-error" (i.e., the largest lambda at which the mean squared error (MSE) is within one standard error of the minimum MSE). This rule is standard practice for improving generalization on hold-out test set (unseen data). [26]

[26] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning.* New York, NY: Springer; 2009.

5. The LASSO model (where lambda is equal to lambda1se) with the selected features was evaluated on the hold-out test set and performance metrics obtained (see **Exhibit 4**).

Exhibit 3: LASSO feature selection (100-fold CV on 80% Training Set)

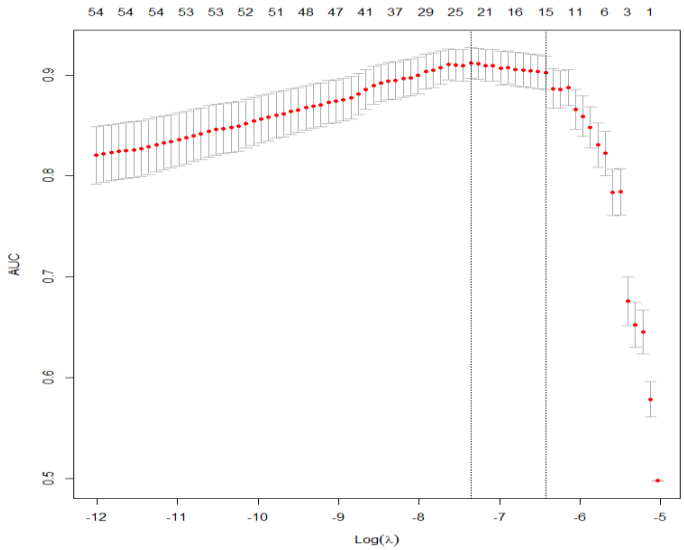
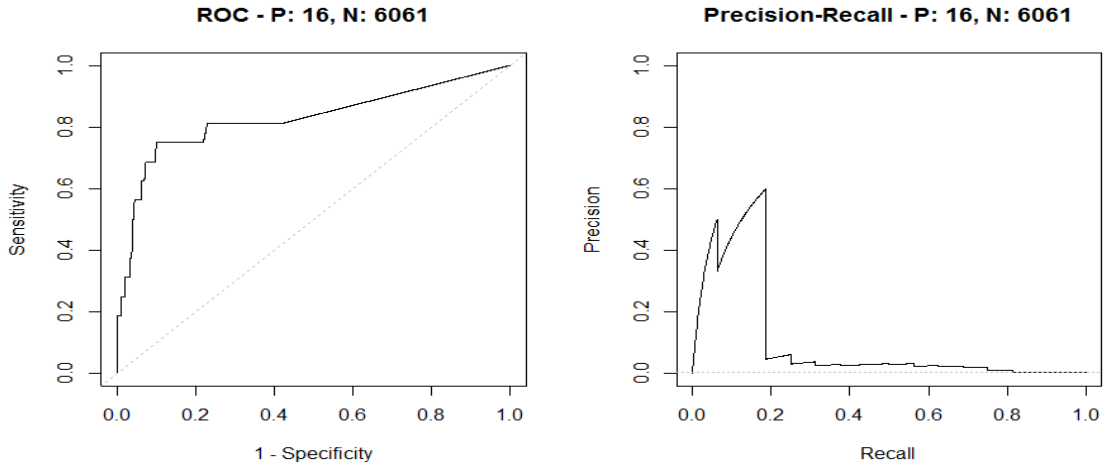


Exhibit 4: Performance of LASSO model with selected features (hold-out Test Data)



6. The estimates from the final risk-adjustment model was generated using a multivariate logit regression model, estimated on the entire dataset using the set of features selected by LASSO through 100-fold cross-validation and tested on the hold-out test set. The predicted values used for calibration plots were generated using a multivariate probit model. Feature selection and RA model performance were evaluated using a variety of metrics such as C-statistics, area under the precision-recall curve and calibration plots.
7. The risk-adjustment (RA) model was also tested with additional social drivers of health variables (Medicaid insurance, Hispanic ethnicity, race), considered individually and collectively. See **Section 7** for results.

8. After feature selection with 100-fold cross-validation and testing on the hold-out test set, the retained risk factors were weight loss POA, deficiency anemias POA, heart failure POA, diabetes with chronic complications POA, moderate to severe liver disease POA, peripheral vascular disease POA, pulmonary circulation disease POA, valvular disease POA, ASA categories (https://www.asahq.org/standards-and-practice-parameters/statement-on-asa-physical-status-classification-system), and lab values for oxygen (partial pressure), leukocytes, albumin, BUN, bilirubin, and pH of arterial blood. We used APACHE II or APACHE III categorizations of these laboratory values, as appropriate, and aggregated categories to achieve the optimal separation of low-risk and high-risk patients. In accord with APACHE categorization methods, missing values were assigned to the "normal" or reference category for each lab test. We tested models by forcing in age and sex (which were not selected by LASSO) but found that it led to no meaningful improvement in any metric of model performance (e.g., AUC, Brier score, AIC/BIC) nor was clinically justifiable.

5.3 Model Performance

After selecting features as described above, we estimated RA models using alternative functional forms to optimize calibration across the risk distribution. The probit specification had the best calibration based on both decile plots and bands, so predicted values for risk adjustment come from the log reg Probit RA model. We show odds ratio estimates from the log reg Logit model because the parameter estimates from a logit model are more interpretable (by exponentiation to yield adjusted odds ratios) than those from a probit model.

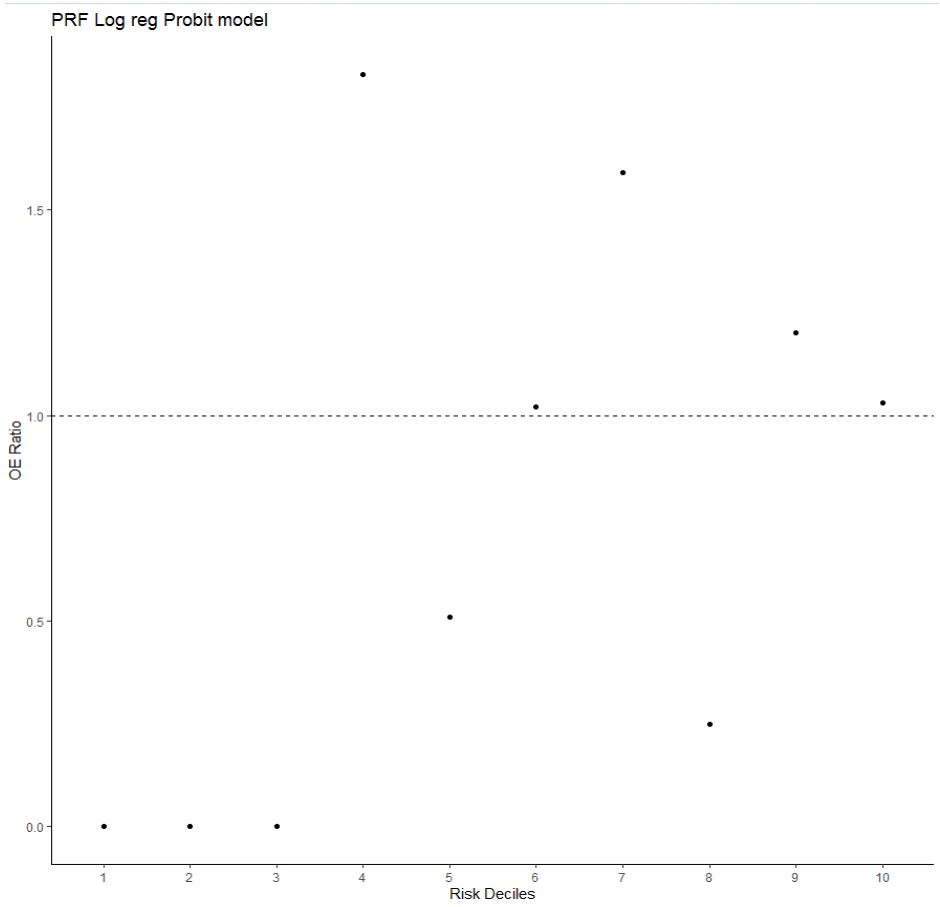
Overall model discrimination was assessed by C-statistic. The C-statistic is the area under the receiver-operator curve (i.e., AUC) that measures the discriminative ability of a regression model across all levels of risk. It also describes the probability that a randomly selected patient who experienced postoperative respiratory failure had a higher expected value than a randomly selected patient who did not experience that event. The AUC was 0.826 in the holdout test set (based on least absolute shrinkage and selection operator or LASSO regression) and 0.912 for the final probit RA model. These values indicate strong discrimination performance, relative to a random classifier with AUC= 0.5.

The precision-recall (PR) curve and the area under the curve (AUPRC) are less sensitive to data imbalance or class imbalance (i.e., very rare events) than the AUC. Given the low overall event rate for this measure, it was advisable to check the values of AUPRC. The AUPRC was 0.098 in the holdout test set (based on LASSO), indicating poor prediction at the individual patient level but good performance relative to a random classifier with AUPRC=0.003.

The RA model calibration was assessed across deciles of patient risk using Hosmer-Lemeshow plots. The deciles of risk are ten mutually exclusive groups containing equal numbers of discharges, ranging from very low-risk patients (according to the model) to high-risk patients. We do not provide Hosmer-Lemeshow test statistics because, given the large sample size of our data, the null hypothesis is almost always rejected. Moreover, the plots provide more detail on model fit than the overall Hosmer-Lemeshow statistic. Because over 78% of events occurred in

the highest-risk decile, and nearly 88% occurred in the highest-risk quintile, the decile analysis is statistically unstable. **Exhibit 5** shows the Hosmer-Lemeshow decile calibration plot from the final risk-adjustment model. The results are unstable due to a small number of observed events (num_obs) and expected events (num_exp) in all of the bottom eight deciles.

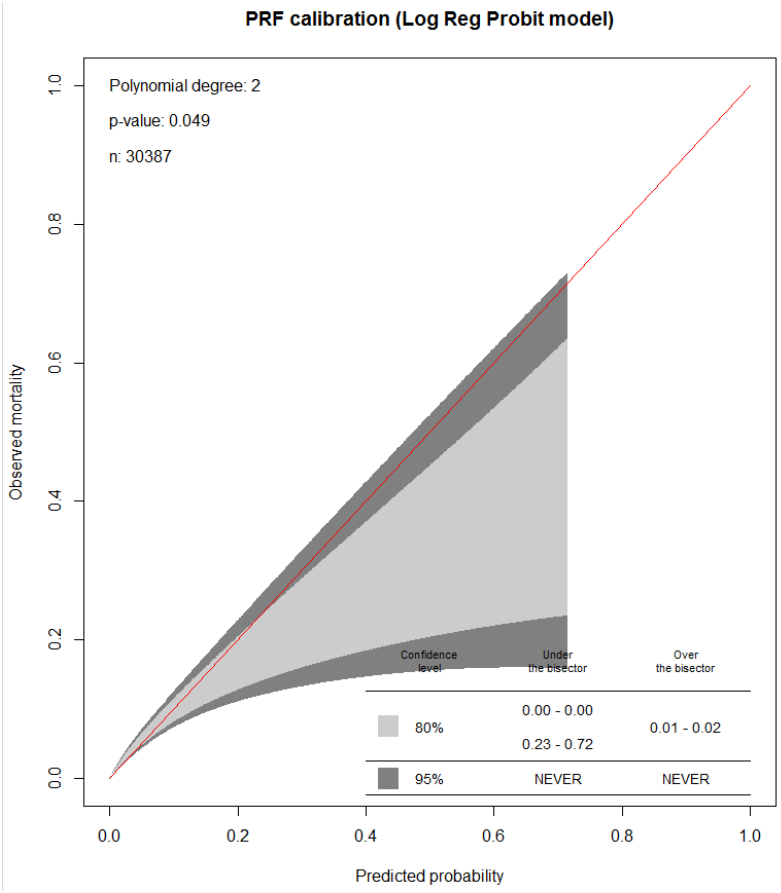
Exhibit 5: Hosmer-Lemeshow Decile Calibration Plot (Final Risk Adjustment Model)



decile	o_e_ratio	num_obs	num_exp	numrecs	Event rate	% of events
1	0	0	0.941	3039	0.0000%	0.00%
2	0	0	0.941	3039	0.0000%	0.00%
3	0	0	0.941	3039	0.0000%	0.00%
4	1.83	3	1.64	3038	0.0987%	3.30%
5	0.51	1	1.95	3039	0.0329%	1.10%
6	1.02	2	1.97	3039	0.0658%	2.20%
7	1.59	4	2.52	3038	0.1317%	4.40%
8	0.25	1	4.05	3039	0.0329%	1.10%
9	1.2	9	7.53	3039	0.2962%	9.89%
10	1.03	71	68.8	3038	2.3371%	78.02%

A preferred approach in this situation is to estimate calibration belts suggested by Nattino et al. (2017).[27] Calibration belts are an advance over the conventional Hosmer-Lemeshow plot, as the latter has the limitation of undue sensitivity to the choice of bins and extreme fluctuations in the observed-to-expected ratios in bins with few harm events. The null hypothesis of perfect calibration is barely rejected at the p<0.05 level (i.e., p=0.049), but the 95% confidence boundaries never cross the bisector. **Exhibit 6** shows the calibration band plot from the final risk-adjustment model.

Exhibit 6: Calibration Band Plot (Final Risk Adjustment Model)



6. Risk Adjustment Model Specification

Table 3 shows the coefficient estimates, standard errors and 95% confidence interval using data points from the full denominator population.

Table 3. Coefficient Estimates, Standard Errors, and Odds Ratios (Full Denominator Population)

	Probit Coef.	Robust S.E Clustered at Hospital	Significance	Odds Ratio (95% CI)	
Deficiency anemias (POA)	0.219	0.108	*	1.840	(1.094, 3.093)

[27] Nattino, G., Lemeshow, S., Phillips, G., Finazzi, S., & Bertolini, G. (2017). Assessing the calibration of dichotomous outcome models with the calibration belt. The Stata Journal, 17(4), 1003-1014

	Probit Coef.	Robust S.E Clustered at Hospital	Significance	Odds Ratio (95% CI)	
Heart failure (POA)	0.309	0.114	**	2.413	(1.376, 4.232)
Diabetes with chronic complications (POA)	0.288	0.105	**	1.781	(1.057, 2.998)
Liver disease, moderate to severe (POA)	0.617	0.227	**	5.065	(1.97, 13.025)
Peripheral vascular disease (POA)	0.243	0.116	*	1.699	(0.967, 2.983)
Pulmonary circulation disease (POA)	0.263	0.142	n.s.	1.771	(0.915, 3.429)
Valvular disease (POA)	0.28	0.111	*	1.927	(1.118, 3.321)
Weight loss (POA)	0.488	0.125	***	3.092	(1.725, 5.545)
ASA Category 3	0.204	0.124	n.s.	2.042	(0.988, 4.223)
ASA Category 4 or 5	0.330	0.160	*	2.594	(1.088, 6.187)
Oxygen (partial pressure <80 mm Hg)	0.639	0.156	***	4.474	(2.293, 8.731)
Leukocyte count (≥20,000/µL)	0.601	0.158	***	4.712	(2.289, 9.698)
Albumin (<2.5 g/dL)	0.661	0.167	***	4.930	(2.453, 9.908)
Blood urea nitrogen (≥14.4 mg/dL)	0.214	0.092	*	1.905	(1.167, 3.109)
Bilirubin (≥2.0 mg/dL)	0.587	0.178	***	3.633	(1.703, 7.752)
pH of arterial blood (<7.25 or >7.49)	1.029	0.167	***	10.305	(5.154, 20.603)

*** p<0.0001; ** p < 0.001; *p < 0.01; n.s. = not significant

Notes: The odds ratio estimates in this table come from logit estimation to improve interpretability. The parameter estimates and their standard errors come from prohibit estimation, which was used to improve calibration and generate unbiased predicted rates. See https://stats.stackexchange.com/questions/137905/interpretation-of-odds-in-probit-regression for explanation of why odds ratio estimates from probit regression are uninterpretable; AIC: 938.48; Number of Fisher Scoring iterations: 10; c-statistic: 0.912; BrierScore 0.0028.

Table 4 shows the denominator count as well as observed and risk-adjusted measure performance rates for every hospital included in the analysis. We calculated risk-adjusted measure rate as:

$$\frac{Observed\ measure\ rate}{Expected\ measure\ rate} \times sample\ average,$$

where the expected measure rate came from the risk-adjustment model and the sample average stands in for the observed measure rate in the reference population.

Table 4. Denominator Count, Observed and Risk-adjusted Measure Rates Per 1,000 Qualified Inpatient Encounters

Hospital	Denominator Count	Observed Measure Rate	Risk-adjusted Measure Rate
1	322	18.63	4.35
2	1,264	3.16	5.45
3	10,909	1.19	2.80
4	4,724	4.23	2.54
5	638	6.27	16.79
6	5,345	6.17	3.10
7	73	0.00	0.00
8	851	1.18	1.79
9	866	1.15	1.84
10	2,643	1.51	2.70
11	995	0.00	0.00
12	1,757	2.85	2.69

Note: Expected measure rate was resulted from the risk-adjustment model and sample average serves as the proxy for the observed measure rate in the reference population.

7. Social Risk Factors

Using data from 12 hospitals we conducted a social disparities analysis and found:

- Hispanic patients have similar risk of PRF (OR=0.96; 95% CI, 0.42-2.20) as non-Hispanic patients, after adjusting for age and other factors in the risk-adjustment model.
- Black patients (OR=1.45; 95% CI, 0.77-2.75) and patients of "other" race (OR=0.92; 95% CI, 0.47-1.78) have similar risk of PRF as White patients, after adjusting for age and other factors in the risk-adjustment model.
- Risk of PRF is unrelated to Medicaid or uninsured status (OR=1.24; 95% CI, 0.72-2.12), or dual eligibility among Medicare beneficiaries, after adjusting for age and other factors in the risk-adjustment model
- Analyses of observed, expected, and risk-adjusted rates in all of the above patient cohorts confirm that the comorbidities and physiologic factors in the risk-adjustment model account for some increased risk of PRF among Black patients (average expected rate 0.330% versus 0.296%), and that any residual bias is not statistically significant.

See **Tables 5 - 8** below for results (individually and collectively).

Table 5: Social Drivers of Health Analysis - Race

	Estimate	Std. Error	z value	Pr(>|z|)
Race: White	REF			
Race: Black	0.42198	0.31784	1.328	0.184286
Race: Other	-0.04282	0.29587	-0.145	0.884933
Race: Unknown	0.41833	0.34320	1.219	0.222881
Deficiency anemias (POA)	0.59141	0.26654	2.219	0.026496 *
Congestive heart failure (POA)	0.85947	0.28690	2.996	0.002738 **
Diabetes with chronic complications (POA)	0.57078	0.26664	2.141	0.032301 *
Liver disease, moderate to severe (POA)	1.68923	0.48339	3.495	0.000475 ***
Peripheral vascular disease (POA)	0.51913	0.28788	1.803	0.071338 .
Pulmonary circulation disease (POA)	0.54845	0.33840	1.621	0.105078
Valvular disease (POA)	0.67373	0.27937	2.412	0.015882 *
Weight loss (POA)	1.15750	0.29797	3.885	0.000102 ***
ASA Category 3	0.70926	0.37089	1.912	0.055835 .
ASA Category 4 or 5	0.94950	0.44356	2.141	0.032303 *
Oxygen (partial pressure <80 mm Hg)	1.51258	0.34270	4.414	1.02e-05 ***
Leukocyte count (≥20,000/µL)	1.54367	0.36959	4.177	2.96e-05 ***
Albumin (<2.5 g/dL)	1.60448	0.35443	4.527	5.99e-06 ***
Blood urea nitrogen (≥14.4 mg/dL)	0.65528	0.25017	2.619	0.008809 **
Bilirubin (≥2.0 mg/dL)	1.30485	0.38801	3.363	0.000771 ***
pH of arterial blood (<7.25 or >7.49)	2.33665	0.35537	6.575	4.86e-11 ***

Notes: *** p<0.0001; ** p < 0.001; *p < 0.01; .p<0.1; Cstat = 0.91234; BrierScore = 0.00285

Table 6: Social Drivers of Health Analysis - Medicaid Insurance

	Estimate	Std. Error	z value	Pr(>|z|)
Medicaid	0.2297	0.2590	0.887	0.375177

	Estimate	Std. Error	z value	Pr(>|z|)
Deficiency anemias (POA)	0.5972	0.2657	2.248	0.024592 *
Congestive heart failure (POA)	0.8894	0.2865	3.105	0.001905 **
Diabetes with chronic complications (POA)	0.5590	0.2665	2.097	0.035957 *
Liver disease, moderate to severe (POA)	1.6151	0.4825	3.347	0.000816 ***
Peripheral vascular disease (POA)	0.5510	0.2880	1.913	0.055687 .
Pulmonary circulation disease (POA)	0.5517	0.3382	1.631	0.102798
Valvular disease (POA)	0.6695	0.2787	2.402	0.016307 *
Weight loss (POA)	1.1156	0.2984	3.739	0.000185 ***
ASA Category 3	0.7142	0.3713	1.923	0.054427 .
ASA Category 4 or 5	0.9543	0.4440	2.149	0.031601 *
Oxygen (partial pressure <80 mm Hg)	1.4911	0.3424	4.355	1.33e-05 ***
Leukocyte count (≥20,000/µL)	1.5463	0.3691	4.189	2.80e-05 ***
Albumin (<2.5 g/dL)	1.6052	0.3563	4.505	6.62e-06 ***
Blood urea nitrogen (≥14.4 mg/dL)	0.6594	0.2505	2.632	0.008487 **
Bilirubin (≥2.0 mg/dL)	1.2983	0.3872	3.353	0.000798 ***
pH of arterial blood (<7.25 or >7.49)	2.3304	0.3536	6.590	4.41e-11 ***

Notes: *** p<0.0001; ** p < 0.001; *p < 0.01; .p<0.1; Cstat = 0.9100; BrierScore = 0.0028

Table 7: Social Drivers of Health Analysis – Hispanic Ethnicity

	Estimate	Std. Error	z value	Pr(>|z|)
Hispanic	-0.07828	0.36506	-0.214	0.830220
Deficiency anemias (POA)	0.62112	0.26556	2.339	0.019341 *
Congestive heart failure (POA)	0.87295	0.28654	3.047	0.002315 **
Diabetes with chronic complications (POA)	0.58614	0.26612	2.202	0.027631 *
Liver disease, moderate to severe (POA)	1.63931	0.48228	3.399	0.000676 ***
Peripheral vascular disease (POA)	0.52783	0.28760	1.835	0.066464 .
Pulmonary circulation disease (POA)	0.56649	0.33692	1.681	0.092693 .
Valvular disease (POA)	0.65168	0.27746	2.349	0.018839 *
Weight loss (POA)	1.13356	0.29794	3.805	0.000142 ***
ASA Category 3	0.71376	0.37067	1.926	0.054155 .
ASA Category 4 or 5	0.94304	0.44363	2.126	0.033527 *
Oxygen (partial pressure <80 mm Hg)	1.50042	0.34061	4.405	1.06e-05 ***
Leukocyte count (≥20,000/µL)	1.54471	0.36799	4.198	2.70e-05 ***
Albumin (<2.5 g/dL)	1.58269	0.35589	4.447	8.70e-06 ***
Blood urea nitrogen (≥14.4 mg/dL)	0.63690	0.25010	2.547	0.010879 *
Bilirubin (≥2.0 mg/dL)	1.30119	0.38671	3.365	0.000766 ***
pH of arterial blood (<7.25 or >7.49)	2.35261	0.35419	6.642	3.09e-11 ***

Notes: *** p<0.0001; ** p < 0.001; *p < 0.01; Cstat = 0.9081; BrierScore = 0.0029

Table 8: Social Drivers of Health Analysis – Combined (Race, Medicaid Insurance, Hispanic Ethnicity)

	Estimate	Std. Error	z value	Pr(>|z|)
Race: White	REF			
Race: Black	0.37249	0.32522	1.145	0.252069
Race: Other	-0.08811	0.33891	-0.260	0.794881
Race: Unknown	0.39299	0.46227	0.850	0.395260
Medicaid	0.21441	0.27367	0.783	0.433354

	Estimate	Std. Error	z value	Pr(>|z|)
Hispanic	-0.03595	0.42000	-0.086	0.931783
Deficiency anemias (POA)	0.58471	0.26741	2.187	0.028774 *
Congestive heart failure (POA)	0.86727	0.28703	3.021	0.002516 **
Diabetes with chronic complications (POA)	0.55863	0.26721	2.091	0.036560 *
Liver disease, moderate to severe (POA)	1.68067	0.48461	3.468	0.000524 ***
Peripheral vascular disease (POA)	0.53547	0.28825	1.858	0.063222 .
Pulmonary circulation disease (POA)	0.52840	0.34007	1.554	0.120232
Valvular disease (POA)	0.68125	0.28006	2.432	0.014995 *
Weight loss (POA)	1.14332	0.29875	3.827	0.000130 ***
ASA Category 3	0.70729	0.37127	1.905	0.056775 .
ASA Category 4 or 5	0.94551	0.44428	2.128	0.033322 *
Oxygen (partial pressure <80 mm Hg)	1.50249	0.34421	4.365	1.27e-05 ***
Leukocyte count (≥20,000/µL)	1.53503	0.37111	4.136	3.53e-05 ***
Albumin (<2.5 g/dL)	1.61332	0.35524	4.541	5.59e-06 ***
Blood urea nitrogen (≥14.4 mg/dL)	0.66539	0.25078	2.653	0.007970 **
Bilirubin (≥2.0 mg/dL)	1.31299	0.38914	3.374	0.000741 ***
pH of arterial blood (<7.25 or >7.49)	2.33601	0.35640	6.554	5.59e-11 ***

Notes: *** p<0.0001; ** p < 0.001; *p < 0.01; Cstat = 0. 9087; BrierScore = 0.0028

8. Conclusion

Using EHR data from 12 hospitals with varying bed size, geographic location, and EHR system, we developed a baseline risk adjustment model for HH-PRF. Importantly, the risk model developed is still in its preliminary stage due to the small sample of hospitals. Risk-adjusted measure rates move closer to a state where comparison of hospital performance is affected as little as possible by factors other than the quality of care.

Acknowledging these limitations, we consider this exercise an important innovation in hospital outcome measures using EHR data on two fronts:

1. Developing a risk adjustment methodology for eCQMs responds to the preference of care providers and stakeholders that physiological data captured at the start of encounter can be valuable for adjusting patient-level risk factors in hospital outcome measures. In this sense, we took a step toward developing a risk-adjusted eCQM that takes full advantage of the rich physiological information existent in the medical record and recorded at the beginning of the episode of care. These data are used by clinicians to evaluate how sick patients are and to guide their treatment plans in real time. The face validity of these data and their use for risk adjustment are well-justified.
2. Use of EHR data in risk adjustment provides new efficiencies in future eCQM development and implementation, in that EHR data are already documented during the process of care and hence, data collection incurs minimal burden on providers. Maximizing the utility of EHR data elements for risk adjustment improves feasibility and data element reliability, and potentially improves harmonization across measures.