Risk-Adjusting Hospital Inpatient Mortality Using Automated Inpatient, Outpatient, and Laboratory Databases

Division of Research
Gabriel J. Escobar, MD
Marla N. Gardner
John D. Greene
Arona Ragins
Peter Scheirer

MIA
Patricia Kipnis, PhD
Peter Scheirer
Jay Soule
David Schlessinger
Hillary Street

UC Santa Cruz
David Draper, PhD
Juan Carlos LaGuardia
Milovan Krnjajic, PhD

Stanford
Nick Bambos, PhD
Dimitris Tsamis
Lawrence Chow
Carri Chan

Management Information & Analysis
Innovation Acceleration Insight

STANFORD ENGINEERING
SYSTEMS RESEARCH INITIATIVE
DISCLOSURES

• All of the work described here has been funded by The Permanente Medical Group, Inc., Kaiser Foundation Health Plan, Inc., and Kaiser Foundation Hospitals, Inc.

• None of the individuals who have contributed to this work have any conflicts of interest to disclose
PRESENTATION OBJECTIVES

• Consider the value of moving beyond “the usual suspects” (data elements found in claims data)

• Describe the Northern California Kaiser Permanente Medical Care Program’s risk adjustment methodology

• Describe ongoing work in Kaiser Permanente that highlights how the distinction between risk adjustment models and clinical or operational predictive models will become blurred in the future
MOVING BEYOND CLAIMS DATA

• Risk-adjustment for hospital mortality has been driven by the high costs of acquiring detailed patient clinical data
• Consequently, efforts have been focused on the best use of claims data (the usual suspects: age + sex + ICD codes + various disposition codes)
• One exception to this has been in the ICU, where highly detailed data have been employed, and a variety of scores now exist (APACHE IV, SAPS III, VA automated system) – however, ICU admissions account for only 10-20% of all hospital admissions
• Recent published work highlights the value of incorporation of non-traditional data elements
WHY SHOULD WE SETTLE FOR THE LOWEST COMMON DENOMINATOR?

• Billing data are driven by need for reimbursement

• High propensity to being “gamed”

• “Coding creep” occurs

• Many diagnoses strongly associated with outcome can be present on admission, e.g., stroke, DVT, pressure ulcers
POA coding modifier showed that diagnoses associated with complications were present on admission only 10 – 22% of the time.

Absence of POA coding led to 33 – 40% of low performing hospitals not being detected for these conditions:

- CABG
- Coronary angioplasty
- Hip replacement
- AMI

“Treating complications as pre-existing conditions gives poor-performing hospitals ‘credit’ for their complications and may cause some hospitals that are delivering low-quality care to be misclassified as average- or high-performing hospitals.”
VALUE OF INCORPORATING PHYSIOLOGIC DATA

- Increasingly available, particularly in hospital chains
- Much less expensive than manual chart abstraction
- Have tremendous face validity with clinicians
- Relatively easy to combine with other data
- Can be used either for disease-specific models (e.g., Fine PSI for community acquired pneumonia) or for global risk adjustment (e.g., VA or Kaiser Permanente risk adjustment methodologies)
RELATIVE CONTRIBUTION OF PHYSIOLOGIC DATA TO OVERALL MODEL PERFORMANCE

- Operational use of laboratory data first occurred in the ICU
- Render et al. – 29,377 consecutive first ICU admissions in 17 VA hospitals
  
  Laboratory data accounted for 74% of model predictive ability
  Diagnosis accounted for 13%

- Zimmerman et al. – 110,558 ICU admissions in 45 U.S. hospitals
  
  Laboratory data accounted for 65% of model predictive ability
  Diagnosis accounted for 16%
Quantified effect of adding POA coding, laboratory data, and vital signs data for 5 conditions and 3 procedures

Not restricted to ICU; N ranged from 5309 for AAA to 200,506 for CHF

Average effect of adding predictors, as evidenced by change in c statistic:

- No risk adjustment: 0.50
- Administrative model: 0.79
- POA model: 0.84
- POA + labs: 0.86
- POA + labs + VSS: 0.88
TABAK ET AL.: DEFINITIVE QUANTIFICATION OF VALUE OF LABORATORY DATA

• Evaluated 6 disease-specific mortality predictive models for pay-for-performance (ischemic & hemorrhagic stroke, pneumonia, CHF, and sepsis)

• 194,903 admissions in 2000-2003 across 71 hospitals that imported laboratory data

• Quantified relative contribution with omega statistic

• Laboratory data were between 2 and 67 times more important in predicting mortality than ICD-9 variables

• Only models where laboratory data were less important were those for stroke, where altered mental status recordings were more important
KAISER PERMANENTE MEDICAL CARE PROGRAM, NORTHERN CALIFORNIA REGION

- Integrated health care delivery system

- Information systems based on common medical record number across care continuum

- 3.2 million members

- Approximately 6,000 physicians care for patients at 20 hospitals

- All 56 clinics now using Epic outpatient EMR; all hospitals will be online for inpatient EMR by March 2010
Overview of KPMCP methodology

- 44 patient-centric models generate predicted mortality risk
- Employs only data preceding hospitalization (to avoid the “forgiveness” problem)
  - Incorporates physiologic data
  - Incorporates outpatient and inpatient diagnostic data from the year preceding hospitalization
- Based on linked hospitalization, not isolated stay
- Published in *Medical Care* in March 2008
- Externally validated in Ottawa, Canada
Population

- All hospitalizations in Northern California KPMCP hospitals
- Hospitalizations for delivery *not* included (but post-delivery complications *are* included)
  - Patients < 15 years of age *not* included
  - Patients initially admitted to non-KP hospitals *not* included (other transports-in *are* included)
- Outcomes are ascribed to *first admitting* KP hospital
## Effect of Inter-Hospital Transport

<table>
<thead>
<tr>
<th>Hospital A</th>
<th>Hospital B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 Hospitalizations</td>
<td>1000 Hospitalizations</td>
</tr>
<tr>
<td>50 Deaths</td>
<td>50 Deaths</td>
</tr>
<tr>
<td>Mortality Rate: 5%</td>
<td>Mortality Rate: 5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hospital A</th>
<th>Hospital B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 Hospitalizations</td>
<td>1100 Hospitalizations</td>
</tr>
<tr>
<td>25 Deaths</td>
<td>75 Deaths</td>
</tr>
<tr>
<td>Mortality Rate: 2.5%</td>
<td>Mortality Rate: 6.8%</td>
</tr>
</tbody>
</table>

100 patients 50 deaths
Linked hospitalization examples

- Vallejo – San Francisco – Oakland
  
  included; if patient dies, death is ascribed to Vallejo

- Sutter* – Vallejo – San Francisco
  
  excluded

- Sacramento – Sutter* – Sacramento – Vallejo
  
  included; if patient dies, death is ascribed to Sacramento;
  LOS @ Sutter (non KPMCP hospital) “bridged” using data from Authorization for Outside Medical Services database
Comorbidity Point Score (COPS): measures the chronic illness burden

- Point score based on inpatient and outpatient utilization during the year preceding hospitalization and “bucketed” into hierarchical condition categories (HCCs) by DxCG software
- 16,090 possible ICD codes are grouped into 184 HCCs, which we collapsed into 41 comorbidity groups
- 41 comorbidity groups put in preliminary regression model
- Regression coefficients transformed into points
- Final COPS can range from 0 to theoretical maximum of 701
COPS – sample point values

AIDS 18 points
CHF 32 points
Metastatic cancer 57 points
Uncomplicated diabetes 8 points
Complicated diabetes 10 points
Head injury 11 points
• COPS is a measure of comorbidity.
• Increases in comorbidity increases LOS for a given age. However, LOS still does not always increase with increases in age.

Restrictions: non-death hospitalizations age≥15
Laboratory Acute Physiology Score (LAPS): measures physiologic derangement at admission

- Point score based on 14 laboratory tests obtained in the 24 hours preceding hospitalization (production model uses 72 hours; analyses show no difference between 24 and 72 hour time frame)
- Points are assigned using two regression models
  - Step 1 – assigns preliminary mortality risk (< 6%, ≥ 6%) based on limited patient data; 83% of population is in the low risk group
  - Step 2 – second model includes all available patient data
- Final score is a point score that can range from 0 to a theoretical maximum of 256
- Missing data for 3 laboratory tests – arterial pH, troponin I, and white blood cell count – are handled differently for the 17% of the patients in the high risk group. Among these patients, *extra points are given for missing data for these three tests*
LAPS – sample point values – usual tests

Creatinine

- < 1.0: 0 points
- 1.0 – 1.9: 1 point
- 2.0 – 3.9: 7 points
- ≥ 4.0: 5 points

BUN / Cr

- < 25: 0 points
- ≥ 25: 6 points
LAPS – sample point values – selected tests

White blood cell count (1000s / cu mm)

- < 2.0: 29 points
- 2.0 – 4.9: 6 points
- 5.0 – 12.9: 0 points
- ≥ 13.0: 15 points
- Missing & high risk: 23 points

Other tests handled this way: arterial pH, troponin I,
LOS Distribution for PNEUM According to laps

Probability of Release on Day s

s (days)
Performance metrics for DOR-MIA model

- **c statistic:** 0.88 (validation dataset)
  - restrict to one hospitalization per patient → c = 0.90
  - use final diagnosis instead of admit diagnosis → c = 0.89
  - use 30-day instead of inpatient mortality → c = 0.86
- Introduce random variation in admit diagnosis (e.g., replace “chest pain” with “sepsis”)
  - 5% of the records → c = 0.87
  - 25% of the records → c = 0.86
- Introduce random variation in LAPS or COPS (5 to 25% of the records)
- No change in c statistic
Calibration: a good c statistic is not enough

- When risk-adjusting, having a high c statistic is not the only requirement
- It is also important to show that the approach predicts well at different levels of risk
  - it *does* make a difference if the predicted risk is 12% as opposed to 5%
- Calibration is assessed using a *calibration curve* and a *Hosmer-Lemeshow* plot
What does the explaining?

<table>
<thead>
<tr>
<th>Feature</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission category</td>
<td>7%</td>
</tr>
<tr>
<td>Age</td>
<td>17%</td>
</tr>
<tr>
<td>COPS</td>
<td>9%</td>
</tr>
<tr>
<td>LAPS</td>
<td>54%</td>
</tr>
<tr>
<td>Sex</td>
<td>1%</td>
</tr>
<tr>
<td>Dx subset (some models only)</td>
<td>6%</td>
</tr>
</tbody>
</table>

Numbers do not add up to 100% because these are averages across 44 models.
Use of KP risk adjustment methodology

- The **Observed to Expected Ratios** or **Observed minus Expected differences** are used to compare mortality and LOS across populations of patients with different risk profiles.

- The **observed mortality** is the number of deaths divided by the number of admissions at the originating facility.

- The **observed LOS** is the average length of stay for an episode at the originating facility.

- The **expected mortality** is the mean mortality risk expressed as an a priori risk of death for the particular hospitalization.

- The **expected LOS** is the likely LOS for the particular hospitalization.
Operational use of system

• Risk adjusted mortality and LOS comparisons now updated quarterly

• Key clinical leaders can now access these comparisons via secure web site

• Secure web site also permits trending
Observed to Expected Mortality Ratios (w/ 95% CIs)
Observed to Expected LOS Ratios (w/ 95% CIs)
LOS Ratios vs Mortality Ratios: All

LOS Ratios

- 0.5
- 0.6
- 0.7
- 0.8
- 0.9
- 1.0
- 1.1
- 1.2
- 1.3
- 1.4
- 1.5

Mortality Ratios

- 0.5
- 0.6
- 0.7
- 0.8
- 0.9
- 1.0
- 1.1
- 1.2
- 1.3
- 1.4
- 1.5

Significant = Statistically Different than 1
Admission or Principal Dx: PNEUM- 2006

Probability of Inpatient Mortality vs. LOS Ratio

- LOS Ratio
- Probability of Inpatient Mortality

The chart shows a scatter plot with points labeled N, C, M, B, L, J, M, Q, K, O, F, G, and I.
PREDICTING OUTCOMES AMONG WARD PATIENTS

• While our risk adjustment model predicts death well, it does not predict death within a narrow time frame very well

• Systems such as APACHE or SAPS, which aim to predict and risk adjust intra-ICU outcome, not that helpful at predicting which ward patient will need to go to the ICU

• Manually assigned scores (e.g., Modified Early Warning Score) do not have good performance characteristics (c statistics are in 0.70 range)
SHORT TERM PREDICTION MODELS

- We have identified one patient subset with extremely high mortality rate: patients not initially admitted to the ICU who experience an unplanned transfer
  Ward → ICU, Ward → TCU, TCU → ICU
- These patients have extremely elevated mortality (on the average, severity-adjusted observed to expected mortality ratios of 2.5 to 4.0), with wide variation across facilities
- These patients constitute 3-4% of admissions but account for:
  24% of all ICU admissions
  22% of all hospital deaths
  13% of all hospital days
FUTURE DIRECTIONS

- Incorporation of additional physiologic data
  - Chief complaint (e.g., “confusion”)
  - Usual vital signs (e.g., T, P, R, oximetry)
  - Lactate
- Incorporation of trends (Δ heart rate rather than “worst” heart rate”)
- Short-term prediction time frame (i.e., death or deterioration within 2 to 12 hours of some some T₀)
- Incorporate process-outcome linkage into reporting structure (e.g., outcomes among patients with sepsis WITH and WITHOUT lactate marker, or lactate within some time interval from blood culture)
Can providers quantify risk?

• Data suggest that our clinicians (and our models) can tell the difference between high risk (e.g., risk of death > 10%) and low risk (risk of death < 1%)

• Problem is in that middle range: patients admitted to ward or TCU are in the 2 – 8% risk range, and current tools (including clinical judgment) cannot discriminate well in this range, particularly within a narrow time frame
Code Blue calls, RRT calls, and transfers to ICU at Facility X (11.1.06 – 1.31.08)

Hospitalizations at Facility X
n = 12,002

<table>
<thead>
<tr>
<th>Event Description</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code Blue Team called, no ICU transfer</td>
<td>65</td>
</tr>
<tr>
<td>RRT Team called, no ICU transfer</td>
<td>126</td>
</tr>
<tr>
<td>OR/PAR → ICU</td>
<td>487</td>
</tr>
<tr>
<td>No CB or RRT call</td>
<td>478</td>
</tr>
<tr>
<td>Ward → ICU</td>
<td>456</td>
</tr>
</tbody>
</table>

Legend
- Code Blue Team called
- Rapid Response Team Called
- Both (Code Blue and Rapid Response) Teams Called

* 26 unlinkable records
Early detection of impending deterioration

• We have determined that “one size fits all” predictive models are not going to work

• Different diagnoses have different vital signs characteristics; signals can cancel each other out

• This may be one reason why RRTs, some of which have employed such scores, have not been effective (current detection systems and early warning scores are not very accurate)
Pneumonia patients: Mean systolic blood pressure in 13 patients who experienced an unplanned transfer to the ICU (---) between 8 and 72 hours in the hospital and 162 who did not experience any unplanned transfer (—)
GI bleeding patients: Mean systolic blood pressure in 8 patients who experienced an unplanned transfer to the ICU (---) between 8 and 72 hours in the hospital and 518 who did not experience any unplanned transfer (—)
Signal loss occurs when the two populations are combined.
WHAT COULD BE DONE WITH KP HEALTHCONNECT?

Admit Decision

Admit diagnosis

KP Health Connect ADT

MD Problem list

Daily algorithm run as background process

VSS Lab

Daily probability display
Doe, John X
MRN 987654321

**WARNING:** High risk patient with no WBC or pulse oximeter data in past 24 hours

Age: 68
Diagnosis: Community Acquired Pneumonia

**Probability of physiologic deterioration**

- Within 24 hours: 16%
- Within 48 hours: 21%

**Likely length of stay (diagnosis only estimate):** 4.2 + 1.1 days

**Likely length of stay (severity + diagnosis):** 8.7 + 2.2 days

**Probability of length of stay exceeding 7 days:** 27%
All Hospitalizations

Deaths

Remove "Comfort Care" patients

Preliminary eligible cohort

Deaths
Deaths

Preliminary eligible cohort

Pneumonia patients

Measure process markers among survivors

Measure process markers among decedents
CONCLUSIONS

• Future risk adjustment needs to move past the current focus on data scarcity

• In hospital systems with EMRs, the same applications that generate risk adjustment models could generate actual predicted probabilities for discrete patient outcomes; these predicted probabilities can be linked tightly to desired clinician actions, in real time

• Greater emphasis needs to be placed on developing methods suitable for comparing hospitals with EMRs rather than expending yet more energy on new ways to recombine ICD codes and disposition codes
Articles Cited:


Holloway RG, Quill TE. Mortality as a measure of quality. Implications for palliative and end-of-life care. JAMA 2007; 298:802-804.


van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. Accepted for publication, Journal of Clinical Epidemiology.

Articles Cited (continued):

van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. Accepted for publication, Journal of Clinical Epidemiology.


Holloway RG, Quill TE. Mortality as a measure of quality. Implications for palliative and end-of-life care. JAMA 2007; 298:802-804.

