

## ORIGINAL RESEARCH

# Early Detection of Impending Physiologic Deterioration Among Patients Who Are Not in Intensive Care: Development of Predictive Models Using Data From an Automated Electronic Medical Record

Gabriel J. Escobar, MD<sup>1,2\*</sup>, Juan Carlos LaGuardia, MS<sup>2</sup>, Benjamin J. Turk, BA<sup>2</sup>, Arona Ragins, MA<sup>2</sup>,  
Patricia Kipnis, PhD<sup>2,3</sup>, David Draper, PhD<sup>4</sup>

<sup>1</sup>Hospital Operations Research, Division of Research, Kaiser Permanente Medical Care Program, Oakland, California; <sup>2</sup>Division of Research, Kaiser Permanente Medical Care Program, Oakland, California; <sup>3</sup>Management, Information and Analysis, Kaiser Foundation Health Plan, Oakland, California; <sup>4</sup>Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, Santa Cruz, California.

**BACKGROUND:** Ward patients who experience unplanned transfer to intensive care units have excess morbidity and mortality.

**OBJECTIVE:** To develop a predictive model for prediction of unplanned transfer from the medical–surgical ward to intensive care (or death on the ward in a patient who was “full code”) using data from a comprehensive inpatient electronic medical record (EMR).

**DESIGN:** Retrospective case-control study; unit of analysis was a 12-hour patient shift. Shifts where a patient experienced an unplanned transfer were event shifts; shifts without a transfer were comparison shifts. Hospitalization records were transformed into 12-hour shift records, with 10 randomly selected comparison shifts identified for each event shift. Analysis employed logistic regression and split validation.

**SETTING:** Integrated healthcare delivery system in Northern California.

**PATIENTS:** Hospitalized adults at 14 hospitals with comprehensive inpatient EMRs.

**MEASUREMENTS:** Predictors included vital signs, laboratory test results, severity of illness scores, longitudinal chronic illness burden scores, transpired hospital length of stay, and care directives. Patients were also given a retrospective, electronically (not manually assigned) Modified Early Warning Score, or MEWS(re). Outcomes were transfer to the intensive care unit (ICU) from the ward or transitional care unit, or death outside the ICU among patients who were “full code”.

**RESULTS:** We identified 4,036 events and 39,782 comparison shifts from a cohort of 102,422 patients’ hospitalizations. The MEWS(re) had a c-statistic of 0.709 in the derivation and 0.698 in the validation dataset; corresponding values for the EMR-based model were 0.845 and 0.775.

**LIMITATIONS:** Using these algorithms requires hospitals with comprehensive inpatient EMRs and longitudinal data.

**CONCLUSIONS:** EMR-based detection of impending deterioration outside the ICU is feasible in integrated healthcare delivery systems. *Journal of Hospital Medicine* 2012;7:388–395. © 2012 Society of Hospital Medicine

Patients in general medical–surgical wards who experience unplanned transfer to the intensive care unit (ICU) have increased mortality and morbidity.<sup>1–3</sup> Using an externally validated methodology permitting assessment of illness severity and mortality risk among all hospitalized patients,<sup>4,5</sup> we recently documented observed-to-expected mortality ratios >3.0 and excess length of stay of 10 days among patients who experienced such transfers.<sup>6</sup>

It is possible to predict adverse outcomes among monitored patients (eg, patients in the ICU or undergoing continuous electronic monitoring).<sup>7,8</sup> However, prediction of unplanned transfers among medical–

surgical ward patients presents challenges. Data collection (vital signs and laboratory tests) is relatively infrequent. The event rate (~3% of hospital admissions) is low, and the rate in narrow time periods (eg, 12 hours) is extremely low: a hospital with 4000 admissions per year might experience 1 unplanned transfer to the ICU every 3 days. Not surprisingly, performance of models suitable for predicting ward patients’ need for intensive care within narrow time frames have been disappointing.<sup>9</sup> The Modified Early Warning Score (MEWS), has a c-statistic, or area under the receiver operator characteristic of 0.67,<sup>10–12</sup> and our own model incorporating 14 laboratory tests, but no vital signs, has excellent performance with respect to predicting inpatient mortality, but poor performance with respect to unplanned transfer.<sup>6</sup>

In this report, we describe the development and validation of a complex predictive model suitable for use with ward patients. Our objective for this work was to develop a predictive model based on clinical and physiologic data available in real time from a comprehensive electronic medical record (EMR), not a

\*Address for correspondence and reprint requests: Gabriel J. Escobar, MD, Division of Research, Kaiser Permanente Medical Care Program, 2000 Broadway Ave, Oakland, CA 94612; Telephone:510-891-3502; Fax: 510-891-3508; E-mail: gabriel.escobar@kp.org

Additional Supporting Information may be found in the online version of this article.

Received: August 25, 2011; Revised: January 23, 2012; Accepted: January 30, 2012

2012 Society of Hospital Medicine DOI 10.1002/jhm.1929

Published online in Wiley Online Library (Wileyonlinelibrary.com).

clinically intuitive, manually assigned tool. The outcome of interest was unplanned transfer from the ward to the ICU, or death on the ward in a patient who was “full code.” This model has been developed as part of a regional effort to decrease preventable mortality in the Northern California Kaiser Permanente Medical Care Program (KPMCP), an integrated healthcare delivery system with 22 hospitals.

## MATERIALS AND METHODS

For additional details, see the Supporting Information, Appendices 1–12, in the online version of this article.

This project was approved by the KPMCP Institutional Board for the Protection of Human Subjects.

The Northern California KPMCP serves a total population of approximately 3.3 million members. All Northern California KPMCP hospitals and clinics employ the same information systems with a common medical record number and can track care covered by the plan but delivered elsewhere. Databases maintained by the KPMCP capture admission and discharge times, admission and discharge diagnoses and procedures (assigned by professional coders), bed histories permitting quantification of intra-hospital transfers, inter-hospital transfers, as well as the results of all inpatient and outpatient laboratory tests. In July 2006, the KPMCP began deployment of the EMR developed by Epic Systems Corporation ([www.epic.com](http://www.epic.com)), which has been adapted for the KPMCP and is known as KP HealthConnect (KPHC) in its hospitals. The last of these 22 hospitals went online in March 2010.

Our setting consisted of 14 hospitals in which the KPHC inpatient EMR had been running for at least 3 months (the KPMCP Antioch, Fremont, Hayward, Manteca, Modesto, Roseville, Sacramento, Santa Clara, San Francisco, Santa Rosa, South Sacramento, South San Francisco, Santa Teresa, and Walnut Creek hospitals). We have described the general characteristics of KPMCP hospitals elsewhere.<sup>4,6</sup> Our initial study population consisted of all patients admitted to these hospitals who met the following criteria: hospitalization began from November 1, 2006 through December 31, 2009; initial hospitalization occurred at a Northern California KPMCP hospital (ie, for inter-hospital transfers, the first hospital stay occurred within the KPMCP); age  $\geq 18$  years; hospitalization was not for childbirth; and KPHC had been operational at the hospital for at least 3 months.

### Analytic Approach

The primary outcome for this study was transfer to the ICU after admission to the hospital among patients residing either in a general medical–surgical ward (“ward”) or transitional care unit (TCU), or death in the ward or TCU in a patient who was “full code” at the time of death (ie, had the patient survived, s/he would have been transferred to the ICU).

The unit of analysis for this study was a 12-hour patient shift, which could begin with a 7 AM  $T_0$  (henceforth, day shift) or a 7 PM  $T_0$  (night shift); in other words, we aimed to predict the occurrence of an event within 12 hours of  $T_0$  using only data available prior to  $T_0$ . A shift in which a patient experienced the primary study outcome is an *event* shift, while one in which a patient did not experience the primary outcome is a *comparison* shift. Using this approach, an individual patient record could consist of both event and comparison shifts, since some patients might have multiple unplanned transfers and some patients might have none. Our basic analytic approach consisted of creating a cohort of event and comparison shifts ( $\sim 10$  comparison shifts were randomly selected for each event shift), splitting the cohort into a derivation dataset (50%) and validation dataset (50%), developing a model using the derivation dataset, then applying the coefficients of the derivation dataset to the validation dataset. Because some event shifts were excluded due to the minimum 4-hour length-of-stay requirement, we also applied model coefficients to these excluded shifts and a set of randomly selected comparison shifts.

Since the purpose of these analyses was to develop models with maximal signal extraction from sparsely collected predictors, we did not block a time period after the  $T_0$  to allow for a reaction time to the alarm. Thus, since some events could occur immediately after the  $T_0$  (as can be seen in the Supporting Information, Appendices, in the online version of this article), our models would need to be run at intervals that are more frequent than 2 times a day.

### Independent Variables

In addition to patients’ age and sex, we tested the following candidate independent variables. Some of these variables are part of the KPMCP risk adjustment model<sup>4,5</sup> and were available electronically for all patients in the cohort. We grouped admission diagnoses into 44 broad diagnostic categories (primary conditions), and admission types into 4 groups (emergency medical, emergency surgical, elective medical, and elective surgical). We quantified patients’ degree of physiologic derangement in the 72 hours preceding hospitalization with a Laboratory-based Acute Physiology Score (LAPS) using 14 laboratory test results prior to hospitalization; we also tested individual laboratory test results obtained after admission to the hospital. We quantified patients’ comorbid illness burden using a Comorbidity Point Score (COPS) based on patients’ preexisting diagnoses over the 12-month period preceding hospitalization.<sup>4</sup> We extracted temperature, heart rate, respiratory rate, systolic blood pressure, diastolic blood pressure, oxygen saturation, and neurological status from the EMR. We also tested the following variables based on specific information extracted from the EMR: shock index (heart rate

**TABLE 1.** Characteristics of Final Study Cohort

	Never Admitted to ICU	Direct Admit to ICU From ED	Unplanned Transfer to ICU*	Other ICU Admission <sup>†</sup>
N	89,269	5963	2880	4310
Age (mean ± SD)	61.26 ± 18.62	62.25 ± 18.13	66.12 ± 16.20	64.45 ± 15.91
Male (n, %)	37,228 (41.70%)	3091 (51.84%)	1416 (49.17%)	2378 (55.17%)
LAPS <sup>‡</sup> (mean ± SD)	13.02 ± 15.79	32.72 ± 24.85	24.83 ± 21.53	11.79 ± 18.16
COPS <sup>§</sup> (mean ± SD)	67.25 ± 51.42	73.88 ± 57.42	86.33 ± 59.33	78.44 ± 52.49
% Predicted mortality risk (mean ± SD)	1.93% ± 3.98%	7.69% ± 12.59%	5.23% ± 7.70%	3.66% ± 6.81%
Survived first hospitalization to discharge <sup>  </sup>	88,479 (99.12%)	5336 (89.49%)	2316 (80.42%)	4063 (94.27%)
Care order on admission				
Full code	78,877 (88.36%)	5198 (87.17%)	2598 (90.21%)	4097 (95.06%)
Partial code	664 (0.74%)	156 (2.62%)	50 (1.74%)	27 (0.63%)
Comfort care	21 (0.02%)	2 (0.03%)	0 (0%)	0 (0%)
DNR	8227 (9.22%)	539 (9.04%)	219 (7.60%)	161 (3.74%)
Comfort care and DNR	229 (0.26%)	9 (0.15%)	2 (0.07%)	2 (0.05%)
No order	1251 (1.40%)	59 (0.99%)	11 (0.38%)	23 (0.53%)
Admission diagnosis (n, %)				
Pneumonia	2385 (2.67%)	258 (4.33%)	242 (8.40%)	68 (1.58%)
Sepsis	5822 (6.52%)	503 (8.44%)	279 (9.69%)	169 (3.92%)
GI bleeding	9938 (11.13%)	616 (10.33%)	333 (11.56%)	290 (6.73%)
Cancer	2845 (3.19%)	14 (0.23%)	95 (3.30%)	492 (11.42%)
Total hospital length of stay (days ± SD)	3.08 ± 3.29	5.37 ± 7.50	12.16 ± 13.12	8.06 ± 9.53

NOTE: All overnight admissions to the study hospitals excluding 66 patients who were removed due to incomplete data. Column categories are mutually exclusive and based on a patient's first hospitalization during the study time period.

Abbreviations: COPS, Comorbidity Point Score; DNR, do not resuscitate; ED, emergency department; GI, gastrointestinal; ICU, intensive care unit; LAPS, Laboratory Acute Physiology Score; SD, standard deviation. \*This group consists of all patients who meet our case definition and includes: 1) patients who had an unplanned transfer to the ICU from the transitional care unit (TCU) or ward; and 2) patients who died on the ward without a DNR order in place at the time of death (ie, who would have been transferred to the ICU had they survived). †This group includes patients admitted directly to the ICU from the operating room, post-anesthesia recovery, or an unknown unit, as well as patients with a planned transfer to the ICU. ‡LAPS point score based on 14 laboratory test results obtained in the 72 hr preceding hospitalization. With respect to a patient's physiologic derangement, the unadjusted relationship of LAPS and inpatient mortality is as follows: a LAPS <7 is associated with a mortality risk of <1%; <7 to 30 with a mortality risk of 1%–5%; 30 to 60 with a mortality risk of 5%–9%; and >60 with a mortality risk of 10% or more. See text and Escobar et al<sup>4</sup> for more details. §COPS point score based on a patient's healthcare utilization diagnoses (during the year preceding admission to the hospital). Analogous to present on admission (POA) coding. Scores can range from 0 to a theoretical maximum of 701, but scores >200 are rare. With respect to a patient's preexisting comorbidity burden, the unadjusted relationship of COPS and inpatient mortality is as follows: a COPS <50 is associated with a mortality risk of <1%; <100 with a mortality risk of 1%–5%; 100 to 145 with a mortality risk of 5%–10%; and >145 with a mortality risk of 10% or more. See text and Escobar et al<sup>4</sup> for more details. ||Numbers for patients who survived last hospitalization to discharge are available upon request.

divided by systolic blood pressure)<sup>13</sup>; care directive status (patients were placed into 4 groups: full code, partial code, do not resuscitate [DNR], and no care directive in place); and a proxy for measured lactate (PML; anion gap/serum bicarbonate × 100).<sup>14–16</sup> For comparison purposes, we also created a retrospective electronically assigned MEWS, which we refer to as the MEWS(re), and we assigned this score to patient records electronically using data from KP HealthConnect.

### Statistical Methods

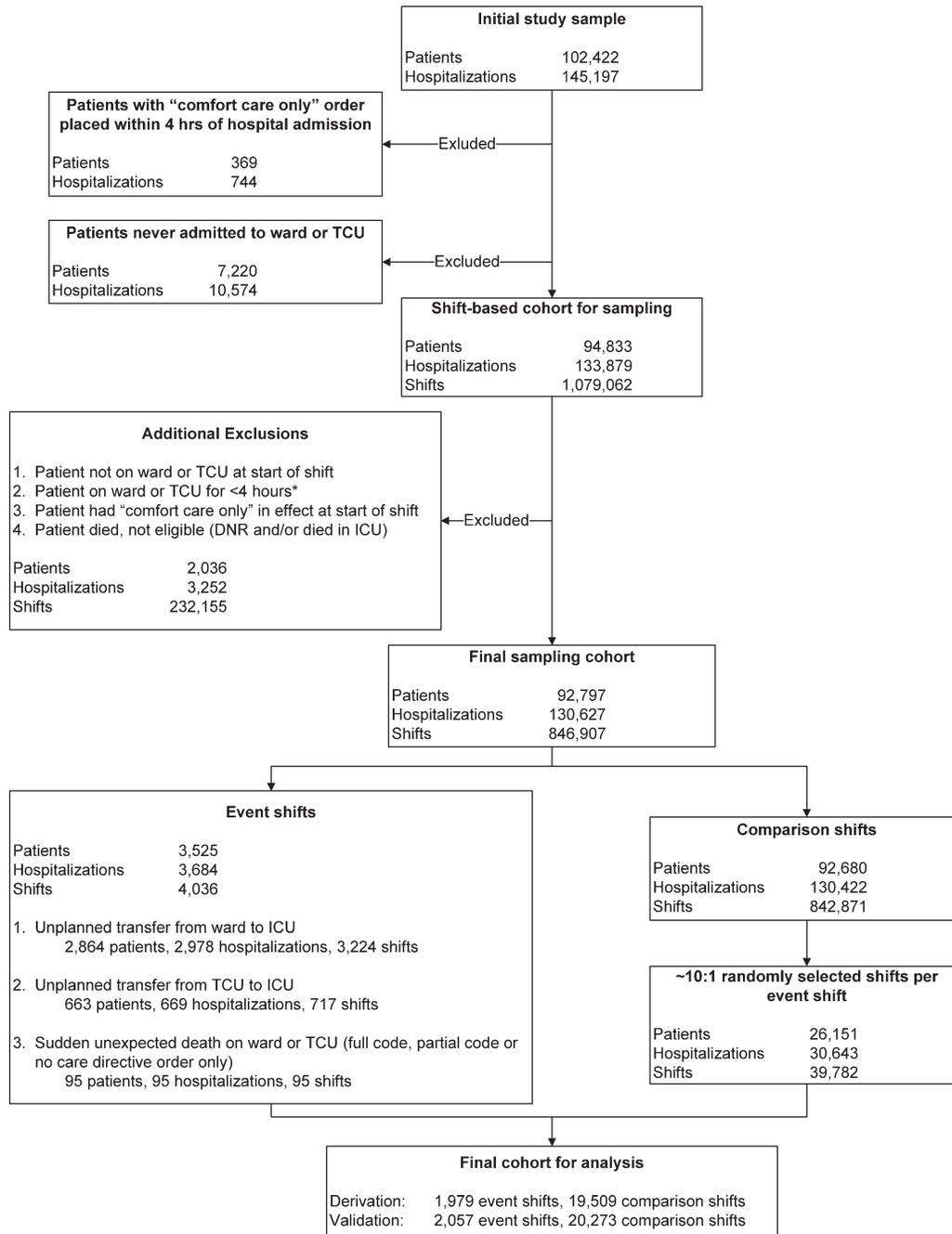
Analyses were performed in SAS 9.1, Stata 10, and R 2.12. Final validation was performed using SAS (SAS Institute Inc., Carey, North Carolina). Since we did not limit ourselves to traditional severity-scoring approaches (eg, selecting the “worst” heart rate in a given time interval), but also included trend terms (eg, change in heart rate over the 24 hours preceding T<sub>0</sub>), the number of potential variables to test was very large. Detailed description of the statistical strategies employed for variable selection is provided in the Supporting Information, Appendices, in the online version of this article. Once variables were selected, our basic approach was to test a series of diagnosis-specific logistic regression submodels using a variety of predictors that included vital signs, vital signs trends (eg, most recent heart rate minus earliest heart rate, heart rate over preceding 24 hours), and other above-mentioned variables.

We assessed the ability of a submodel to correctly distinguish patients who died, from survivors, using the c-statistic, as well as other metrics recommended by Cook.<sup>17</sup> At the end of the modeling process, we pooled the results across all submodels. For vital signs, where the rate of missing data was <3%, we tested submodels in which we dropped shifts with missing data, as well as submodels in which we imputed missing vital signs to a normal value. For laboratory data, where the rate of missing data for a given shift was much greater, we employed a probabilistic imputation method that included consideration of when a laboratory test result became available.

### RESULTS

During the study period, a total of 102,488 patients experienced 145,335 hospitalizations at the study hospitals. We removed 66 patients with 138 hospitalizations for data quality reasons, leaving us with our initial study sample of 102,422 patients whose characteristics are summarized in Table 1. Table 1, in which the unit of analysis is an individual patient, shows that patients who experienced the primary outcome were similar to those patients described in our previous report, in terms of their characteristics on admission as well as in experiencing excess morbidity and mortality.<sup>6</sup>

Figure 1 shows how we developed the analysis cohort, by removing patients with a comfort-care-only



**FIG. 1.** Development of sampling cohort. \*There are 429 event shifts excluded; see text for details. Abbreviations: DNR, do not resuscitate; ICU, intensive care unit; TCU, transitional care unit.

order placed within 4 hours after admission (369 patients/744 hospitalizations) and patients who were never admitted to the ward or TCU (7,220/10,574). This left a cohort consisting of 94,833 patients who experienced 133,879 hospitalizations spanning a total of 1,079,062 shifts. We then removed shifts where: 1) a patient was not on the ward at the start of a shift, or was on the ward for <4 hours of a shift; 2) the patient had a comfort-care order in place at the start of the shift; and 3) the patient died and was ineligible to be a case (the patient had a DNR order in place or died in the ICU). The final cohort eligible for sampling

consisted of 846,907 shifts, which involved a total of 92,797 patients and 130,627 hospitalizations. There were a total of 4,036 event shifts, which included 3,224 where a patient was transferred from the ward to the ICU, 717 from the TCU to the ICU, and 95 where a patient died on the ward or TCU without a DNR order in place. We then randomly selected 39,782 comparison shifts. Thus, our final cohort for analysis included 4,036 event shifts (1,979 derivation/2,057 validation) and 39,782 comparison shifts (19,509/20,273). As a secondary validation, we also applied model coefficients to the 429 event

**TABLE 2.** Event and Comparison Shifts

Predictor	Event Shifts	Comparison Shifts	P
Number	4036	39,782	
Age (mean ± SD)	67.19 ± 15.25	65.41 ± 17.40	<0.001
Male (n, %)	2007 (49.73%)	17,709 (44.52%)	<0.001
Day shift	1364 (33.80%)	17,714 (44.53%)	<0.001
LAPS*	27.89 ± 22.10	20.49 ± 20.16	<0.001
COPS†	116.33 ± 72.31	100.81 ± 68.44	<0.001
Full code‡ (n, %)	3496 (86.2%)	32,156 (80.8%)	<0.001
ICU shift during hospitalization§	3964 (98.22%)	7197 (18.09%)	<0.001
Unplanned transfer to ICU during hospitalization¶	353 (8.8%)	1466 (3.7%)	<0.001
Temperature (mean ± SD)	98.15 (1.13)	98.10 (0.85)	0.009
Heart rate (mean ± SD)	90.30 (20.48)	79.86 (5.27)	<0.001
Respiratory rate (mean ± SD)	20.36 (3.70)	18.87 (1.79)	<0.001
Systolic blood pressure (mean ± SD)	123.65 (23.26)	126.21 (19.88)	<0.001
Diastolic blood pressure (mean ± SD)	68.38 (14.49)	69.46 (11.95)	<0.001
Oxygen saturation (mean ± SD)	95.72% (3.00)	96.47% (2.26)	<0.001
MEWS(re)¶ (mean ± SD)	3.64 (2.02)	2.34 (1.61)	<0.001
% <5	74.86%	92.79%	
% ≥5	25.14%	7.21%	<0.001
Proxy for measured lactate# (mean ± SD)	36.85 (28.24)	28.73 (16.74)	<0.001
% Missing in 24 hr before start of shift**	17.91%	28.78%	<0.001
Blood urea nitrogen (mean ± SD)	32.03 (25.39)	22.72 (18.9)	<0.001
% Missing in 24 hr before start of shift	19.67%	20.90%	<0.001
White blood cell count × 1000 (mean ± SD)	12.33 (11.42)	9.83 (6.58)	<0.001
% Missing in 24 hr before start of shift	21.43%	30.98%	<0.001
Hematocrit (mean ± SD)	33.08 (6.28)	33.07 (5.25)	0.978
% Missing in 24 hr before start of shift	19.87%	29.55%	<0.001

NOTE: Code status, vital sign, and laboratory values measures closest to the start of the shift (7 AM or 7 PM) are used.

Abbreviations: COPS, COmorbidity Point Score; ICU, intensive care unit; LAPS, Laboratory Acute Physiology Score; MEWS(re), Modified Early Warning Score (retrospective electronic); SD, standard deviation. \*LAPS; see Table 1, text, and Escobar et al<sup>4</sup> for more details. †COPS; see Table 1, text, and Escobar et al<sup>4</sup> for more details. ‡Refers to patients who had an active “full code” order at the start of the sampling time frame. §See text for explanation of sampling time frame, and how both cases and controls could have been in the ICU. ¶See text for explanation of how both cases and controls could have experienced an unplanned transfer to the ICU. #MEWS(re); see text and Subbe et al<sup>10</sup> for a description of this score. #Anion gap ÷ bicarbonate) × 100. \*\*Rates of missing data for vital signs are not shown because <3% of the shifts were missing these data.

shifts excluded due to the <4-hour length-of-stay requirement.

Table 2 compares event shifts with comparison shifts. In the 24 hours preceding ICU transfer, patients who were subsequently transferred had statistically significant, but not necessarily clinically significant, differences in terms of these variables. However, missing laboratory data were more common, ranging from 18% to 31% of all shifts (we did not incorporate laboratory tests where ≥35% of the shifts had missing data for that test).

After conducting multiple analyses using the derivation dataset, we developed 24 submodels, a compromise between our finding that primary-condition-specific models showed better performance and the fact that we had very few events among patients with certain primary conditions (eg, pericarditis/valvular heart disease), which forced us to create composite categories (eg, a category pooling patients with pericarditis, atherosclerosis, and peripheral vascular disease). Table 3 lists variables included in our final submodels.

**TABLE 3.** Variables Included in Final Electronic Medical Record-Based Models

Variable	Description
Directive status	Full code or not full code
LAPS*	Admission physiologic severity of illness score (continuous variable ranging from 0 to 256). Standardized and included as LAPS and LAPS squared
COPS†	Comorbidity burden score (continuous variable ranging from 0 to 701). Standardized and included as COPS and COPS squared.
COPS status‡	Indicator for absent comorbidity data
LOS at T <sub>0</sub>	Length of stay in the hospital (total time in hours) at the T <sub>0</sub> ; standardized.
T <sub>0</sub> time of day	7 AM or 7 PM
Temperature	Worst (highest) temperature in 24 hr preceding T <sub>0</sub> ; variability in temperature in 24 hr preceding T <sub>0</sub> .
Heart rate	Most recent heart rate in 24 hr preceding T <sub>0</sub> ; variability in heart rate in 24 hr preceding T <sub>0</sub> .
Respiratory rate	Most recent respiratory rate in 24 hr preceding T <sub>0</sub> ; worst (highest) respiratory rate in 24 hr preceding T <sub>0</sub> ; variability in respiratory rate in 24 hr preceding T <sub>0</sub> .
Diastolic blood pressure	Most recent diastolic blood pressure in 24 hr preceding T <sub>0</sub> transformed by subtracting 70 from the actual value and squaring the result. Any value above 2000 is subsequently then set to 2000, yielding a continuous variable ranging from 0 to 2000.
Systolic pressure	Variability in systolic blood pressure in 24 hr preceding T <sub>0</sub> .
Pulse oximetry	Worst (lowest) oxygen saturation in 24 hr preceding T <sub>0</sub> ; variability in oxygen saturation in 24 hr preceding T <sub>0</sub> .
Neurological status	Most recent neurological status check in 24 hr preceding T <sub>0</sub> .
Laboratory tests‡	Blood urea nitrogen Proxy for measured lactate = (anion gap ÷ serum bicarbonate) × 100 Hematocrit Total white blood cell count

Abbreviations: COPS, COmorbidity Point Score; LAPS, Laboratory Acute Physiology Score; LOS, length of stay. \*LAPS based on 14 laboratory test results obtained in the 72 hr preceding hospitalization. See text and Escobar et al<sup>4</sup> for details. †COPS based on a patient's diagnoses in the 12 mo preceding hospitalization. See text and Escobar et al<sup>4</sup> for details. ‡Indicator variable (for patients in whom a COPS could not be obtained) also included in models. §See text and Supporting Information, Appendices, in the online version of this article for details on imputation strategy employed when values were missing. See Wrenn<sup>14</sup> and Rocktaeschel et al<sup>16</sup> for justification for use of the combination of anion gap and serum bicarbonate.

Table 4 summarizes key results in the validation dataset. Across all diagnoses, the MEWS(re) had c-statistic of 0.709 (95% confidence interval, 0.697–0.721) in the derivation dataset and 0.698 (0.686–0.710) in the validation dataset. In the validation dataset, the MEWS(re) performed best among patients with a set of gastrointestinal diagnoses (c = 0.792; 0.726–0.857) and worst among patients with congestive heart failure (0.541; 0.500–0.620). In contrast, across all primary conditions, the EMR-based models had a c-statistic of 0.845 (0.826–0.863) in the derivation dataset and 0.775 (0.753–0.797) in the validation dataset. In the validation dataset, the EMR-based models also performed best among patients with a set of gastrointestinal diagnoses (0.841; 0.783–0.897) and worst among patients with congestive heart failure (0.683; 0.610–0.755). A negative correlation (R = -0.63) was evident between the number of event shifts in a submodel and the drop in the c-statistic seen in the validation dataset.

We also compared model performance when our datasets were restricted to 1 randomly selected

**TABLE 4.** Best and Worst Performing Submodels in the Validation Dataset

Diagnoses Group*	No. of Shifts in Validation Dataset		c-Statistic	
	Event	Comparison	MEWS(re) <sup>†</sup>	EMR Model <sup>‡</sup>
Acute myocardial infarction	36	169	0.541	0.572
Diseases of pulmonary circulation and cardiac dysrhythmias	40	329	0.565	0.645
Seizure disorders	45	497	0.594	0.647
“Rule out” myocardial infarction	77	727	0.602	0.648
Pneumonia	163	847	0.741	0.801
GI diagnoses, set A <sup>§</sup>	58	942	0.755	0.803
GI diagnoses, set B <sup>  </sup>	256	2,610	0.772	0.806
GI diagnoses, set C <sup>¶</sup>	46	520	0.792	0.841
All diagnosis	2,032	20,106	0.698	0.775

**Abbreviations:** EMR, electronic medical record; GI, gastrointestinal; MEWS(re), Modified Early Warning Score (retrospective electronic). \*Specific International Classification of Diseases (ICD) codes used are detailed in the Supporting Information, Appendices, in the online version of this article. <sup>†</sup>MEWS(re); see text, Supporting Information, Appendices, in the online version of this article, and Subbe et al<sup>10</sup> for more details. <sup>‡</sup>Model based on comprehensive data from EMR; see text, Table 3, and Supporting Information, Appendices, in the online version of this article for more details. <sup>§</sup>This group of diagnoses includes appendicitis, cholecystitis, cholangitis, hernias, and pancreatic disorders. <sup>||</sup>This group of diagnoses includes: gastrointestinal hemorrhage, miscellaneous disorders affecting the stomach and duodenum, diverticulitis, abdominal symptoms, nausea with vomiting, and blood in stool. <sup>¶</sup>This group of diagnoses includes inflammatory bowel disease, malabsorption syndromes, gastrointestinal obstruction, and enteritides.

observation per patient; in these analyses, the total number of event shifts was 3,647 and the number of comparison shifts was 29,052. The c-statistic for the MEWS(re) in the derivation dataset was 0.709 (0.694–0.725); in the validation dataset, it was 0.698 (0.692–0.714). The corresponding values for the EMR-based models were 0.856 (0.835–0.877) and 0.780 (0.756–0.804). We also tested models in which, instead of dropping shifts with missing vital signs, we imputed missing vital signs to their normal value. The c-statistic for the EMR-based model with imputed vital sign values was 0.842 (0.823–0.861) in the derivation dataset and 0.773 (0.752–0.794) in the validation dataset. Lastly, we applied model coefficients to a dataset consisting of 4,290 randomly selected comparison shifts plus the 429 shifts excluded because of the 4-hour length-of-stay criterion. The c-statistic for this analysis was 0.756 (0.703–0.809).

As a general rule, the EMR-based models were more than twice as efficient as the MEWS(re). For example, a MEWS(re) threshold of  $\geq 6$  as the trigger for an alarm would identify 15% of all transfers to the ICU, with 34.4 “false alarms” for each transfer; in contrast, using the EMR-based approach to identify 15% of all transfers, there were 14.5 “false alarms” for each transfer. Applied to the entire KPMCP Northern California Region, using the MEWS(re), a total of 52 patients per day would need to be evaluated, but only 22 per day using the EMR-based approach. If one employed a MEWS(re) threshold of  $\geq 4$ , this would lead to identification of 44% of all transfers, with a ratio of 69 “false alarms” for each transfer; using the EMR, the ratio would be 34 to 1. Across the entire KPMCP, a total of 276 patients per day (or about

19.5 a day per hospital) would need to be evaluated using the MEWS(re), but only 136 (or about 9.5 per hospital per day) using the EMR.

## DISCUSSION

Using data from a large hospital cohort, we have developed a predictive model suitable for use in non-ICU populations cared for in integrated healthcare settings with fully automated EMRs. The overall performance of our model, which incorporates acute physiology, diagnosis, and longitudinal data, is superior to the predictive ability of a model that can be assigned manually. This is not surprising, given that scoring systems such as the MEWS make an explicit tradeoff losing information found in multiple variables in exchange for ease of manual assignment. Currently, the model described in this report is being implemented in a simulated environment, a final safety test prior to piloting real-time provision of probability estimates to clinicians and nurses. Though not yet ready for real-time use, it is reasonable for our model to be tested using the KPHC “shadow” server, since evaluation in a simulated environment constitutes a critical evaluation step prior to deployment for clinical use. We also anticipate further refinement and revalidation to occur as more inpatient data become available in the KPMCP and elsewhere.

A number of limitations to our approach must be emphasized. In developing our models, we determined that, while modeling by clinical condition was important, the study outcome was rare for some primary conditions. In these diagnostic groups, which accounted for 12.5% of the event shifts and 10.6% of the comparison shifts, the c-statistic in the validation dataset was  $< 0.70$ . Since all 22 KPMCP hospitals are now online and will generate an additional 150,000 adult hospitalizations per year, we expect to be able to correct this problem prior to deployment of these models for clinical use. Having additional data will permit us to improve model discrimination and thus decrease the evaluation-to-detection ratio. In future iterations of these models, more experimentation with grouping of International Classification of Diseases (ICD) codes may be required. The problem of grouping ICD codes is not an easy one to resolve, in that diagnoses in the grouping must share common pathophysiology while having a grouping with a sufficient number of adverse events for stable statistical models.

Ideally, it would have been desirable to employ a more objective measure of deterioration, since the decision to transfer a patient to the ICU is discretionary. However, we have found that key data points needed to define such a measure (eg, vital signs) are not consistently charted when a patient deteriorates—this is not surprising outside the research setting, given that nurses and physicians involved in a transfer may be focusing on caring for the patient rather than immediately charting. Given the complexities of

end-of-life-care decision-making, we could not employ death as the outcome of interest. A related issue is that our model does not differentiate between reasons for needing transfer to the ICU, an issue recently discussed by Bapojé et al.<sup>18</sup>

Our model does not address an important issue raised by Bapojé et al.<sup>18</sup> and Litvak, Pronovost, and others,<sup>19,20</sup> namely, whether a patient should have been admitted to a non-ICU setting in the first place. Our team is currently developing a model for doing exactly this (providing decision support for triage in the emergency department), but discussion of this methodology is outside the scope of this article.

Because of resource and data limitations, our model also does not include newborns, children, women admitted for childbirth, or patients transferred from non-KPMCP hospitals. However, the approach described here could serve as a starting point for developing models for these other populations.

The generalizability of our model must also be considered. The Northern California KPMCP is unusual in having large electronic databases that include physiologic as well as longitudinal patient data. Many hospitals cannot take advantage of all the methods described here. However, the methods we employed could be modified for use by hospital systems in countries such as Great Britain and Canada, and entities such as the Veterans Administration Hospital System in the United States. The KPMCP population, an insured population with few barriers to access, is healthier than the general population, and some population subsets are underrepresented in our cohort. Practice patterns may also vary. Nonetheless, the model described here could serve as a good starting point for future collaborative studies, and it would be possible to develop models suitable for use by stand-alone hospitals (eg, recalibrating so that one used a Charlson comorbidity<sup>21</sup> score based on present on-admission codes rather than the COPS).

The need for early detection of patient deterioration has played a major role in the development of “rapid response teams,” as well as scores such as the MEWS. In particular, entities such as the Institute for Healthcare Improvement have advocated the use of early warning systems.<sup>22</sup> However, having a statistically robust model to support an early warning system is only part of the solution, and a number of new challenges must then be addressed. The first is actual electronic deployment. Existing inpatient EMRs were not designed with complex calculations in mind, and we anticipate that some degradation in performance will occur when we test our models using real-time data capture. As Bapojé et al point out, simply having an alert may be insufficient, since not all transfers are preventable.<sup>18</sup> Early warning systems also raise ethical issues (for example, what should be done if an alert leads a clinician to confront the fact that an end-of-life-care discussion needs to occur?). From a research perspective, if

one were to formally test the benefits of such models, it would be critical to define outcome measures other than death (which is strongly affected by end-of-life-care decisions) or ICU transfer (which is often desirable).

In conclusion, we have developed an approach for predicting impending physiologic deterioration of hospitalized adults outside the ICU. Our approach illustrates how organizations can take maximal advantage of EMRs in a manner that exceeds “meaningful use” specifications.<sup>23,24</sup> Our study highlights the possibility of using fully automated EMR data for building and applying sophisticated statistical models in settings other than the highly monitored ICU without the need for additional equipment. It also expands the universe of severity scoring to one in which probability estimates are provided in real time and throughout an entire hospitalization. Model performance will undoubtedly improve over time, as more patient data become available. Although our approach has important limitations, it is suitable for testing using real-time data in a simulated environment. Such testing would permit identification of unanticipated problems and quantification of the degradation of model performance due to “real life” factors, such as delays in vital signs charting or EMR system “brownouts.” It could also serve as the springboard for future collaborative studies, with a broader population base, in which the EMR becomes a tool for care, not just documentation.

#### Acknowledgements

We thank Ms Marla Gardner and Mr John Greene for their work in the development phase of this project. We are grateful to Brian Hoberman, Andrew Hwang, and Marc Flagg from the RIMS group; to Colin Stobbs, Sriram Thiruvengkatachari, and Sundeep Sood from KP IT, Inc; and to Dennis Andaya, Linda Gliner, and Cyndi Vlasallo for their assistance with data-quality audits. We are also grateful to Dr Philip Madvig, Dr Paul Feigenbaum, Dr Alan Whippy, Mr Gregory Adams, Ms Barbara Crawford, and Dr Marybeth Sharpe for their administrative support and encouragement; and to Dr Alan S. Go, Acting Director of the Kaiser Permanente Division of Research, for reviewing the manuscript.

Disclosures: Drs Escobar, Kipnis, Ragins, LaGuardia, and Turk are Kaiser Foundation employees, and Dr Escobar is a member of the Permanente Medical Group, Inc. This work was supported by The Permanente Medical Group, Inc, the Sidney Garfield Memorial Fund (grant 115-9518, “Early detection of impending physiologic deterioration in hospitalized patients”), the Agency for Healthcare Research and Quality (grant 1R01HS018480-01, “Rapid clinical snapshots from the EMR among pneumonia patients”), and the Gordon and Betty Moore Foundation (grant #2663, “Early detection of impending physiologic deterioration: electronic early warning system”).

#### References

- Barnett MJ, Kaboli PJ, Sirio CA, Rosenthal GE. Day of the week of intensive care admission and patient outcomes: a multisite regional evaluation. *Med Care*. 2002;40(6):530–539.
- Ensminger SA, Morales IJ, Peters SG, et al. The hospital mortality of patients admitted to the ICU on weekends. *Chest*. 2004;126(4):1292–1298.
- Luyt CE, Combes A, Aegerter P, et al. Mortality among patients admitted to intensive care units during weekday day shifts compared with “off” hours. *Crit Care Med*. 2007;35(1):3–11.
- Escobar G, Greene J, Scheirer P, Gardner M, Draper D, Kipnis P. Risk adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care*. 2008;46(3):232–239.
- van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. *J Clin Epidemiol*. 2010;63(7):798–803.
- Escobar GJ, Greene JD, Gardner MN, Marelich GP, Quick B, Kipnis P. Intra-hospital transfers to a higher level of care: contribution to

- total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *J Hosp Med*. 2011;6(2):74–80.
7. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med*. 1999;25(12):1360–1366.
  8. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci Transl Med*. 2010;2(48):48ra65.
  9. Subbe CP, Gao H, Harrison DA. Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med*. 2007;33(4):619–624.
  10. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a Modified Early Warning Score in medical admissions. *Q J Med*. 2001;94:521–526.
  11. Subbe CP, Davies RG, Williams E, Rutherford P, Gemmel L. Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions. *Anaesthesia*. 2003;58(8):797–802.
  12. MERIT Study Investigators. Introduction of the medical emergency team (MET) system: a cluster-randomized controlled trial. *Lancet*. 2005;365(9477):2091–2097.
  13. Keller AS, Kirkland LL, Rajasekaran SY, Cha S, Rady MY, Huddleston JM. Unplanned transfers to the intensive care unit: the role of the shock index. *J Hosp Med*. 2010;5(8):460–465.
  14. Wrenn K. The delta (delta) gap: an approach to mixed acid-base disorders. *Ann Emerg Med*. 1990;19(11):1310–1313.
  15. Williamson JC. Acid-base disorders: classification and management strategies. *Am Fam Physician*. 1995;52(2):584–590.
  16. Rocktaeschel J, Morimatsu H, Uchino S, Bellomo R. Unmeasured anions in critically ill patients: can they predict mortality? *Crit Care Med*. 2003;31(8):2131–2136.
  17. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928–935.
  18. Bapoje SR, Gaudiani JL, Narayanan V, Albert RK. Unplanned transfers to a medical intensive care unit: causes and relationship to preventable errors in care. *J Hosp Med*. 2011;6(2):68–72.
  19. Litvak E, Pronovost PJ. Rethinking rapid response teams. *JAMA*. 2010;304(12):1375–1376.
  20. Winters BD, Pham J, Pronovost PJ. Rapid response teams—walk, don't run. *JAMA*. 2006;296(13):1645–1647.
  21. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal populations: development and validation. *J Chronic Dis*. 1987;40:373–383.
  22. Institute for Healthcare Improvement. Early Warning Systems: *The Next Level of Rapid Response*. 2011. <http://www.ihl.org/IHI/Programs/AudioAndWebPrograms/ExpeditionEarlyWarningSystems/TheNextLevelofRapidResponse.htm?player=wmp>. Accessed 4/6/11.
  23. Bowes WA. Assessing readiness for meeting meaningful use: identifying electronic health record functionality and measuring levels of adoption. *AMIA Annu Symp Proc*. 2010;2010:66–70.
  24. Medicare and Medicaid Programs; Electronic Health Record Incentive Program. Final Rule. *Fed Reg*. 2010;75(144):44313–44588.