

SPECIAL ARTICLE

Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration

Gabriel J. Escobar, M.D., Vincent X. Liu, M.D., Alejandro Schuler, Ph.D.,
Brian Lawson, Ph.D., John D. Greene, M.A., and Patricia Kipnis, Ph.D.

ABSTRACT

BACKGROUND

Hospitalized adults whose condition deteriorates while they are in wards (outside the intensive care unit [ICU]) have considerable morbidity and mortality. Early identification of patients at risk for clinical deterioration has relied on manually calculated scores. Outcomes after an automated detection of impending clinical deterioration have not been widely reported.

METHODS

On the basis of a validated model that uses information from electronic medical records to identify hospitalized patients at high risk for clinical deterioration (which permits automated, real-time risk-score calculation), we developed an intervention program involving remote monitoring by nurses who reviewed records of patients who had been identified as being at high risk; results of this monitoring were then communicated to rapid-response teams at hospitals. We compared outcomes (including the primary outcome, mortality within 30 days after an alert) among hospitalized patients (excluding those in the ICU) whose condition reached the alert threshold at hospitals where the system was operational (intervention sites, where alerts led to a clinical response) with outcomes among patients at hospitals where the system had not yet been deployed (comparison sites, where a patient's condition would have triggered a clinical response after an alert had the system been operational). Multivariate analyses adjusted for demographic characteristics, severity of illness, and burden of coexisting conditions.

RESULTS

The program was deployed in a staggered fashion at 19 hospitals between August 1, 2016, and February 28, 2019. We identified 548,838 non-ICU hospitalizations involving 326,816 patients. A total of 43,949 hospitalizations (involving 35,669 patients) involved a patient whose condition reached the alert threshold; 15,487 hospitalizations were included in the intervention cohort, and 28,462 hospitalizations in the comparison cohort. Mortality within 30 days after an alert was lower in the intervention cohort than in the comparison cohort (adjusted relative risk, 0.84, 95% confidence interval, 0.78 to 0.90; $P < 0.001$).

CONCLUSIONS

The use of an automated predictive model to identify high-risk patients for whom interventions by rapid-response teams could be implemented was associated with decreased mortality. (Funded by the Gordon and Betty Moore Foundation and others.)

From the Systems Research Initiative, Kaiser Permanente Division of Research, Oakland (G.J.E., V.X.L., A.S., B.L., J.D.G., P.K.), the Intensive Care Unit, Kaiser Permanente Medical Center, Santa Clara (V.X.L.), and Unlearn.AI, San Francisco (A.S.) — all in California. Address reprint requests to Dr. Escobar at the Systems Research Initiative, Kaiser Permanente Division of Research, 2000 Broadway Ave., Oakland, CA 94612, or at gabriel.escobar@kp.org.

N Engl J Med 2020;383:1951-60.

DOI: 10.1056/NEJMsa2001090

Copyright © 2020 Massachusetts Medical Society.

ADULTS WHOSE CONDITION DETERIORATES in general medical–surgical wards have considerable morbidity and mortality.^{1–5} Efforts at early detection of clinical deterioration in inpatients who are outside the intensive care unit (ICU) have used manually calculated scores (e.g., the National Early Warning Score⁶) in which chart abstraction of vital signs and point assignment that is based on these values are performed manually; if a patient's score exceeds a threshold, a rapid-response team is called. Some studies have described automated vital-signs triggers⁷ and automated versions of the National Early Warning Score⁸ or similar⁹ scores. Several investigators have developed complex predictive models, suitable for real-time use with electronic health records (EHRs), for early detection of deterioration in a patient's condition,^{10–12} including one model that was tested in a randomized trial.¹³ These EHR-based models include laboratory tests and information about coexisting conditions; they can involve complex calculations.

We previously described an automated early warning system that identifies patients at high risk for clinical deterioration. Detection is achieved with the use of a predictive model (the Advance Alert Monitor [AAM] program)¹⁴ that identifies such patients. Beginning in November 2013, we conducted a pilot test of this program in 2 hospitals in Kaiser Permanente Northern California (KPNC), an integrated health care delivery system that owns 21 hospitals.^{15–18} The system generates AAM scores that predict the risk of unplanned transfer to the ICU or death in a hospital ward among patients who have “full code” orders (i.e., patients who wish to have cardiopulmonary resuscitation performed in the event that they have a cardiac arrest). The alerts provide 12-hour warnings¹⁴ and do not require an immediate response from clinicians. Given encouraging results from this program,^{18,19} the KPNC leadership deployed the AAM program in its 19 remaining hospitals on a staggered schedule.

In this article, we describe the effect of the program deployment in these 19 hospitals over a 3.5-year period. We compared outcomes at sites where the program was operational (intervention population) with outcomes at sites where it had not yet been deployed (comparison population, which involved patients whose conditions would have triggered alerts had the system been operational).

METHODS

PREDICTIVE MODEL TO IDENTIFY PATIENTS AT HIGH RISK FOR CLINICAL DETERIORATION

Our study included all 21 hospitals (including the 2 pilot sites) in the KPNC system^{2,4,20–23} that had been using the Epic EHR system (www.epic.com) since mid-2010. We used a discrete-time, logistic-regression model to generate hourly AAM scores.¹⁴ The model was based on 649,418 hospitalizations (including 19,153 hospitalizations in which the patients' condition deteriorated) involving 374,838 patients 18 years of age or older who had been admitted to KPNC hospitals between January 1, 2010, and December 31, 2013. Predictors included laboratory tests, individual vital signs, neurologic status, severity of illness and longitudinal indexes of coexisting conditions, care directives, and health services indicators (e.g., length of stay). As instantiated in the Epic EHR system, an AAM score of 5 (alert threshold) indicates a 12-hour risk of clinical deterioration of 8% or more. At this threshold, the model generates one new alert per day per 35 patients, with a C statistic of 0.82 and 49% sensitivity.

STUDY POPULATION AND INTERVENTION

The eligible population consisted of adults 18 years of age or older who had initially been admitted to a general medical–surgical ward or step-down unit, including patients who had initially been admitted to a surgical area and who were then subsequently admitted to one of these units. The target population included eligible patients whose condition reached the alert threshold at sites where the program was operational (intervention cohort; alerts led to a clinical response) or not (comparison cohort; usual care, with no alerts). The comparison cohort also included all the patients who had been admitted to any of the study hospitals in the 1 year before the introduction of the intervention in the first hospital (historical controls). The nontarget population included all the patients whose condition did not reach the alert threshold.

The automated system scanned a patient's data and assigned separate scores for the following three variables: vital signs and laboratory test results (assessed at admission and hourly with the Laboratory-based Acute Physiology Score, version 2 [LAPS2], on a scale from 0 to 414, with higher scores indicating greater physiologic in-

stability), chronic coexisting conditions at admission (Comorbidity Point Score, version 2 [COPS2], on which 12-month scores range from 0 to 1014, with higher scores indicating a worse burden of coexisting conditions), and deterioration risk (according to the AAM, with risk scores ranging from 0 to 100%, and higher scores indicating a greater risk of clinical deterioration).^{14,21} The LAPS2 and COPS2 scores, which are assigned to all hospitalized adults, are scalar values that facilitate the characterization and description of patients' vital signs plus laboratory test results and their coexisting conditions separately. Patients with the care directives "full code," "partial code" (i.e., patients allow some, but not all, resuscitation procedures), and "do not resuscitate" are assigned AAM scores; AAM scores are not assigned to patients in the ICU or to patients who have a care directive of "comfort care only," who were excluded from our main analyses.

In order to minimize alert fatigue, automated-system results were not shown directly to hospital staff. Specially trained registered nurses monitored alerts remotely. If the AAM score reached the threshold, the nurses working remotely performed an initial chart review and contacted the rapid-response nurse on the ward or step-down unit, who then initiated a structured assessment and contacted the patient's physician. The physician then could initiate a clinical rescue protocol (which could include proactive transfer to the ICU), an urgent palliative care consultation, or both. Subsequently, the nurses working remotely monitored patients' status, ensuring adherence to the performance standards of the AAM program.²⁴ At active sites, registered nurses on the rapid-response team were staffed 24 hours a day, 7 days a week, and did not have regular patient assignments. Implementation teams ensured that the clinical staff at the study sites received training on all the components of the program (Section S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org).^{15-18,24} This study was approved by the KPNC Institutional Review Board for the Protection of Human Subjects.

STAGGERED DEPLOYMENT

After implementing the program at 2 hospitals (pilot sites), we ranked the remaining 19 hospitals according to expected numbers of outcomes. The health system leadership agreed that the 3 hospitals with the highest expected numbers

would go first, using a random sequence. The staggered deployment sequence for the 16 other hospitals was based on operational and geographic criteria, which were dictated by the limited availability of the implementation teams. All 21 hospitals ultimately adopted the program.

ANALYTIC STRATEGY

We estimated that the study would have power to detect a decrease in 90-day mortality attributed to the program, assuming that two or three hospitals (excluding the two pilot sites) would adopt the program on the same date every 2 months. We calculated the study power to detect 90-day mortality that was 10% and 20% lower in the target population at the intervention sites than in the population at the comparison sites. Assuming the deployment of a new pair or triad of hospitals every 2 months, we estimated that the study would have more than 80% power to detect an effect size of more than 20% with a 10-month follow-up. Subsequently, because of internal requests to evaluate the program as soon as possible, we elected to use a 30-day time frame for outcomes.

The study period (August 1, 2015, to February 28, 2019) included the year before the intervention was deployed (August 1, 2015, to July 31, 2016). The principal independent variable was the status of the AAM program (operational or not) at a hospital. The unit of analysis was a patient's hospitalization, which in some cases involved linking multiple hospital stays within KPNC, in which interhospital transport is common.^{21,23}

OUTCOMES

The principal dependent variable was mortality within 30 days after an AAM alert. We also analyzed the following secondary outcomes: ICU admission; length of stay in the hospital; 30-day mortality after admission; and favorable status at 30 days (defined as the patient being alive, not in the hospital, and not having been rehospitalized), which was analyzed post hoc.

Attention that was given to patients whose condition triggered an alert could harm other patients.²⁵ Given resource limitations, our ability to quantify unintended harm was limited to analyzing mortality and length of stay in the hospital among patients in a ward or step-down unit whose condition did not trigger an alert (eligible, nontarget population) and among patients who had initially been admitted to the

ICU. For these analyses involving patients without alerts, we used the time of hospital admission as the starting point (T_0).

DATA COLLECTION

We captured data regarding the demographic characteristics of the patients (including Kaiser Foundation Health Plan [KFHP] coverage), sequence of hospital units (ward, ICU, etc.) where the patients stayed, length of stay in the hospital, mortality, and rehospitalization.^{2,4,20-23} We grouped patients' diagnoses into 30 primary conditions according to the *International Classification of Diseases, 9th and 10th Revisions*,^{21,22} using classification software from the Healthcare Cost and Utilization Project (www.ahrq.gov/data/hcup). We classified patients' care directives as "full code" or "not full code" (which included the directives "partial code," "do not resuscitate," and "comfort care only")²¹ and assigned Charlson comorbidity index scores to patients.²⁶ We excluded hospital records that did not include data on stays in a ward or step-down unit; missing or clearly erroneous data for the AAM score, LAPS2, and COPS2 were imputed or truncated (Section S5).^{14,21}

At the pilot sites, scores that were calculated on an external server were displayed every 6 hours on EHR dashboards that were visible to clinicians.¹⁶ After the decision was made to use hourly remote monitoring that would not be displayed to hospital clinicians, score calculation was moved to different servers and then to the predictive modeling module of the Epic EHR system. This meant that, although the predicted risk was the same, the apparent threshold that was used by the nurses working remotely changed. Moreover, before the activation of the AAM program at the first hospital, AAM scores were not available, nor were they available at any site before the activation of the program. To ensure uniform measurement in the periods before and after deployment of the program, we based analyses on retrospective data from the Epic Clarity data warehouse, not on scores assigned by the system. We assigned all the patients at all the sites retrospective LAPS2, COPS2, and AAM scores on admission, every hour, any time a hospital unit change occurred, and (for patients at active sites) any time an alert was issued. For our principal analyses, T_0 was the time of the first retrospective or actual alert for a patient during a hospitalization. For outcomes among patients in the

nontarget population and among those admitted directly to the ICU, T_0 was the time of hospital admission.

STATISTICAL ANALYSIS

Our primary analysis assessed the effect of the intervention in the target cohort (with regard to ICU admission, mortality, length of stay, and favorable status after discharge). Using the same approach, we also evaluated the intervention effect in the nontarget population and the ICU cohort to assess for possible harm. Hospitalizations were assigned to intervention or comparison status according to their first alert date (for hospitalizations in the target cohort) or admission date (for those in the nontarget population and the ICU cohort).

We used generalized linear models to estimate the intervention effect with a fixed estimate for the intervention and fixed hospital effects, controlling for secular trends²⁷ and patient covariates for the target, nontarget, and ICU cohorts separately. We modeled secular time trends as restricted cubic spline functions with five knots placed at equally spaced percentiles of time and a truncated-power-function basis.²⁸⁻³⁰

All the models were adjusted for age, sex, season (whether admission occurred in December through March), KFHP coverage, care directive (at the time of first alert for the target cohort and on admission for the eligible and entire hospitalized cohorts), COPS2, LAPS2 at admission, and diagnosis. The target-population models also include the first alert value and elapsed hours from admission to the first alert.

We modeled binary outcomes (mortality, ICU admission, and favorable status within 30 days) using a Poisson distribution with a log link to estimate the adjusted relative risk of the intervention episodes, as compared with the no-intervention episodes.³¹ Missing or censored values were set to the favorable event for binary outcomes. We used a competing-risk Cox proportional-hazards model to assess the effect of the intervention on the length of stay in the hospital (time to hospital discharge), with death as a competing risk.^{32,33} In this model, the hazard rate ratio refers to the intervention instantaneous rate of discharge from the hospital divided by the instantaneous rate of discharge from the hospital in the comparison group. Thus, a rate ratio of more than 1 indicates that the intervention shortened the time to discharge. Cox proportional-

hazards models were used in a post hoc analysis of the effect of the intervention on mortality over the course of the entire study period.³⁴⁻³⁷ We assessed the proportional-hazards assumption by testing whether the Schoenfeld residuals were associated with time.³⁸

We applied these models to 1000 bootstrap samples that were generated in two stages. First, we obtained a random sample of n_i patients (where n_i is the number of unique patients in hospital i) with replacement within each hospital; we then retained all hospitalizations for selected patients who had multiple hospitalizations.³⁹ We constructed 95% confidence intervals by taking the 2.5th and 97.5th percentiles as the lower and upper boundaries. We did not adjust the confidence intervals for multiple comparisons, and we report a two-sided P value only for the primary outcome.

Since the intervention protocols differed from the protocol that was used in the 2 pilot-site hospitals, we elected to conduct our analyses using only the other 19 hospitals. As a sensitivity analysis, we replicated our analyses using data from all 21 hospitals starting in November 2012, which was 1 year before the program began at the first pilot site.

The analyses were conducted with the use of R software, version 3.5.2, and SAS software, version 9.4 (SAS Institute). Quantification of the process measures is described in Section S6.

RESULTS

STUDY POPULATION

We identified 633,430 hospitalizations during the study period. After excluding 17,042 hospitalizations for which the hospital location could not be determined and 769 hospitalizations for obstetric reasons, 615,619 hospitalizations (548,838 in the eligible population and 66,781 in the ICU cohort) involving 354,489 patients were included in the analysis (Fig. 1). Events for which a patient's condition triggered an alert at a hospital different from the one at which the patient had been admitted were rare (<1%).

UNADJUSTED ANALYSES

Table 1 shows the eligible population, divided into the target and nontarget populations, with the intervention and comparison cohorts in the target population (Section S3). Patients who reached the alert threshold were sicker than pa-

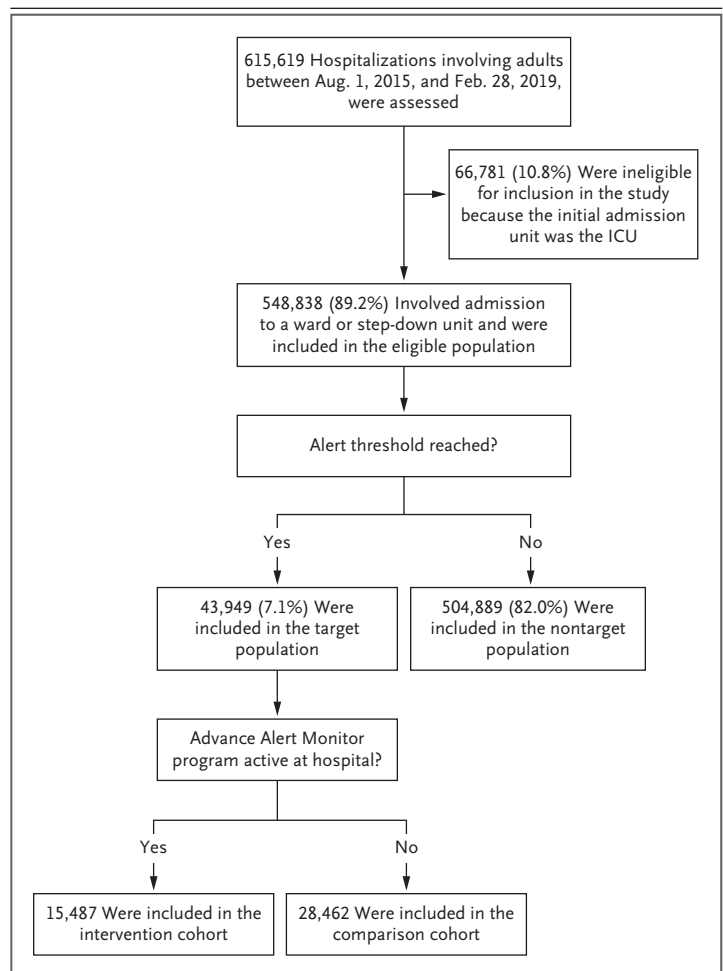


Figure 1. Study Cohorts.

Eligible patients were those whose initial hospital unit after admission was the ward or step-down unit and those who had initially been admitted to a surgical area and whose first postsurgical unit was the ward or step-down unit. Patients with direct admission to the intensive care unit (ICU) were not eligible. (Information about patients who had been admitted to the ICU is provided in Section S3.) Eligible hospitalizations were assigned to the target or nontarget population on the basis of the study definitions. The target population consisted of patients who had a hospitalization in which the patient's condition reached the alert threshold of the Advance Alert Monitor program. If the alert occurred at a hospital where the program was active, the hospitalization was included in the intervention cohort; if the program was not active at the hospital, the hospitalization was included in the comparison cohort.

tients in the ward or step-down unit who did not reach the alert threshold. Table 1 also shows the unadjusted comparisons between 15,487 hospitalizations (involving 13,274 patients) in the intervention cohort and 28,462 hospitalizations (involving 23,797 patients) in the comparison cohort. Patients in the intervention cohort, as compared with those in the comparison cohort, had a lower unadjusted incidence of ICU admis-

sion (17.7% vs. 20.9%), a shorter length of stay among survivors (6.5 days vs. 7.2 days), and lower mortality within 30 days after an event reaching the alert threshold (15.8% vs. 20.4%).

ADJUSTED ANALYSES

Figure 2 shows the staggered deployment sequence of the hospitals, the numbers of patients in the target population, and unadjusted mortality within 30 days after an event reaching the alert threshold. Table 2 shows adjusted results for the target population and the nontarget population. The intervention condition (alerts led to a clinical response) was associated with lower mortality within 30 days after an event reaching the alert threshold than the comparison condition (usual care, with no alerts) (adjusted relative risk, 0.84; 95% confidence interval [CI], 0.78 to 0.90; $P < 0.001$), without worse outcomes in the nontarget population. We estimated an absolute difference of 3.8 percentage points in mortality within 30 days after an event reaching the alert threshold between the intervention cohort and the comparison cohort. This difference translated into 3.0 deaths (95% CI, 1.2 to 4.8) avoided per 1000 eligible patients or to 520 deaths (95% CI, 209 to 831) per year over the 3.5-year study period among approximately 153,000 annual hospitalizations. The intervention was also associated with a lower incidence of ICU admission, a higher percentage of patients with a favorable status 30 days after the alert, a shorter length of stay, and longer survival.

Examination of individual hospital coefficients showed limited variation in mortality outcomes across the study hospitals (Table S8). The individual hospital risk ratios for mortality within 30 days after an alert relative to the first hospital at which the AAM program became active varied between 0.9 and 1.5.

SENSITIVITY ANALYSES AND PROCESS MEASURES

Results from our sensitivity analyses were similar to those of the primary analyses and were also favorable in the intervention cohort. Patients in the intervention cohort were less likely than those in the comparison cohort to die without a referral for palliative care. We did not observe clinically significant differences in vital-sign measurements, therapies, and changes in care directives between the intervention cohort and the comparison cohort (Tables S11 through S15). This result indicates that we were not able to

identify changes in process measures that may have led to the observed improvements in outcomes associated with the intervention.

DISCUSSION

In this study, we quantified beneficial hospital outcomes — lower mortality, a lower incidence of ICU admission, and a shorter length of hospital stay — that were associated with staggered deployment of an automated predictive model that identifies patients at high risk for clinical deterioration. Unlike many scores currently in use, AAM is fully automated, takes advantage of detailed EHR data, and does not require an immediate response by hospital staff. These factors facilitated its incorporation into a rapid-response system that uses remote monitoring, thus shielding providers from alert fatigue. The AAM program is based on standardized workflows for all aspects of care of patients whose condition is deteriorating in wards, including aspects of care involving clinical rescue, palliative care, provider education, and system administration.²⁴

The magnitude of the effects we found are consistent with other studies of rapid-response systems. With respect to complex automated scores, it is important to consider the rigorously executed randomized trial conducted by Kollef et al.¹³ Those authors found a shorter length of stay in the hospital among patients who had an early-warning-system alert displayed to clinicians than among patients for whom alerts were not displayed, but they also found no change in the incidence of ICU admission or in in-hospital mortality. In a very large study of a rapid-response system that used manual scores, Chen et al.⁴⁰ found that preexisting favorable trends in mortality in 292 Australian hospitals continued, with additional improvement among low-risk patients. Priestley et al.⁴¹ compared outcomes in a single hospital before and after the implementation of a rapid-response system that used manual scoring. They found lower inpatient mortality than we did but did not report 30-day mortality and had equivocal findings regarding the length of stay in the hospital. Bedoya et al.⁸ used an automated version of the National Early Warning Score and found no change in the incidence of ICU admission or in in-hospital mortality. Bedoya et al. also found that there was clinician frustration with excessive alerts and noted that the score was largely ignored by frontline nursing staff.

Table 1. Characteristics of the Study Population.*

Characteristic	Eligible Population	Nontarget Population	Target Population	
			Intervention Cohort	Comparison Cohort
No. of hospitalizations	548,838	504,889	15,487	28,462
No. of patients	326,816	313,115	13,274	23,797
Inpatient admission (%)	75.6	74.4	88.7	90.4
Admission for observation (%)	24.4	25.6	11.3	9.6
Age (yr)	66.1±17.5	66.0±17.6	68.4±15.6	67.2±16.2
Male sex (%)	48.0	47.4	55.1	54.3
Admission from emergency department (%)	75.9	74.3	94.3	93.5
Charlson comorbidity index†				
Median score (IQR)	3.0 (1.0–5.0)	2.0 (1.0–5.0)	4.0 (2.0–7.0)	4.0 (2.0–7.0)
Score ≥4 (%)	40.9	39.5	57.9	56.3
COPS2‡				
Mean score	50.6±49.8	48.6±48.5	74.7±60.1	71.9±58.2
Score ≥65 (%)	30.9	29.3	49.6	48.4
Admission LAPS2§				
Mean score	58.7±37.9	55.3±35.9	97.8±36.7	98.3±38.1
Score ≥110 (%)	10.7	8.3	37.0	38.5
Full code on admission (%)¶	94.0	94.0	93.2	95.4
Patients with any admission to the ICU during current hospitalization (%)	4.5	3.2	17.7	20.9
Patients who died in the hospital (%)	2.1	1.1	9.8	14.4
Length of stay in the hospital (days)				
Among patients who survived	3.5±4.5	3.2±3.7	6.5±8.0	7.2±10.1
Among patients who died	7.5±10.7	6.0±7.5	9.1±12.0	9.0±13.3
Outcome within 30 days after admission (%)				
Death	5.4	4.2	15.5	19.9
Favorable outcome	83.2	84.9	66.3	62.5
Outcome within 30 days after first alert (%)				
Death	—	—	15.8	20.4
Favorable outcome	—	—	66.3	62.5

* Plus–minus values are means ±SD. The unit of analysis is a patient's hospitalization, which could include multiple linked hospital stays for patients who were transported between hospitals. Since patients could have multiple hospitalizations, the values for patients' characteristics are those at the time of a given hospitalization. Additional details are provided in Section S3. ICU denotes intensive care unit, and IQR interquartile range.

† Scores on the Charlson comorbidity index range from 0 to 40, with higher scores indicating a greater burden of coexisting conditions. Scores were calculated using the methods of Deyo et al.²⁶

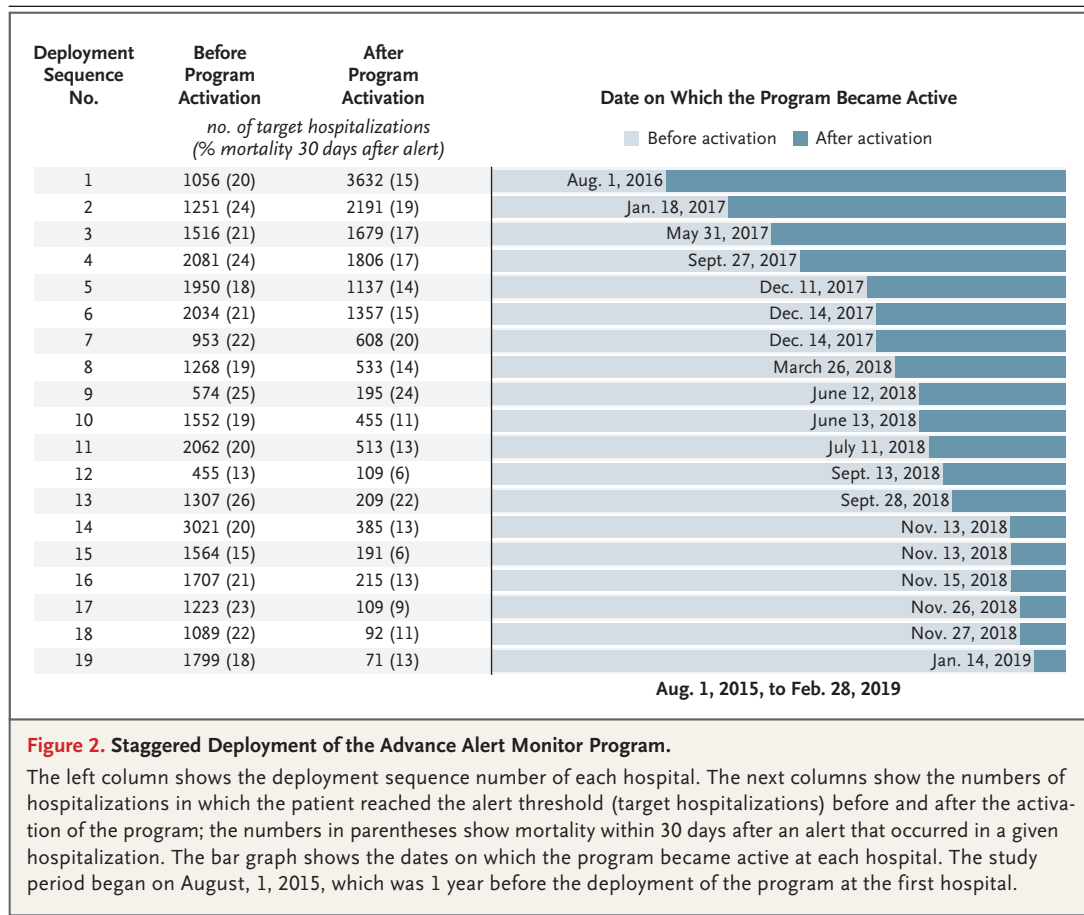
‡ The scale for the Comorbidity Point Score, version 2 (COPS2), ranges from 0 to 1014, with higher scores indicating an increasing burden of coexisting conditions. The score is assigned on the basis of all the diagnoses received by a patient in the 12 months before the index hospitalization. The univariate relationship of COPS2 to 30-day mortality is as follows: a score of 0 to 39 is associated with 1.7% mortality; a score of 40 to 64 with 5.2%; and a score of 65 or higher with 9.0%. Details are provided by Escobar et al.²¹

§ Scores on the Laboratory-based Acute Physiology Score, version 2 (LAPS2), range from 0 to 414, with higher scores indicating increasing physiologic abnormalities. The score is assigned on the basis of a patient's worst vital signs, results on pulse oximetry, neurologic status, and 16 laboratory test results obtained in the preceding 24 hours (hourly and discharge LAPS2) or 72 hours (admission LAPS2). The univariate relationship of an admission LAPS2 to 30-day mortality is as follows: a score of 0 to 59 is associated with 1.0% mortality; a score of 60 to 109 with 5.0%; and a score of 110 or higher with 13.7%. Details are provided by Escobar et al.²¹

¶ Care directives were classified as "full code" or "not full code" (which included the care directives "partial code," "do not resuscitate," and "comfort care only").

|| Favorable status at 30 days indicates that, at 30 days after an alert (target population) or at 30 days after admission (remaining eligible hospitalizations; nontarget population) the patient was alive, was not in the hospital, and had not been readmitted at any time.

Our study has various strengths. It was based on a large, multicenter cohort and included multiple outcomes. We adjusted the analyses for severity of illness, which is assessed hourly with the LAPS2 in KPNC. Our results are statistically robust, with all the sensitivity analyses trending in the same direction. Furthermore, KPNC has made a major effort to ensure consistent implementation of the program.²⁴



Our study also has important limitations. It does not have the methodologic rigor of the randomized, controlled trial conducted by Kollef et al.; that trial had a much smaller sample size, however.¹³ It is not possible for us to identify the exact timing and nature of the clinicians' actions that occurred after an alert.²⁴ Although it has recently become feasible,⁴² we did not have the technical capability or organizational approval for patient-level randomization. Moreover, given data limitations, technical obstacles, and a complex intervention with multiple components, we cannot determine the exact causes of the observed associations. This limitation is also a result of the fact that iterative improvements in computing infrastructure, documentation practices, and workflows were needed to ensure sustainable rollout. Furthermore, process-measure analyses did not show consistent significant associations with the intervention. Thus, we cannot rule out the possibility that improved outcomes were the result of broad institutional cultural change rather than the intervention per se.

Another limitation to generalizability is the study cohort, which consisted of a population of insured patients who were cared for in a highly integrated system in which baseline hospitalization rates were decreasing.²³ Given an increasing threshold for admission, our population of patients may differ from those in other hospital cohorts.

Reflection on our findings suggests future directions for research. One direction is to quantify the relative contributions of the predictive model and the clinical rescue and palliative care processes. Automated scores have statistical performance that is superior to scores such as the National Early Warning Score,⁴³ but it is unclear whether scores that use newer approaches (e.g., so-called bespoke models⁴⁴) will necessarily result in better outcomes. This is because, as Bedoya et al.⁸ point out, clinicians might not use the predictions. A second area relates to how notifications are handled. Although the use of remote monitoring in KPNC appears to have been successful, it is not the only way in which

Table 2. Adjusted Outcomes in the Eligible Population, with Comparison between the Intervention Cohort and Comparison Cohort.*

Variable	Study Population	Adjusted Relative Risk or Hazard Rate Ratio (95% CI)
Target population		
No. of hospitalizations	43,949	
No. of patients	35,669	
ICU admission within 30 days after alert		0.91 (0.84–0.98)
Death within 30 days after alert		0.84 (0.78–0.90)
Favorable status at 30 days after alert†		1.04 (1.02–1.06)
Hospital discharge, as assessed by proportional-hazards analysis		1.07 (1.03–1.11)
Survival, as assessed by proportional-hazards analysis		0.83 (0.78–0.89)
Nontarget population		
No. of hospitalizations	504,889	
No. of patients	313,115	
ICU admission within 30 days after admission		0.94 (0.89–0.99)
Death within 30 days after admission		0.97 (0.93–1.02)
Favorable status 30 days after admission†		1.00 (0.99–1.00)
Hospital discharge, as assessed by proportional-hazards analysis		0.98 (0.97–0.99)
Survival, as assessed by proportional-hazards analysis		0.99 (0.96–1.03)

* The analysis included 548,838 hospitalizations and 326,816 patients (a patient could be included in both the target and nontarget populations, so the numbers of patients do not sum to 326,816). For the first three analyses (ICU admission, mortality, and favorable status within 30 days after an alert), the adjusted relative risk is for whether the patient was in the intervention condition (alerts led to a clinical response), as compared with patients in the comparison condition (usual care, with no alerts). For the nontarget population, the analytic approach was the same as for the target population, except that the cohorts involved patients on the ward whose condition did not trigger an alert; since there was no alert, we used 30-day mortality. We used Cox proportional-hazard models to assess the effects of the intervention on the hospital length of stay, with censoring of a patient's data at the time of death, and long-term survival (median follow-up in the target population, 0.8 years [IQR, 0.1 to 1.8]; median follow-up in the nontarget population, 1.4 years [IQR, 0.5 to 2.4]; maximum follow-up in both populations, 3.6 years). For the hospital length of stay, the hazard rate ratio refers to the instantaneous rate of discharge from the hospital divided by the instantaneous rate of discharge from the hospital in the comparison group; a rate ratio greater than 1 indicates that the intervention shortened the time to discharge. The hazard rate ratio corresponding to long-term survival refers to the long-term mortality in the intervention group as compared with the comparison group; a ratio lower than 1 indicates lower mortality in the intervention group. Additional details are provided by Harrell,³⁴ Basu et al.,³⁵ Hosmer and Lemeshow,³⁶ and Mihaylova et al.³⁷ Confidence intervals (CIs) were calculated with the use of bootstrapping to control for within-facility and within-patient correlations; see Goldstein et al.³⁹

† Favorable status at 30 days indicates that, at 30 days after an alert (in the target population) or at 30 days after admission (in the nontarget population) the patient was alive, was not in the hospital, and had not been readmitted at any time.

one might ensure compliance without alert fatigue. The program is amortized across 21 hospitals, which permits economies of scale, such as the use of nurses working remotely who attempt to mitigate alert fatigue as well as monitor compliance with clinical rescue and palliative care workflows. This approach may not be feasible for many hospitals.

In this study, we found that in conjunction with careful implementation, the use of automated predictive models was associated with lower hospital mortality, a lower incidence of ICU admission, and a shorter length of stay in the hospital.

Supported by grants (2663 and 2663.01) from the Gordon and Betty Moore Foundation, by the Sidney Garfield Memorial Fund, by a grant (1R01HS018480-01) from the Agency for Healthcare Research and Quality, by the Permanente Medical Group, by Kaiser Foundation Hospitals, and by grants (K23GM112018 and R35GM128672, to Dr. Liu) from the National Institutes of Health.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Barbara Crawford, Michelle Caghey, Philip Madvig, Stephen Parodi, Brian Hoberman, Robin Betts, and Marybeth Sharpe for administrative assistance and encouragement; Tracy Lieu and Catherine Lee for reviewing an earlier version of the manuscript; Ruth Ann Bertsch and Laura Myers for proofreading the revised manuscript; Kathleen Daly for formatting an earlier version of the manuscript; and the many Kaiser Permanente clinicians and staff who have contributed to the success of the Advance Alert Monitor program.

REFERENCES

1. Bapojé 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;6:68-72.
2. 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;6:74-80.
3. Delgado MK, Liu V, Pines JM, Kipnis P, Gardner MN, Escobar GJ. Risk factors for unplanned transfer to intensive care within 24 hours of admission from the emergency department in an integrated health-care system. *J Hosp Med* 2013;8:13-9.
4. Liu V, Kipnis P, Rizk NW, Escobar GJ. Adverse outcomes associated with delayed intensive care unit transfers in an integrated healthcare system. *J Hosp Med* 2012;7:224-30.
5. Churpek MM, Wendlandt B, Zadavec FJ, Adhikari R, Winslow C, Edelson DP. Association between intensive care unit transfer delay and hospital mortality: a multicenter investigation. *J Hosp Med* 2016;11:757-62.
6. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* 2013;84:465-70.
7. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. *Crit Care* 2017;21:52.
8. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. *Crit Care Med* 2019;47:49-55.
9. Evans RS, Kuttler KG, Simpson KJ, et al. Automated detection of physiologic deterioration in hospitalized patients. *J Am Med Inform Assoc* 2015;22:350-60.
10. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. 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. *J Hosp Med* 2012;7:388-95.
11. Rothman MJ, Rothman SI, Beals J IV. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. *J Biomed Inform* 2013;46:837-48.
12. Churpek MM, Yuen TC, Winslow C, et al. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med* 2014;190:649-55.
13. Kollef MH, Chen Y, Heard K, et al. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *J Hosp Med* 2014;9:424-9.
14. Kipnis P, Turk BJ, Wulf DA, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform* 2016;64:10-9.
15. Escobar GJ, Dellinger RP. Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016;11:Suppl 1:S5-S10.
16. Escobar GJ, Turk BJ, Ragins A, et al. Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. *J Hosp Med* 2016;11:Suppl 1:S18-S24.
17. Dummett BA, Adams C, Scruth E, Liu V, Guo M, Escobar GJ. Incorporating an early detection system into routine clinical practice in two community hospitals. *J Hosp Med* 2016;11:Suppl 1:S25-S31.
18. Granich R, Sutton Z, Kim YS, et al. Early detection of critical illness outside the intensive care unit: clarifying treatment plans and honoring goals of care using a supportive care team. *J Hosp Med* 2016;11:Suppl 1:S40-S47.
19. Escobar GJ, Liu V, Kim YS, et al. Early detection of impending deterioration outside the ICU: a difference-in-differences (DiD) study. *Am J Respir Crit Care Med* 2016;93:A7614. abstract.
20. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care* 2008;46:232-9.
21. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care* 2013;51:446-53.
22. Escobar GJ, Ragins A, Scheirer P, Liu V, Robles J, Kipnis P. Nonelective rehospitalizations and postdischarge mortality: predictive models suitable for use in real time. *Med Care* 2015;53:916-23.
23. Escobar GJ, Plimier C, Greene JD, Liu V, Kipnis P. Multiyear rehospitalization rates and hospital outcomes in an integrated health care system. *JAMA Netw Open* 2019;2(12):e1916769.
24. Paulson SS, Dummett BA, Green J, Scruth E, Reyes V, Escobar GJ. What do we do after the pilot is done? Implementation of a hospital early warning system at scale. *Jr Comm J Qual Patient Saf* 2020;46:207-16.
25. Volchenboum SL, Mayampurath A, Göksu-Gürsoy G, Edelson DP, Howell MD, Churpek MM. Association between in-hospital critical illness events and outcomes in patients on the same ward. *JAMA* 2016;316:2674-5.
26. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613-9.
27. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007;28:182-91.
28. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer-Verlag, 2009.
29. Harrell F Jr. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag, 2010.
30. Davey C, Hargreaves J, Thompson JA, et al. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. *Trials* 2015;16:358.
31. McNutt L-A, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;157:940-3.
32. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. 2nd ed. New York: John Wiley, 2011.
33. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med* 2017;36:4391-400.
34. Harrell FE Jr. Semiparametric modeling of health care cost and resource utilization. Presented at the 24th Annual Midwest Biopharmaceutical Statistics Workshop, Muncie, IN, May 21-23 2001.
35. Basu A, Manning WG, Mullahy J. Comparing alternative models: log vs Cox proportional hazard? *Health Econ* 2004;13:749-65.
36. Hosmer DW, Lemeshow S. Applied survival analysis: regression modelling of time to event data. New York: Wiley, 2008.
37. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ* 2011;20:897-916.
38. Collett D. Modelling survival data in medical research. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press, 2014.
39. Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. *Stat Med* 2002;21:3291-315.
40. Chen J, Ou L, Flabouris A, Hillman K, Bellomo R, Parr M. Impact of a standardized rapid response system on outcomes in a large healthcare jurisdiction. *Resuscitation* 2016;107:47-56.
41. Priestley G, Watson W, Rashidian A, et al. Introducing Critical Care Outreach: a ward-randomised trial of phased introduction in a general hospital. *Intensive Care Med* 2004;30:1398-404.
42. Horwitz LI, Kuznetsova M, Jones SA. Creating a learning health system through rapid-cycle, randomized testing. *N Engl J Med* 2019;381:1175-9.
43. Linnen DT, Escobar GJ, Hu X, Scruth E, Liu V, Stephens C. Statistical modeling and aggregate-weighted scoring systems in prediction of mortality and ICU transfer: a systematic review. *J Hosp Med* 2019;14:161-9.
44. Morgan DJ, Bame B, Zimand P, et al. Assessment of machine learning vs standard prediction rules for predicting hospital readmissions. *JAMA Netw Open* 2019;2(3):e190348.

Copyright © 2020 Massachusetts Medical Society.