

# Derivation, Validation, Sustained Performance, and Clinical Impact of an Electronic Medical Record–Based Perioperative Delirium Risk Stratification Tool

Elizabeth L. Whitlock, MD, MSc,\* Matthias R. Braehler, MD, PhD,\* Jennifer A. Kaplan, MD,† Emily Finlayson, MD,‡ Stephanie E. Rogers, MS, MPH, MD,‡ Vanja Douglas, MD,§ and Anne L. Donovan, MD||

**BACKGROUND:** Postoperative delirium is an important problem for surgical inpatients and was the target of a multidisciplinary quality improvement project at our institution. We developed and tested a semiautomated delirium risk stratification instrument, Age, WORLD backwards, Orientation, iLlness severity, Surgery-specific risk (AWOL-S), in 3 independent cohorts from our tertiary care hospital and describe its performance characteristics and impact on clinical care.

**METHODS:** The risk stratification instrument was derived with elective surgical patients who were admitted at least overnight and received at least 1 postoperative delirium screen (Nursing Delirium Screening Scale [NuDESC] or Confusion Assessment Method for the Intensive Care Unit [CAM-ICU]) and preoperative cognitive screening tests (orientation to place and ability to spell WORLD backward). Using data pragmatically collected between December 7, 2016, and June 15, 2017, we derived a logistic regression model predicting probability of delirium in the first 7 postoperative hospital days. A priori predictors included age, cognitive screening, illness severity or American Society of Anesthesiologists physical status, and surgical delirium risk. We applied model odds ratios to 2 subsequent cohorts (“validation” and “sustained performance”) and assessed performance using area under the receiver operator characteristic curves (AUC-ROC). A post hoc sensitivity analysis assessed performance in emergency and preadmitted patients. Finally, we retrospectively evaluated the use of benzodiazepines and anticholinergic medications in patients who screened at high risk for delirium.

**RESULTS:** The logistic regression model used to derive odds ratios for the risk prediction tool included 2091 patients. Model AUC-ROC was 0.71 (0.67–0.75), compared with 0.65 (0.58–0.72) in the validation (n = 908) and 0.75 (0.71–0.78) in the sustained performance (n = 3168) cohorts. Sensitivity was approximately 75% in the derivation and sustained performance cohorts; specificity was approximately 59%. The AUC-ROC for emergency and preadmitted patients was 0.71 (0.67–0.75; n = 1301). After AWOL-S was implemented clinically, patients at high risk for delirium (n = 3630) had 21% (3%–36%) lower relative risk of receiving an anticholinergic medication perioperatively after controlling for secular trends.

**CONCLUSIONS:** The AWOL-S delirium risk stratification tool has moderate accuracy for delirium prediction in a cohort of elective surgical patients, and performance is largely unchanged in emergent/preadmitted surgical patients. Using AWOL-S risk stratification as a part of a multidisciplinary delirium reduction intervention was associated with significantly lower rates of perioperative anticholinergic but not benzodiazepine, medications in those at high risk for delirium. AWOL-S offers a feasible starting point for electronic medical record–based postoperative delirium risk stratification and may serve as a useful paradigm for other institutions. (Anesth Analg 2020;131:00–00)

From the \*Department of Anesthesia and Perioperative Care, †Department of Surgery, ‡Division of Geriatrics, Department of Medicine, §Department of Neurology, and ||Division of Critical Care Medicine, Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, California.

Accepted for publication June 11, 2020.

**Funding:** Support was provided from institutional and/or departmental sources. A.L.D. has received research support from the Network for the Investigation of Delirium: Unifying Scientists (Subcontract 91511 of R24AG054259). M.R.B. has received funding support from the University of California, San Francisco Clinical and Translational Science Institute Pilot Awards Program. E.F. has received research support from the National Institute on Aging (NIA R01 AG044425, NIA P30 AG04428, and NIA R21AG054208).

Copyright © 2020 International Anesthesia Research Society  
DOI: 10.1213/ANE.0000000000005085

S.E.R. has received research support from the Centers for Disease Control and Prevention (CDC-OPOIIDS-2017-001 and CDC-STEADI-2016-001). V.D. has received funding support from the Sara & Evan Williams Foundation Endowed Neurohospitalist Chair. E.L.W. has received research support from the National Institutes of Health (R03AG059822, P30AG044281, and KL2TR001879) and the Foundation for Anesthesia Education and Research.

**Conflicts of Interest:** See Disclosures at the end of the article.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal’s website ([www.anesthesia-analgesia.org](http://www.anesthesia-analgesia.org)).

Reprints will not be available from the authors.

Address correspondence to Elizabeth L. Whitlock, MD, MSc, Department of Anesthesia and Perioperative Care, University of California, San Francisco, 500 Parnassus Ave, Campus Box 0648, San Francisco, CA 94143. Address e-mail to [elizabeth.whitlock@ucsf.edu](mailto:elizabeth.whitlock@ucsf.edu).

**KEY POINTS**

- **Question:** How can we accomplish electronic medical record–based postoperative delirium risk stratification to target resources toward high-risk patients?
- **Findings:** The Age, WORLD backwards, Orientation, iLlness severity, Surgery-specific risk (AWOL-S) semiautomated postoperative risk stratification tool has moderate, stable performance in a single-center cohort of surgical patients, and implementation was associated with a small decrease in anticholinergic, but not benzodiazepine, medication use in high-risk patients.
- **Meaning:** Although AWOL-S was developed in a single center and has its own limitations, it offers a useful paradigm for broad-based delirium stratification, which may impact clinical care.

**GLOSSARY**

**ASAPS** = American Society of Anesthesiologists physical status; **AUC-ROC** = area under the receiver operator characteristic curve; **AWOL** = Age, WORLD backwards, Orientation, iLlness severity delirium risk stratification tool; **AWOL-S** = Age, WORLD backwards, Orientation, iLlness severity, Surgery-specific risk; **CAM-ICU** = Confusion Assessment Method for the ICU; **CI** = confidence interval; **EMR** = electronic medical record; **ICU** = intensive care unit; **IQR** = interquartile range; **IRB** = institutional review board; **NuDESC** = Nursing Delirium Screening Scale; **PACU** = postanesthesia care unit; **UCSF** = University of California, San Francisco

Postoperative delirium is a common complication of major surgery and is associated with injurious falls, prolonged hospitalization, institutional discharge, and death.<sup>1</sup> Evidence-based methods of preventing postoperative delirium exist, although effective preventative methods such as the Hospital Elder Life Program<sup>2</sup> tend to be resource intensive, limiting their broad application across an unselected population. Many risk stratification tools have been published in the perioperative setting,<sup>3–7</sup> but performance of published instruments is widely variable, and internal validation studies are rarely undertaken, making it difficult to confirm validity even in a single-center population. Furthermore, while predicting postoperative delirium is an important goal, implementation of a risk prediction tool in clinical practice offers a major opportunity for quality improvement,<sup>1</sup> ideally including a mechanism to make the risk stratification results available to all care providers via the electronic medical record (EMR) and facilitating deployment of institutional resources to prevent delirium in those at high risk.

The University of California, San Francisco (UCSF) Delirium Reduction Campaign was a health system-wide initiative to broadly apply the Age, WORLD backward, Orientation, iLlness severity<sup>8</sup> (AWOL) delirium risk stratification tool to medical inpatients. To extend this initiative to elective surgical patients, we adapted the AWOL tool, now termed Age, WORLD backwards, Orientation, iLlness severity, Surgery-specific risk (AWOL-S), and describe its performance in 3 distinct epochs: the derivation period, a validation period, and finally, a cohort of patients following broad roll-out of postoperative delirium prevention and diagnosis protocols to assess sustained performance despite institutional practice changes.

Because emergent surgical patients and those preadmitted for optimization before planned surgery were not included in the derivation of the tool, we also describe the incidental performance of the tool in these important high-risk patients. Finally, in a post hoc analysis, we examine trends of benzodiazepine and anticholinergic use in patients screened at high risk for delirium to assess whether the use of these medications decreased.

**METHODS****Setting and Ethical Approval**

UCSF is an academic medical center consisting of 2 inpatient hospitals and 706 inpatient beds within the city of San Francisco. This study was performed as a quality improvement initiative at UCSF Medical Center, and the creation and testing of the AWOL-S tool were deemed to be not human subject research as defined by the UCSF Committee on Human Research, as they occurred entirely under the auspices of a quality improvement activity (institutional review board [IRB] #16-21073). The Committee on Human Research also approved the research into the performance characteristics and clinical impact of this tool and waived the requirement for written informed consent from subjects (IRB #19-27578).

**Local Context**

At the time we began this study, the UCSF Delirium Reduction Campaign—an institutionally funded initiative focusing on the prevention and identification of delirium at a hospital system-wide level—had begun risk stratifying all patients admitted to a rolling schedule of hospital floors using the AWOL delirium risk stratification tool<sup>8</sup> and testing at least once per nursing shift for the presence of delirium using

the Nursing Delirium Screening Scale (NuDESC)<sup>9</sup> or Confusion Assessment Method for the Intensive Care Unit (CAM-ICU),<sup>10</sup> which have similar sensitivity and specificity.<sup>11</sup> The training program for the approximately 1000 acute care and 500 intensive care nurses at UCSF Health included 30 minutes of formal training sessions, one-on-one education and daily huddles for the first week of clinical delirium screening, and weekly feedback emails for the first year of screening. Further details about the perioperative Delirium Reduction Campaign efforts are detailed elsewhere.<sup>12</sup>

The AWOL risk stratification tool, was developed at this institution and validated in medically hospitalized patients (ie, nonsurgical) with acceptable performance.<sup>8</sup> We felt that the greatest threat to face validity of the AWOL tool in the perioperative population was the well-demonstrated fact that certain procedures are more likely to precipitate postoperative delirium.<sup>13</sup> Thus, we provisionally planned to test predictive power of the AWOL components, add a term for the postoperative delirium risk associated with the surgical procedure itself, and consider a perioperative-specific adaptation of illness severity (ie, American Society of Anesthesiologists physical status [ASAPS]) in our risk stratification tool, while rederiving the risk estimates conferred by each characteristic to best predict delirium risk in our perioperative population.

**Patients**

All patients 18 years of age and older admitted to UCSF Medical Center for a surgical procedure requiring at least 1 night inpatient stay were considered. Patients undergoing emergency surgery or those who were hospitalized on the day before their surgery were excluded. Because preadmitted patients presenting for surgery (elective or emergent) may have already undergone risk stratification using the AWOL tool on admission, we did not include these patients in our derivation data. Non-English-speaking patients were largely excluded in the derivation cohort because of the evolving ability to perform AWOL and NuDESC assessment in these patients; during the derivation period, adaptations of the WORLD assessment for

non-English-speaking patients were being trialed at a hospital level independent of our perioperative work.

**Timing**

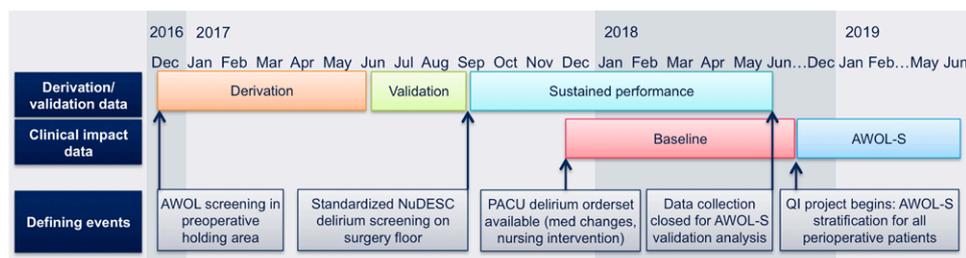
The derivation and the assessment of the risk score occurred concurrently with the gradual planned roll-out of the UCSF delirium pathway. On December 7, 2016, nurses in the preoperative area were instructed to perform the items of the AWOL screen (described further later) as a part of the standardized check-in process. On June 15, 2017, data collection was closed, and the resulting “derivation” dataset was used to estimate odds ratios for the model (described later). From June 16, 2017 to September 6, 2017, a “validation” dataset was collected. Then, on September 7, 2017, the surgical inpatient floors began routinized AWOL screening and mandatory NuDesc assessments, representing a change in practice for data collection and management of high delirium-risk patients; patients treated between September 7, 2017 and June 5, 2018, made up the “sustained performance” dataset. Finally, the AWOL-S screening tool was made available in the EMR, and preoperative nurses were trained in its use on June 6, 2018. Major events and timeline for the derivation/validation data are graphically described in Figure 1 (“Derivation/validation data”).

**AWOL Screen**

Briefly, the AWOL screen assigns points for characteristics associated with the development of delirium; a sum of points  $\geq 2$  stratifies the patient into a high-risk group. The characteristics are as follows: age  $\geq 80$  years, inability to spell WORLD backward, lack of orientation to place (country, city, hospital name, and hospital floor), and illness severity greater than or equal to “moderate,” as assessed by a registered nurse.<sup>8</sup>

**Variables**

The primary predictors were determined a priori and included age, nursing illness severity assigned at the day-of-surgery nursing assessment, and WORLD and orientation to place as components of the AWOL screen.



**Figure 1.** Graphical depiction of hospital QI initiative, data collection periods from this study (derivation/validation of AWOL-S and post hoc assessment of clinical impact), and significant events in the evolution of delirium-related screening and risk stratification. AWOL indicates age, WORLD backward, orientation, illness severity delirium risk stratification tool; AWOL-S, Age, WORLD backwards, Orientation, iLlness severity, Surgery-specific risk; NuDESC, Nursing Delirium Screening Scale; PACU, postanesthesia care unit; QI, quality improvement.

We also evaluated the predictive power of ASAPS, as assigned in the routine provision of anesthesia care, as a potential alternative to nursing illness severity. ASAPS was assessed by a resident physician training in anesthesiology, a physician anesthesiologist, or a certified registered nurse anesthetist according to guidance from the American Society of Anesthesiologists.<sup>14</sup>

A priori, we stratified all surgical procedures into low, moderate, and high risk based on the described potential to precipitate postoperative delirium using data from the National Surgical Quality Improvement Program<sup>13</sup> and intrinsic UCSF rates of postoperative delirium by surgical booking code (ie, an iterative process by which we identified unusually deliriogenic procedures in the derivation dataset and ensured that they were coded appropriately). A high delirium–risk procedure was generally thought of as one with a >10% risk of postoperative delirium; moderate risk, 5%–10%; and low risk, <5%, with some flexibility according to clinical judgment in acknowledgment that the literature rates may differ from institutional rates and experience and depending on the patient composition of the study in question. For example, of general surgery cases, appendectomy and cholecystectomy have low literature risk of postoperative delirium and hence were designated as “low intrinsic risk,” but colectomy, esophagectomy, and pancreatectomy are typically associated with >10% risk of delirium, so were given a “high intrinsic risk” designation. While ventral hernia repair is typically a low delirium–risk procedure, our institution performs an unusually high volume of complex hernia reconstructions, and, thus, ventral hernia repair is a moderate-risk operation in our population. The full list of delirium risk determinations for surgical cases is presented in Supplemental Digital Content, Appendix 1, <http://links.lww.com/AA/D158>.

The primary outcome was development of delirium in the first 7 days postoperatively, as assessed by NuDESC delirium screens that were performed at least once per shift of nursing care on acute care wards or CAM-ICU for patients in the ICU.

### **AWOL-S Performance in Exclusion Population**

Patients undergoing nonelective surgical cases (“emergency cases” as defined by the American Society of Anesthesiologists’ “E” modifier)<sup>14</sup> and those admitted for optimization before an elective/nonemergent surgery were deliberately excluded from rule derivation; however, this is an important subpopulation for delirium risk stratification. We also describe the performance of the AWOL-S tool in these patients for the period December 7, 2016 to June 5, 2018.

### **Clinical Implementation**

The AWOL-S logistic regression equation was programmed into a nursing flow sheet in our Epic EMR,

to be completed by the preoperative nurse during check-in on the day of surgery. The flow sheet automatically imported values for patient age, any ASAPS that had been entered, and the assigned procedure risk score; nursing staff entered the results of the WORLD and orientation screening. If an ASAPS had not yet been entered, the nursing staff rated illness severity on a 5-point scale, as for the original AWOL. Once the flow sheet rows were completed, the EMR automatically calculated and displayed the absolute predicted probability of developing delirium (ie, output of logistic regression equation derived earlier). For clinical utility, a dichotomous cutoff differentiating between “high” and “low” delirium risk was set at a predicted probability of 5%. Those with a predicted probability of delirium >5% (ie, those who were stratified as “high” risk for delirium during this perioperative encounter) received an orange cue (“flag”) on their EMR chart, which was visible to care providers from anesthesiology, surgery, and nursing who were using a perioperative Epic context. Choices of medications administered intraoperatively and in the postanesthesia care unit (PACU) could, therefore, explicitly be impacted by the AWOL-S risk stratification flag (high or low).

On July 1, 2018, clinical implementation of delirium best practice interventions, based on AWOL-S risk stratification results, was made the subject of a perioperative quality improvement initiative. Details of clinical implementation are described in Supplemental Digital Content, Appendix 2, <http://links.lww.com/AA/D158>, but briefly, anesthesiology resident physicians, nurse anesthetists, and attending physicians were instructed to avoid deliriogenic medications and prescribe nonpharmacologic delirium prevention precautions for all patients assessed by AWOL-S to be at high risk of delirium, including those not anticipating an overnight hospital stay, in the PACU. A PACU delirium order set facilitating these measures was implemented in December 2017 (Figure 1, “Significant events”), but its use was not emphasized until the start of the quality improvement period. The content of the order set is described elsewhere,<sup>12</sup> but briefly, nonpharmacologic delirium prevention measures like frequent reorientation were offered as a nursing order, and haloperidol replaced prochlorperazine and metoclopramide as rescue antiemetic selections. It was recommended, though not mandatory, that the surgical team prescribe similar delirium prevention measures to be implemented on the inpatient floor after PACU discharge.

### **Post Hoc Anticholinergic Medication and Benzodiazepine Use Analysis**

We evaluated the association between an AWOL-S positive screen and the perioperative (ie, “anesthesia

start” to PACU discharge) use of medications typically avoided in patients at high delirium risk. We compared the rates of benzodiazepines (midazolam and lorazepam) and anticholinergic medications (diphenhydramine, meperidine, prochlorperazine, promethazine, and scopolamine) before (December 1, 2017 to June 30, 2018) and after (July 1, 2018 to June 30, 2019) the quality improvement initiative focusing on clinical implementation of delirium best practices based on the AWOL-S score (Figure 1, “Clinical impact data”), using rates in patients who did not screen high risk to control for secular trends.

### Statistical Analysis

All analyses were performed in Stata version 14.2 (StataCorp LLC, College Station, TX), and a threshold of  $P < .05$  was considered to indicate statistical significance. The threshold for statistical significance was not adjusted for multiple comparisons because the impact of the intervention on delirigenic medication prescribing was assumed to be strongly correlated among different medications, and the post hoc exploratory nature of the analysis means findings should be interpreted cautiously.

A priori, we planned to include age, illness severity (assessed on a 1–5 scale by nursing staff<sup>8,15</sup> or by ASAPS), cognitive ability (as assessed by WORLD backward and orientation), and an indicator of procedure-specific delirium risk. We evaluated performance of a logistic regression model to predict delirium incorporating different versions of an age coefficient (eg, as a continuous variable, as an ordinal variable with cutoffs at 65 and 80 years, and as a dichotomous variable with cutoff at 70 years) and illness severity. Items were evaluated in the model according to clinical judgment, ease of interpretability for a clinical audience, and impact on model fit and prediction accuracy. Model fit was assessed using the area under the receiver operator characteristic curve (AUC-ROC). All tested versions of coefficients were evaluated to ensure model assumptions were met (eg, linearity in the logit for continuous predictors). To address nonindependence of multiple surgical cases in a single patient, variances were adjusted for clustering by patient identification number.

The final model, termed AWOL-S, was selected on the basis of model fit and alignment with AWOL components. Model odds ratios were then applied to the validation and sustained performance cohorts. Calibration of predicted probabilities in the cohorts not used to derive the AWOL-S model (ie, the validation and sustained performance cohorts of elective patients) was assessed with predicted versus observed probability plots using the user-defined Stata package “pmmcalplot.”<sup>15</sup>

Univariate statistics were performed according to the variable type and the distribution to compare key characteristics among the derivation, validation, and sustained performance cohorts; the exclusion cohort was not included in the statistical comparison because of obvious and expected differences in the distribution of risk profiles.

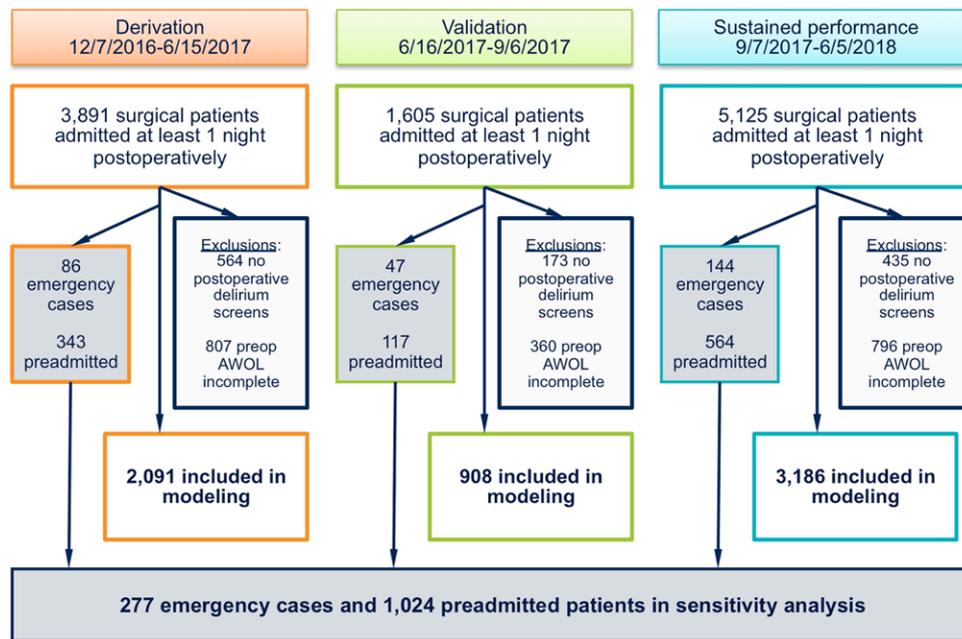
Rates of benzodiazepine and anticholinergic medications during perioperative care—that is, start of anesthesia team care until discharge from the PACU or, for patients who went directly to ICU, the end of anesthesia team care—were compared before and after the clinical implementation of AWOL-S using binomial regression with an interaction term (ie, a term in the model to indicate whether high-risk patients were treated differentially after AWOL-S implementation, after controlling for secular trends in medication use in low-risk patients). Variances were adjusted for clustering by patient since multiple surgical encounters per patient could be considered. All elective and emergent surgical procedures were included for this analysis.

### RESULTS

There were 10,621 surgical cases performed on patients who stayed at least 1 night postoperatively between December 7, 2016 and June 5, 2018. Of these, 6105 were elective procedures for patients who received preoperative risk stratification and at least 1 postoperative delirium assessment. An additional 1301 cases were performed in preadmitted patients or on an emergent basis who also received preoperative and postoperative screening. (Figure 2; Table 1) There were modest but statistically significant differences in the distribution of ASAPS, preoperative orientation, and procedure risk value among the 3 main cohorts. As expected, the exclusion cohort (emergent and preadmitted cases) markedly differed from the target population for the derivation, validation, and sustained performance cohorts.

Performance of the original AWOL cutoff for elevated delirium risk was poor in the elective perioperative population (AUC-ROC, 0.56 [0.52–0.60] for the derivation cohort; Table 2). Odds ratios for the AWOL-S equation were obtained via logistic regression performed on 2091 patients in the derivation cohort. Rounded odds ratios and baseline risk (intercept) for the prediction algorithm are presented in Table 3.

Odds ratios for each predictor variable from the derivation cohort were applied to the baseline risk for validation and sustained performance cohorts, and performance of AWOL-S (predicted probability model and performance of the 5% predicted probability cutoff used to dichotomize “high” versus “low” delirium risk) across the 3 time epochs is



**Figure 2.** Diagram of included and excluded patients in the derivation, validation, and sustained performance cohorts. AWOL indicates age, WORLD backward, orientation, illness severity delirium risk stratification tool.

**Table 1. Description of the Derivation, Validation, and Implementation of Elective Surgical Cohorts, and Differential Performance of the Models for Delirium Risk Stratification**

Description of Cohort	Derivation	Validation	Sustained Performance	Exclusion	P Value
Number of encounters	2091	908	3186	1301	
Age (y)	58 ± 16	59 ± 16	59 ± 15	60 ± 16	.64
Unable to spell WORLD backward (%)	16.8	16.7	15.6	29.7	.46
Not oriented to place (%)	6.7	6.7	5.0	10.4	<b>.019</b>
ASA physical status (%)					
I	8.3	5.3	6.0	2.6	<b>.004</b>
II	53.8	52.4	52.7	32.6	
III	35.5	39.3	39.0	53.4	
IV	2.5	3.0	2.4	11.4	
V	0	0	0	0.1	
Surgery-specific delirium risk (%)					
Low	16.7	19.9	18.6	18.5	<b>.001</b>
Moderate	63.8	60.4	65.3	55.9	
High	19.5	19.7	16.0	25.6	
Number of delirium assessments	5 [IQR 2–9]	5 [IQR 2–9]	6 [IQR 3–11]	10 IQR [4–14]	<b>&lt;.001</b>
Delirium in first 7 postop days (%)	5.5	5.7	5.6	15.1	.97
<b>Model Performance</b>	<b>Derivation</b>	<b>Validation</b>	<b>Sustained Performance</b>	<b>Exclusion</b>	
AWOL ≥2 cutoff					
AUC	0.56 (0.52–0.60)	0.56 (0.50–0.61)	0.57 (0.54–0.60)	0.67 (0.64–0.71)	
Sensitivity (%)	22.6	21.2	22.0	49.2	
Specificity (%)	88.9	90.0	91.9	85.4	
AWOL-S					
Regression model AUC	0.71 (0.67–0.75)	0.65 (0.58–0.72)	0.75 (0.71–0.78)	0.71 (0.67–0.75)	
5% cutoff: sensitivity	75.7	65.4	75.3	80.2	
5% cutoff: specificity	58.9	59.2	59.6	44.8	

The “exclusion” cohort was not included in the univariate statistical comparison for differences in cohort description among the derivation, validation, and implementation cohorts. AWOL-S model performance is reported as both the full regression model (“regression model AUC”) and the clinically implemented cutoff of 5% (“5%”).

Bold text represents  $P < .05$ .

Abbreviations: ASA, American Society of Anesthesiologists; AUC, area under the receiver operator characteristic curve; AWOL, Age, WORLD backward, Orientation, iLlness severity delirium risk stratification tool; AWOL-S, Age, WORLD backwards, Orientation, iLlness severity, Surgery-specific risk; IQR, interquartile range.

presented in Table 1, with ROC curves for the predicted probabilities in Figure 3A. Performance was slightly worse in the validation cohort due to lower

sensitivity in that cohort; specificity of the 5% cutoff was stable at approximately 59%. Model calibration in the combined validation and sustained

**Table 2. Clinical Impact: Description of Cohort and Medication Administration**

	Baseline		AWOL-S		P
	Low Risk	High Risk	Low Risk	High Risk	
Number of surgical encounters	7368	2049	13,704	3630	
Age (y)	53 ± 16	67 ± 7.2	53 ± 316	68 ± 812	.85
Unable to spell WORLD backward	10%	35%	13%	38%	<b>&lt;.001</b>
Not oriented to place	2.5%	15%	1.7%	13%	<b>&lt;.001</b>
High illness severity (ASA III+)	29%	77%	29%	76%	.12
Surgery-specific delirium risk					
Low	74%	7.4%	74%	7.1%	.32
Moderate	23%	70%	23%	70%	
High	2.6%	23%	3.0%	23%	
Delirium in first 7 postop days	1.3%	10%	1.3%	11.4%	.31

Perioperative Medication Use	Baseline		AWOL-S		Differential RR	P
	Low Risk	High Risk	Low Risk	High Risk		
Any benzodiazepine	54%	37%	50%	33%	0.95 (0.88–1.03)	.24
Midazolam	54%	36%	50%	32%	0.95 (0.87–1.03)	.19
Lorazepam	1.8%	2.0%	1.7%	1.8%	0.98 (0.63–1.53)	.63
Any anticholinergic	12%	8.6%	12%	6.5%	<b>0.79 (0.64–0.97)</b>	<b>.02</b>
Diphenhydramine	1.7%	1.5%	2.2%	1.4%	0.74 (0.45–1.22)	.23
Meperidine	2.8%	2.0%	3.9%	1.9%	0.68 (0.45–1.03)	.07
Prochlorperazine	6.6%	4.5%	5.1%	3.1%	0.89 (0.67–1.19)	.44
Promethazine	0.1%	0.1%	0.06%	0.03%	0.52 (0.04–7.01)	.63
Scopolamine	2.1%	1.1%	1.7%	0.6%	0.72 (0.39–1.34)	.31

“Differential RR” is the relative risk of receiving a medication in the high-delirium-risk group after AWOL-S implementation, after controlling for secular trends using rates of medication use in the low-risk group.

Bold text represents  $P < .05$ .

Abbreviations: ASA, American Society of Anesthesiologists; AWOL-S, Age, WORLD backwards, Orientation, iLLness severity, Surgery-specific risk.

**Table 3. Rounded Odds Ratios for Components in the AWOL-S Prediction Equation**

Baseline Risk <sup>a</sup>	Odds Ratio
<b>0.28%</b>	
Age	
Per year >65	Odds ratio 1.02
Per year >65	Odds ratio 0.98
Unable to spell WORLD backward	Odds ratio 1.5
DisOriented to place	Odds ratio 1.7
iLLness severity	
ASA II	Odds ratio 4.3
ASA III or higher	Odds ratio 8.3
Surgical risk	
Moderate risk	Odds ratio 3.4
High risk	Odds ratio 4.6

Abbreviations: ASA, American Society of Anesthesiologists; AWOL-S, Age, WORLD backwards, Orientation, iLLness severity, Surgery-specific risk.

<sup>a</sup>Baseline risk refers to the predicted probability of delirium for a hypothetical patient at the reference value for all categories; ie, a 65-year-old patient who is able to spell WORLD backward, is oriented to place, and is ASA I with a low-risk planned surgery. The baseline risk is multiplied by the relevant odds ratios to obtain an individual’s predicted probability of delirium.

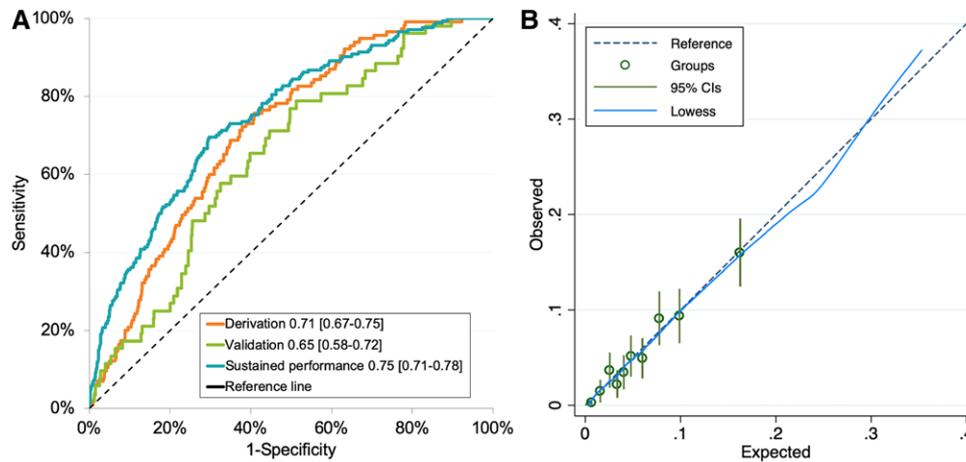
performance cohorts was excellent; the calibration slope was 1.032 (ideal, 1.0) and the intercept was -0.023 (ideal, 0; Figure 3B). For the entire cohort of 10,621 patients, overall AUC-ROC of the regression model was 0.72 (0.70–0.75). The 5% predicted probability threshold that differentiated “low-risk” from “high-risk” patients performed at a sensitivity and specificity in the entire cohort of 73.9% and 59.3%, respectively.

Performance of the AWOL-S tool in emergency and preadmitted patients specifically excluded from the derivation population was also moderate, with slightly higher sensitivity but lower specificity (Table 1).

Changes in medication administration after AWOL-S entered clinical use as a part of a multicomponent quality improvement program were evaluated in a total of 26,751 surgical encounters: 9417 before the clinical implementation of AWOL-S screening (baseline) and 17,334 after implementation. We calculated a differential odds ratio (interaction term; Table 2), which shows the additional change in medication use in high-risk patients after using in low-risk patients to adjust for secular trends. The point estimate for all medications studied was <1, suggesting an overall lower rate of deliriogenic medications in patients who screened high risk for delirium, although confidence intervals were typically wide (Table 2). There was a statistically significant decrease in anticholinergic use; patients at high risk for delirium had a 21% lower relative risk (95% confidence interval, 3–36;  $P = .02$ ) of receiving perioperative anticholinergics after AWOL-S screening was clinically implemented after adjusting for the rate of change in low-risk patients.

**DISCUSSION**

We demonstrate moderate performance of an EMR-based postoperative delirium screening tool that incorporates both automatically generated (eg, age) and physician- and nursing-based (ie, ASAPS or illness severity, and cognitive screening) components to calculate a summary probability of postoperative delirium and delirium risk stratification, which is visible to all perioperative care providers via the EMR. We have implemented this tool, termed AWOL-S, for risk stratification in our perioperative care area,



**Figure 3.** Discrimination and calibration of AWOL-S. A, ROC curves for the logistic regression model predicting postoperative delirium in the derivation, validation, and sustained performance cohorts. Area under the ROC curve is displayed. B, Calibration curve for the combined validation and sustained performance cohorts. AWOL-S indicates Age, WORLD backwards, Orientation, iLLness severity, Surgery-specific risk; CI, confidence interval; ROC, receiver operator characteristic.

aligning our practice with recent recommendations from the American Society for Enhanced Recovery and Perioperative Quality Initiative.<sup>1</sup> AWOL-S risk stratification was associated with a modest but statistically significant reduction in anticholinergic medication administration to high delirium-risk patients, after adjusting for secular trends. The AWOL-S tool represents an improvement over the existing AWOL tool for predicting postoperative delirium and, we feel, offers a useful paradigm for EMR-based delirium risk stratification in a broad population of surgical patients.

While the performance of the AWOL-S tool is moderate, it remained roughly stable across 3 additional, independent cohorts from the same target population. Importantly, it represented a major improvement from the AWOL tool, which has similar, acceptable performance in a medical population<sup>8,16</sup> but proved poor at predicting postoperative delirium. One notable exception to the generally moderate performance of broadly applied postoperative delirium risk stratification tools was published by Kim et al,<sup>6</sup> who demonstrated an AUC-ROC of 0.91 with their 7-factor model in a same-center validation cohort. However, their tool was tested only in general, vascular, and trauma surgeries, and some of the factors in their model, particularly C-reactive protein and ICU admission, may not be universally assessed or reliably known at the start of surgery. In the medical population, an EMR-based clinical decision support tool for delirium achieved a AUC-ROC of 0.76 when applied to older adults, but surgical patients were not the focus of this tool<sup>5</sup> or of a similar EMR-based tool applied to the US Veteran's Affairs medical population.<sup>17</sup> An EMR-based tool for predicting postoperative delirium in a semiautomated fashion was necessary to harmonize the perioperative departments with our institutional

Delirium Reduction Campaign goals, prompting this study.

The AWOL-S tool was derived in a cohort that excluded emergency cases and patients who were preadmitted before their surgery; thus, the performance might be expected to be poorer in those patients. We found moderate performance in a post hoc sensitivity analysis, which was not different from the performance in the target population of elective surgical patients. We had initially planned not to include patients undergoing nonelective surgery in the AWOL-S risk stratification quality improvement pathway, but after it became evident that the AWOL-S tool better predicted delirium than the unmodified AWOL tool, our practice is now to use AWOL-S risk stratification for every patient who enters the preanesthesia care unit, regardless of emergent/preadmitted status.<sup>12</sup>

Implementation of the AWOL-S tool was associated with an approximately 2% absolute risk reduction (from 8.6% to 6.5%) in the administration of anticholinergic medications to those at high delirium risk, translating to 73 high-risk patients who did not receive an anticholinergic medication during the year-long AWOL-S implementation period. During the study period, rates of benzodiazepine administration decreased in our population, regardless of delirium risk. While this study was not designed to prove causality, the ability to adjust for concurrent local trends in medication administration (ie, controlling for medication use rates in low-risk patients) makes it less likely that the finding was due to secular trends alone.

There are important limitations to this study. First, the incidence of delirium in the population of non-emergent surgical inpatients, which formed the target population, for this risk stratification tool was approximately 5% throughout the time period under study,

which likely represents undercounting. Hypoactive delirium is particularly vulnerable to undercounting. The NuDESC tool may be one of the more sensitive options to detect delirium in the postoperative setting,<sup>18</sup> although it, and the CAM-ICU, has elsewhere been criticized for low sensitivity.<sup>11</sup> However, screening all postoperative patients for delirium in an institution of this size using the gold standard (a face-to-face interview with a qualified clinician using the Diagnostic and Statistical Manual of Mental Disorders) was infeasible. Second, the AWOL and AWOL-S tools were not consistently adapted for use in non-English-speaking patients; accordingly, data from these patients are mostly unavailable, although adapting the tools has been a focus for refinement of our ongoing quality improvement initiative. Third, delirium was defined as any positive delirium screen in the first 7 postoperative days, which implies a heterogeneous underlying mixture of delirium severity and duration, which may be identified by this tool. Fourth, the clinical impact of the intervention was assessed using a post hoc, exploratory analysis that should be interpreted conservatively; there is no established threshold for clinical significance of anticholinergic medication use, and, further, we are unable to make any conclusions about the impact of AWOL-S risk stratification on delirium because of evolving delirium interventions independent of the perioperative setting. Finally, delirium prediction tools are notorious for their broad failure to hold external validity.<sup>3</sup> This study was designed as a quality improvement initiative, and, thus, external validity was not an explicit goal; if this tool is implemented in other settings, reweighting the odds ratios according to local needs should be strongly considered.

In summary, we describe the derivation, validation, and sustained performance of an automated EMR-based postoperative delirium prediction tool that we implemented as part of a quality improvement initiative focusing on delirium. The tool has moderate performance across 3 independent elective surgical cohorts and a nonelective cohort in our population. Patient stratification to a high delirium-risk group was associated with a reduction in perioperative use of delirium-precipitating medications, particularly anticholinergics, suggesting risk stratification impacted clinical care. AWOL-S may serve as a useful paradigm for other institutions interested in implementing broad perioperative delirium risk stratification. ■

#### ACKNOWLEDGMENTS

The authors thank Jon Spinner, Rachele Armstrong, Adam Jacobson, Amy Hephner, and Carly Deibler of UCSF Clinical Services and Anesthesia Information Technology, University of California, San Francisco, San

Francisco, CA, for their assistance in leveraging the electronic medical record to make this study possible.

#### DISCLOSURES

**Name:** Elizabeth L. Whitlock, MD, MSc.

**Contribution:** This author helped design the work; acquire, analyze, and interpret the data; draft the main manuscript; and approve the final manuscript.

**Conflicts of Interest:** None.

**Name:** Matthias R. Braehler, MD, PhD.

**Contribution:** This author helped design the work, acquire the data, and critically revised and approved the final manuscript.

**Conflicts of Interest:** None.

**Name:** Jennifer A. Kaplan, MD.

**Contribution:** This author helped acquire the data, offered critical revisions of the manuscript, and approved the final manuscript.

**Conflicts of Interest:** None.

**Name:** Emily Finlayson, MD.

**Contribution:** This author helped design the work, offered critical revisions of the manuscript, and approved the final manuscript.

**Conflicts of Interest:** E. Finlayson is a founding shareholder of Ooney, Inc.

**Name:** Stephanie E. Rogers, MS, MPH, MD.

**Contribution:** This author helped design the work, interpret the data, and critically revised and approved the final manuscript.

**Conflicts of Interest:** None.

**Name:** Vanja Douglas, MD.

**Contribution:** This author helped design the work, interpret the data, and critically revised and approved the final manuscript.

**Conflicts of Interest:** None.

**Name:** Anne L. Donovan, MD.

**Contribution:** This author helped design the work; acquire, analyze, and interpret data; draft the main manuscript and offer critical revisions; and approved the final manuscript.

**Conflicts of Interest:** None.

**This manuscript was handled by:** Robert Whittington, MD.

#### REFERENCES

- Hughes CG, Boncyk CS, Culley DJ, et al; Perioperative Quality Initiative (POQI) 6 Workgroup. American society for enhanced recovery and perioperative quality initiative joint consensus statement on postoperative delirium prevention. *Anesth Analg*. 2020;130:1572–1590.
- Hsieh TT, Yang T, Gartaganis SL, Yue J, Inouye SK. Hospital elder life program: systematic review and meta-analysis of effectiveness. *Am J Geriatr Psychiatry*. 2018;26:1015–1033.
- Jansen CJ, Absalom AR, de Bock GH, van Leeuwen BL, Izaks GJ. Performance and agreement of risk stratification instruments for postoperative delirium in persons aged 50 years or older. *PLoS One*. 2014;9:e113946.
- van Meenen LC, van Meenen DM, de Rooij SE, ter Riet G. Risk prediction models for postoperative delirium: a systematic review and meta-analysis. *J Am Geriatr Soc*. 2014;62:2383–2390.
- de Wit HA, Winkens B, Mestres Gonzalvo C, et al. The development of an automated ward independent delirium risk prediction model. *Int J Clin Pharm*. 2016;38:915–923.
- Kim MY, Park UJ, Kim HT, Cho WH. DELirium prediction based on hospital information (Delphi) in general surgery patients. *Medicine (Baltimore)*. 2016;95:e3072.
- Choi NY, Kim EH, Baek CH, Sohn I, Yeon S, Chung MK. Development of a nomogram for predicting the probability of postoperative delirium in patients undergoing free flap reconstruction for head and neck cancer. *Eur J Surg Oncol*. 2017;43:683–688.

8. Douglas VC, Hessler CS, Dhaliwal G, et al. The AWOL tool: derivation and validation of a delirium prediction rule. *J Hosp Med.* 2013;8:493–499.
9. Gaudreau JD, Gagnon P, Harel F, Tremblay A, Roy MA. Fast, systematic, and continuous delirium assessment in hospitalized patients: the nursing delirium screening scale. *J Pain Symptom Manage.* 2005;29:368–375.
10. Ely EW, Margolin R, Francis J, et al. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Crit Care Med.* 2001;29:1370–1379.
11. Neufeld KJ, Leoutsakos JS, Sieber FE, et al. Evaluation of two delirium screening tools for detecting post-operative delirium in the elderly. *Br J Anaesth.* 2013;111:612–618.
12. Donovan AL, Braehler MR, Robinowitz DL, et al. An implementation-effectiveness study of a perioperative delirium prevention initiative for older adults. *Anesth Analg.* 2020;131:1911–1922.
13. Berian JR, Zhou L, Russell MM, et al. Postoperative delirium as a target for surgical quality improvement. *Ann Surg.* 2018;268:93–99.
14. American Society of Anesthesiologists. ASA Physical Status Classification System, Standards and Guidelines. October 23, 2019. Available at: <https://www.asahq.org/standards-and-guidelines/asaphysical-status-classification-system>. Accessed April 16, 2020.
15. Ensor J, Kym IE. PMCALPLOT: Stata module to produce calibration plot of prediction model performance. *Stat Software Compon, Boston College Depart Economic.* 2018: S458486.
16. Brown EG, Josephson SA, Anderson N, Reid M, Lee M, Douglas VC. Predicting inpatient delirium: the AWOL delirium risk-stratification score in clinical practice. *Geriatr Nurs.* 2017;38:567–572.
17. Rudolph JL, Doherty K, Kelly B, Driver JA, Archambault E. Validation of a delirium risk assessment using electronic medical record information. *J Am Med Dir Assoc.* 2016;17: 244–248.
18. Radtke FM, Franck M, Schust S, et al. A comparison of three scores to screen for delirium on the surgical ward. *World J Surg.* 2010;34:487–494.