



ELSEVIER

Contents lists available at ScienceDirect

Air Medical Journal

journal homepage: <http://www.airmedicaljournal.com/>

Original Research

The Rapid Emergency Medicine Score: A Critical Appraisal of Its Measurement Properties and Applicability to the Air Retrieval Environment

Brodie Nolan, MD^{1,*}, Homer Tien, MD, MSc², Barbara Haas, MD, PhD²,
Refik Saskin, MSc³, Avery Nathens, MD, PhD²

¹ Department of Medicine, Sunnybrook Health Sciences Center, University of Toronto, Toronto, Ontario, Canada

² Department of Surgery, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada

³ Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada



A B S T R A C T

Objective: The Rapid Emergency Medicine Score (REMS) was designed to predict in-hospital mortality using variables that are available in the prehospital setting. The objective of this article is to critically appraise the development and summarize the evidence regarding the measurement properties (sensitivity, reliability and validity) of the REMS.

Methods: A literature search was performed identifying all studies describing the REMS. The original validation study was critically appraised for its development. All other studies that reported any measurement properties of the REMS were also appraised for evidence of calibration, reliability, and validity.

Results: In total, 26 studies reported on the measurement properties of the REMS. Overall, the REMS was developed with robust methodology and has good sensibility with adequate content and face validity. It is easy to understand and feasible to be calculated within minutes of patient assessment. The REMS has the necessary measurement properties to be both a predictive and evaluative clinical index to measure prehospital severity of illness; however, no studies have adequately addressed the intra or inter-rater reliability of the score.

Conclusions: There is evidence to support the use of the REMS as a predictive or evaluative instrument. In most studies, it performed as well or better than other illness severity scores in predicting mortality.

© 2019 Air Medical Journal Associates. Published by Elsevier Inc. All rights reserved.

The evaluation of helicopter emergency medical services (HEMS) clinical outcomes and response to treatment is fraught with challenges. Often, regional HEMS systems do not have access to in-hospital outcomes or treatments, making assessments of performance limited. Additionally, most HEMS organizations only measure time response intervals and out-of-hospital cardiac arrest survival.¹ There are several proposed performance measures for specific disease processes in pre-hospital care,¹ but those affect a small amount of patient transports and mostly evaluate compliance with standards rather than patient response. To better understand the effect of HEMS care on clinical outcomes, a tool that can both predict mortality and measure clinical change is needed. Furthermore, a reliable measure of severity of illness could be used to improve triage of patients who would benefit from launching an HEMS retrieval team.

A tool that predicts mortality needs to account for severity of illness. Severity of illness can be understood as the degree of physiologic derangement caused by an acute disease process.^{2–5} There are many factors that have been conceptualized as part of severity of illness including a patient's vital signs, Glasgow coma scale (GCS), age, laboratory investigations, and chronic health conditions.^{2,3} However many illness severity scores are either too disease specific or require data that are unavailable in the prehospital setting.⁵

There are various scoring systems that have been developed to predict mortality including the Acute Physiology and Chronic Health Evaluation (APACHE II),³ the Rapid Acute Physiology Score (RAPS),⁵ and the Rapid Emergency Medicine Score (REMS).⁴ Of those, REMS shows the most promise as being a useful prehospital measure of severity of illness and evaluative tool for monitoring clinical change. The objective of this article is to critically appraise the development and summarize the evidence regarding the measurement properties (sensitivity, reliability, and validity)

* Address for correspondence: Brodie Nolan, MD, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Toronto, Ontario, Canada M4N 3M5.

E-mail address: brodie.nolan@mail.utoronto.ca (B. Nolan).

of the REMS. We also discuss potential applications of the REMS in the HEMS environment.

Methods

Literature Search

We aimed to identify articles describing the development of the REMS and all randomized/quasi-randomized trials, cohort, case-control, cross-sectional, and case series studies that evaluated the REMS as a measure of illness severity. A literature search was performed on March 7, 2018, using the PubMed and MEDLINE databases. Using the PubMed database, a search for the term “rapid emergency medicine score” yielded 35 articles. An additional search on MEDLINE using the terms “severity of illness scores” AND “transportation of patients” yielded 673 articles. A manual search was also performed of the bibliography of the included primary studies. All titles and abstracts were screened for relevance, and full texts were reviewed if studies met our criteria. Abstracts, dissertations/theses, commentaries, and letters were excluded. Studies that did not report the measurement properties of the REMS were also excluded. Descriptions of the final 26 studies that were included in our appraisal are provided in [Table 1](#).

Development of the REMS

Previous Severity Illness Scores

The APACHE II uses a point score based on 12 routine physiological variables together with age and previous health status for use in intensive care unit (ICU) patients. It has been validated to predict mortality in both medical and surgical intensive care patients.^{4,6} However, the APACHE II score includes several blood chemistry variables and is therefore not suitable for rapid scoring in the prehospital or emergency department environments. The RAPS is an abbreviated version of the APACHE II, which includes the physiologic variables available in the prehospital setting.⁵ It has an identical point value system as the APACHE II, except for GCS, which was adjusted to keep its proportional contribution similar between the 2 scores.⁵ The RAPS was initially evaluated as a prehospital severity scoring system in a population of helicopter-transported patients.⁵ Its obvious advantage over the APACHE II is the availability of all its components in a prehospital environment.

Item Generation

Item generation refers to bringing together a set of variables that are likely to be included in the instrument. Items should be representative of important domains, have a statistical association with the outcome, be sufficient to capture any important clinical change, and have universal applicability. The variables included in the APACHE II system are heart rate, respiratory rate, mean arterial pressure (MAP), GCS, body temperature, oxygenation of arterial blood, arterial pH, serum sodium, serum potassium, serum creatinine, hematocrit, white blood count, and GCS.⁴ The maximal APACHE II score is 71. The RAPS was developed by taking variables of APACHE II that are easily obtained in the prehospital setting (ie, heart rate, respiratory rate, MAP, and GCS).⁵ The scoring range for each variable was 0 to 4 with a maximal score of 16. The REMS was created with the intent on improving the predictive power of RAPS. The authors included the initial 4 variables of RAPS and added age (as weighted in the APACHE II from 0-6 points) and oxygen saturation (weighted on a score 0-4) ([Appendix 1](#)).⁴ Overall item generation seems adequate by including all variables that are both available in the prehospital setting and have previously been predictive of mortality.

Derivation and Internal Validation

A multivariable logistic regression analysis was performed. The REMS was predictive of in-hospital mortality, resulting in an odds

ratio of 1.58 for each point increase (95% confidence interval, 1.48-1.70).⁴ Internal validation was done using a 50/50 split-sample technique, resulting in similar odds ratios and 95% confidence intervals for the 2 samples.

Weights and Scoring

Each variable has unique ranges that determine the patient's score for that variable, 0 to 4 for heart rate, respiratory rate, MAP, oxygen saturation, and GCS and 0 to 6 for age. The REMS is calculated by taking the sum of scores for each variable. Scoring is straightforward and easily calculated in real time. The result is a continuous REMS score of integer values ranging from 0 to 26 with a higher score having higher odds of death. One benefit of these gradations is allowing for greater or lower physiologic derangement to be tracked over time to evaluate change to a patient's clinical status (ie, deterioration or improvement).

Sensibility Assessment

Purpose, Population, and Setting

The authors clearly stated their purpose was to determine if a modified illness severity scoring system (REMS) could be created using variables only available in the prehospital setting and have the scoring system perform as well as the APACHE II in predicting in-hospital mortality for nonsurgical patients presenting to the emergency department.⁴ Its intended purpose was predictive according to the framework described by Kirschner and Guyatt.⁸ Although the study setting was not a prehospital environment, patients were recruited directly from the emergency department and would have likely arrived by emergency medical services (EMS). Furthermore, by focusing on a population of patients admitted to the hospital, it was more feasible to link the REMS to mortality, an outcome that is not commonly available in prehospital databases. Additionally, further studies performed have used REMS for an evaluative purpose ([Table 1](#)).

Content Validity

Content validity is the idea that all the dimensions of a particular concept are adequately captured in the instrument. The REMS score adequately captures physiologic derangement and level of consciousness; however, it does not assess chronic health conditions. Although this may be a limitation to the REMS, this information is not often available in the prehospital or emergency department environment, especially in acutely critically unwell patients. Therefore, the omission of data regarding chronic health conditions from the REMS is appropriate given the intended population for its use. As discussed previously, item generation was statistically rigorous and clinically sound. Each variable included has been associated with increased mortality. Weighing of each variable is appropriate given the degree to which each is likely to contribute to mortality risk.

Face validity

Face validity refers to whether the variables of a clinical index make sense. As mentioned earlier, variable selection was relevant, appropriate, and applicable to severity of illness. Each variable of REMS is independent of the other and phrased clearly. The responses for each variable are discrete integers, and score cutoffs are well-defined. Determination of the final score by the sum of individual variable scores is appropriate.

Feasibility

The measured variables for REMS are easy to understand and routinely captured on all critically unwell patients in a prehospital and emergency department population. Scoring of individual variables according to cutoff values and final calculation of the REMS is simple to perform. Definitions of variables are intuitive. Instructions are provided on how to collect all needed information and how missing

Table 1
The Main Characteristics of Studies That Provided Evidence of Reliability or Validity for the Rapid Emergency Medicine Score (REMS)

First Author	Study Population	Clinimetric Property Reported	Study Design and Setting	Calibration and Reliability Evidence	Validity Evidence
Olsson ⁴	1,027 nonsurgical patients	Predictive	Prospective, single center First report of the scale, derivation, and validation Comparison predictive power of the APACHE II, RAPS, and REMS on in-hospital mortality	Adequate calibration: H-L χ^2 statistic = 9.3 ($P = .23$)	Predictive criterion validity: REMS (ROC AUC = 0.910 \pm 0.015) REMS predictive of mortality OR = 1.58 (95% CI, 1.58-1.70) for each 1-point increase Internal validity: Similar results with split-sample technique Concurrent construct validity: REMS similar predictive power of mortality as APACHE II (ROC AUC = 0.901 \pm 0.015), $r_s = 0.87$ REMS superior to RAPS
Olsson ⁷	11,751 nonsurgical patients	Predictive	Prospective, single-centre Comparison predictive power of REMS to RAPS on in-hospital mortality and length of stay	Poor calibration: H-L $\chi^2 = 62$ ($P \leq .001$)	Predictive criterion validity: REMS (ROC AUC = 0.852 \pm 0.014) REMS predictive of mortality OR = 1.40 (95% CI, 1.36-1.45) for each 1-point increase Internal validity: Similar results with split-sample technique Concurrent construct validity: REMS modestly correlated to length of stay ($r_s = 0.47$) REMS superior to RAPS in predicting in-hospital mortality (AUC = 0.852 \pm 0.014 for REMS compared with 0.652 \pm 0.019 for RAPS)
Olsson ²⁰	11,751 nonsurgical patients	Predictive	Prospective, single center Association between REMS and short and long-term mortality	Not reported	Predictive criterion validity: REMS can predict long-term mortality HR = 1.26 ($P < .001$)
Olsson ²¹	865 nonsurgical patients	Predictive	Prospective, single center Could addition of CCI to REMS improve predictive ability of short- and long-term mortality	Not reported	Predictive criterion validity: REMS can predict short- and long-term mortality Concurrent construct validity: REMS weakly correlated to CCI ($r_s = 0.38$) Addition of CCI to REMS does not increase predictive power of mortality
Goodacre ²²	5,583 nonsurgical patients	Predictive	Retrospective, multi-EMS services Comparison predictive power of REMS to RAPS on in-hospital mortality	Not reported	Predictive criterion validity: REMS (ROC AUC = 0.74 \pm 0.04) Concurrent construct validity: REMS superior to RAPS in predicting in-hospital mortality (AUC = 0.74 \pm 0.04 for REMS compared with 0.64 \pm 0.05 for RAPS)
Howell ²³	2,132 patients with suspected infection	Predictive	Prospective, single center Comparison predictive power of REMS, MEDS, and CURB-65 on in-hospital mortality Used "modified" REMS where altered LOC given 1 point, no breakdown for GCS	Interrater reliability: Kappa 0.9 for screening of infection No specific mention of interrater reliability for REMS	Predictive criterion validity: REMS (AUC = 0.802 \pm 0.05) REMS predictive of mortality OR = 1.40 (95% CI, 1.28-1.45) for each 1-point increase Concurrent construct validity: REMS similar predictive strength to MEDS and CURB-65 (AUC = 0.849 \pm 0.04 for MEDS, AUC = 0.788 \pm 0.04 for CURB-65)
Cattermole ²⁴	330 medical, surgical, and trauma patients	Predictive	Prospective, single center Comparison predictive power of REMS, APACHE II, RTS, MEWS, and PEDS on in-hospital mortality	Not reported	Predictive criterion validity: REMS (AUC = 0.696 \pm 0.053) Concurrent construct validity: REMS similar predictive strength to APACHE II, RTS, MEWS
Crowe ²⁵	216 patients with severe sepsis or septic shock	Predictive	Prospective, single center Comparison predictive power of REMS, MEDS, and CURB-65 on in-hospital mortality Used "modified" REMS where altered LOC given 1 point, no breakdown for GCS	Not reported	Predictive criterion validity: REMS (AUC = 0.62 \pm 0.07) Concurrent construct validity: REMS similar predictive strength to MEDS and CURB-65
Ghanem-Zoubi ¹⁰	1,072 patients with sepsis	Predictive	Prospective, single center Comparison predictive power of REMS, MEWS, and MEDS on in-hospital mortality	Adequate calibration: H-L $P = 0.23$ -0.85 across different time intervals	Concurrent construct validity: REMS similar predictive strength to MEDS and MEWS

Table 1 (Continued)

First Author	Study Population	Clinimetric Property Reported	Study Design and Setting	Calibration and Reliability Evidence	Validity Evidence
Carrillo ¹⁷	432 patients requiring interfacility transfer	Predictive and evaluative	Prospective, single center Compare initial REMS, final REMS, and change in REMS with need for hemodynamic or ventilator support	Test-retest reliability: $r = 0.78$, no intraclass correlation coefficient stated	Predictive criterion validity: Higher REMS higher likelihood of need for hemodynamic or ventilator support Longitudinal construct validity: Increase in REMS over time has higher likelihood for hemodynamic or ventilator support
Bulut ¹⁸	2,000 medical and surgical patients	Predictive	Prospective, multicenter Compare REMS and MEWS in predicting ICU admission and in-hospital mortality	Not reported	Predictive criterion validity: REMS (AUC = 0.707 ± 0.020) Concurrent construct validity: REMS modestly correlated to MEWS ($r = 0.422$) REMS superior to MEWS in predicting in-hospital mortality (AUC = 0.630 ± 0.021 for MEWS)
Kuo ²⁶	171 patients with <i>Vibrio vulnificus</i> infection	Predictive	Retrospective, single center Evaluate performance of REMS on predicting in-hospital mortality	Performed but not reported	Predictive criterion validity: Higher REMS associated with higher mortality (AUC = 0.895 ± 0.05)
Moseson ¹¹	227 medical and surgical patients admitted to ICU	Predictive	Prospective, single center Compare REMS with other emergency department and ICU severity scores in predicting in-hospital mortality	Adequate calibration: H-L $P = 0.28-0.77$ across different populations	Concurrent construct validity: REMS had similar predictive performance to other emergency department severity scores
Imhoff ²⁷	3,680 trauma patients	Predictive	Retrospective, single center Evaluate performance of REMS on predicting in-hospital mortality and compare with RTS, SI, and ISS	Not reported	Predictive criterion validity: REMS (AUC = 0.91 ± 0.02) REMS predictive of mortality OR = 1.51 (95% CI, 1.45-1.58) for each 1-point increase Concurrent construct validity: REMS superior to ISS and SI in predicting in-hospital mortality (AUC = 0.89 ± 0.04 for RTS, AUC = 0.87 ± 0.01 for ISS, and AUC = 0.55 ± 0.31 for SI)
Yu ¹²	656 patients admitted with infection	Evaluative	Retrospective, nested case control, single center Evaluate performance of REMS and other severity scores at predicting clinical deterioration by repeated measures	Adequate calibration: H-L $P > .05$	Longitudinal construct validity: Patients who experienced a clinical deterioration had an increase of REMS within 12-24 hours before clinical deterioration
Hilderink ²⁸	600 patients with sepsis	Predictive	Retrospective, single center Evaluate performance of REMS and other severity scores at predicting in-hospital mortality	Not reported	Predictive criterion validity: REMS (AUC = 0.78 ± 0.060) Concurrent construct validity: REMS had similar predictive performance to other emergency department and ICU severity scores
Polita ²⁹	163 trauma patients	Predictive	Retrospective, single center Evaluate performance of REMS, RAPS, and simplified APACHE II at predicting mortality and compare with APACHE II	Not reported	Predictive criterion validity: REMS (AUC 0.761 ± 0.073) Concurrent construct validity: REMS strongly correlated to APACHE II ($r_s = 0.92$) REMS similar predictive performance to RAPS and simplified APACHE II
Ha ¹³	1,746 medical patients	Predictive	Prospective, single center Compare performance of REMS with WPS	Adequate calibration: H-L $P > .05$	Predictive criterion validity: REMS (AUC = 0.712 ± 0.074) REMS predictive of mortality HR = 1.28 (95% CI, 1.23-1.34) for each 1-point increase Concurrent construct validity: REMS modestly correlated to WPS ($r_s = 0.45$) REMS less predictive of mortality than WPS (AUC = 0.797 ± 0.034 for WPS)
El-Sarnagawy ³⁰	104 poisoned patients	Predictive	Prospective, single center Evaluate performance of REMS, RAPS, GCS, and APACHE II at predicting need for intubation	Not reported	Predictive criterion validity: REMS (AUC = 0.758) Concurrent construct validity: REMS had similar predictive performance to RAPS, GCS, and APACHE II
Alter ¹⁹	61,346 noncardiac arrest EMS transports	Evaluative	Retrospective, single EMS service Initial and final REMS calculated to assess change in clinical status and compared to dispatch priority	Not reported	Longitudinal construct validity: Patients with higher dispatch priority had higher changes in REMS (toward direction of lower final REMS score or clinical improvement)

Table 1 (Continued)

First Author	Study Population	Clinimetric Property Reported	Study Design and Setting	Calibration and Reliability Evidence	Validity Evidence
Nakhjavan-Shahraki ¹⁴	2,148 trauma patients	Evaluative	Prospective, multicenter Compare predictive power of REMS with RAPS on in-hospital mortality	Adequate calibration: calibration plot slope = 0.98, intercept 0.001	Predictive criterion validity: REMS (AUC = 0.93 ± 0.02) Concurrent construct validity: REMS similar to RAPS in predicting in-hospital mortality
Miller ³¹	2,680 trauma patients	Evaluative	Retrospective, multicenter Comparison of REMS with modified REMS and other trauma severity scores on predicting mortality	Not reported	Predictive criterion validity: REMS (AUC = 0.911) Concurrent construct validity: REMS similar to modified REMS in predicting in-hospital mortality (AUC = 0.921 for modified REMS)
Seak ¹⁵	66 patients with hepatic portal venous gas	Evaluative	Retrospective, single center Comparison predictive power of REMS, RAPS, and MEWS on in-hospital mortality	Adequate calibration: H-L <i>P</i> = .232	Predictive criterion validity: REMS (AUC 0.929) Concurrent construct validity: REMS similar to RAPS and MEWS in predicting in-hospital mortality
Brabrand ³²	5,784 medical patients	Evaluative	Retrospective, single center Comparison of predictive power of REMS with other emergency severity scores on 24-hour and in-hospital mortality	Adequate calibration: H-L <i>P</i> = .58 (24-hour mortality); H-L <i>P</i> = .07 (in-hospital mortality)	Predictive criterion validity: REMS (AUC = 0.84 ± 0.12) Concurrent construct validity: REMS had similar predictive performance to other emergency department and ICU severity scores
Hung ²³	114 patients with splenic abscess	Evaluative	Retrospective, multicenter Comparison predictive power of REMS with MEDS, RAPS, and MEWS	Not reported	Predictive criterion validity: REMS (AUC = 0.671 for REMS) Concurrent construct validity: REMS had lower predictive performance compared to other scores (AUC = 0.918 for MEDS, AUC 0.757 for MEWS, AUC 0.676 for RAPS)
Park ³⁴	6,905 trauma patients	Evaluative	Retrospective, single center Comparison predictive power of REMS with EMTRAS, RAPS, and ISS	Not reported	Predictive criterion validity: REMS (AUC = 0.900 for REMS) Concurrent construct validity: REMS similar to EMTRAS in predicting in-hospital mortality

APACHE II = Acute Physiology and Chronic Health Evaluation; AUC = area under the curve; CCI = Charlson comorbidity index; CI = confidence interval; EMS = emergency medical services; EMTRAS = emergency trauma score; H-L = Hosmer-Lemeshow goodness-of-fit test; HR = hazards ratio; ISS = Injury Severity Score; MEDS = Mortality in Emergency Department Sepsis Score; MEWS = Modified Early Warning System; OR = odds ratio; PEDS = Prince of Wales Emergency Department Score; *r* = Pearson correlation coefficient; *r_s* = Spearman rank coefficient; RAPS = Rapid Acute Physiology Score; REMS = Rapid Emergency Medicine Score; ROC = receiver operating characteristic curve; RTS = Revised Trauma Score; SI = Shock Index; WPS = Worthing Physiological Scoring System.

information was obtained. The study authors recommend using repeated measures for better predictability or for an evaluative use. The score should be acceptable to patients because it does not require any additional measurements or tests beyond what constitutes standard care provided by the clinician. It should also be acceptable to clinicians because it is simple to use and intuitive to interpret, with higher scores having a higher odds of mortality. No special tests or skills are required. Furthermore, with increasing technologic advances and electronic patient records, it is possible that the REMS could be calculated automatically without any additional time from the clinician.

Reliability

Reliability refers to the extent to which repeat observations give identical results. This can be either with the same rater (intrarater reliability) or a different rater (interrater reliability). Commonly used measures of reliability include proportional agreement, kappa, and the intraclass correlation coefficient.¹⁶ Both of these measures vary from 0 to 1, with values closer to 1 indicating higher reliability. None of the studies of the REMS to date have formally assessed intra- or interrater reliability, and it is important that these factors be studied.

There is inherent variability of a patient's vital signs that can change minute to minute depending on physiologic responses and treatment (ie, resolution of tachycardia with the administration of analgesia). This variability does cause some concern in regard to its overall reliability. However, this potential variability also makes it a

useful tool to measure minute-to-minute physiologic change to the patient. The authors do suggest using the highest REMS to reduce variability and suggest this score may be most predictive; however, the authors do not offer any evidence for this suggestion.

Because the REMS is a clinical index of vital signs, GCS, and age, it is possible to assess reliability by looking at examples from the literature of these individual measures. Prasad¹⁶ found that the reliability of GCS is high in the end-of-range scores (GCS 3–6 and 14–15) as long as the stimulation techniques are standardized and the users are trained and experienced.¹⁶ The reliability in the middle range depends largely on the users' experience and if pseudoscore (assigned a score of 1 for any unmeasurable feature) was used.¹⁶ A study by Carrillo and Urrutia¹⁷ did show good test-retest reliability of the REMS with a Pearson coefficient of 0.78.

Validity

Validity is defined as the extent to which an instrument measures what it is intended to measure. There are many ways of testing validity, the choice depending on the purpose of the measurement. Because the REMS is used for both predictive and evaluative purposes, it is necessary to evaluate its criterion validity, concurrent construct validity, and longitudinal construct validity.

Criterion Validity and Discrimination

Criterion validity evaluates whether your measure correlates with the existing "gold standard" when both are measured at the same

time.¹⁶ In the setting of the REMS, it refers to the ability of the REMS to distinguish whether a patient lived or died. This may also be referred to as the discriminatory power of a clinical index.⁹ For a binary outcome, such as mortality, the receiver operating characteristic curve plots the sensitivity (true positive rate) against 1 – specificity (false-positive rate) for consecutive cutoffs for the probability of an outcome.⁹ Using the receiver operating characteristic curve, an area under the curve (AUC) value of 1 indicates perfect discrimination, with an AUC of greater than 0.8 being considered adequate for models predicting mortality.⁴ Olsson and Lind⁴ found the REMS to have an AUC of 0.911 in their derivation study of medical patients. Additional studies involving medical, trauma, and surgical populations have shown AUCs of 0.62 to 0.93 for the REMS (Table 1). The majority of those studies have an AUC greater than 0.8; therefore, the REMS appears to have good criterion validity and discriminatory power.

Concurrent Construct Validity

Concurrent construct validity refers to the degree to which the measurement of one instrument correlates to another instrument when measuring the same patients at the same time.⁹ One way this can be measured is by using the Pearson correlation coefficient (r) or the Spearman rank coefficient (r_s). There are statistical tests of correlation in which a strong relationship is indicated by an r_s of greater than 0.7, a moderate relationship is an r_s of 0.3 to 0.7, and a weak relationship is an r_s of less than 0.3.⁴ The derivation study found an r_s of 0.87 when assessing correlation to the APACHE II.⁴ The REMS has also been moderately correlated with length of stay,⁷ the modified early warning system,¹⁸ and the Worthing physiological scoring system.¹³ Additionally, multiple studies have found similar or higher discriminatory power to the REMS compared with other commonly used illness severity scores in medical, trauma, and surgical patient populations (Table 1). Therefore, the REMS should be considered to have good concurrent construct validity.

Responsiveness and Longitudinal Construct Validity

Responsiveness has been defined as the ability to accurately detect change when it has occurred and is sometimes also referred to as longitudinal construct validity. Assessing longitudinal construct validity involves correlating changes in scores of the instrument over time with an a priori prediction of the degree of the expected correlation. In the case of the REMS, it would be expected that the most critically ill patients who are likely to deteriorate would have a higher REMS over time.

Only 3 studies have assessed the longitudinal construct validity of the REMS. Carrillo et al¹⁷ studied 432 patients requiring interfacility transfer and compared the initial REMS, the final REMS, and the change in REMS with the need for hemodynamic or ventilator support. They found that a higher REMS over time predicted a higher likelihood for hemodynamic or ventilator support.¹⁷ A study by Yu et al¹² examining 656 patients admitted with infection found that patients who experienced a clinical deterioration had an increase of their REMS within 12 to 24 hours before clinical deterioration. Lastly, a study by Alter et al¹⁹ found that in a population of 61,346 noncardiac arrest EMS transports, patients with a higher dispatch priority were found to have a higher change in the REMS over time (in the direction of a lower final REMS score or clinical improvement). Therefore, although the evidence of longitudinal construct validity is limited, there does appear to be good evidence for its evaluative purpose.

Applicability of REMS in HEMS Setting

Decision Aid to Launch HEMS

One of the persistent challenges with HEMS is identifying patients who benefit most from HEMS transport. Air retrieval teams are costly, resource intensive, and not without risk.^{35–38} The REMS offers

another data point to assist in the decision to launch a retrieval team. It would be possible to integrate the REMS into an application for a land-based EMS service that would automatically calculate the REMS based on the initial vital signs of the land paramedics. If the calculated REMS was above a determined cutoff, a notice could be pushed to the land crew, prompting them to request an HEMS retrieval team.

Triage Decisions

Another potential use for the REMS in the EMS environment would be assisting in triage decisions. In the situation in which there was only a single resource to service multiple retrieval requests, the calculation of the REMS could serve as another data point in assisting with which patient should be transported first. If the REMS was automatically integrated and calculated by the HEMS dispatch operations software, it could allow for comparisons of severity of illness between multiple patients. Although an understanding of complexities of each case is required, the REMS may assist in identifying patients who are more likely to deteriorate without urgent transport.

Assessment of Response to Treatment

As discussed previously, a common limitation to the HEMS database is the lack of in-hospital and mortality data. One benefit of multiple measurements of the REMS could be an assessment for the clinical response to treatment by the retrieval team. By taking repeated measurements over time, it would be possible to track patient improvement or deterioration. This could serve as an additional point of providing feedback to individual providers or assessment of organizational protocols.

Conclusion

The REMS was designed to predict in-hospital mortality for admitted medical patients using only variables that are available in the emergency department and prehospital setting.⁴ Critical appraisal of its development highlighted several strengths in its development and a few limitations. The tool was developed with robust methodology. Item generation was based on known predictors of mortality that are available in the emergency department. Item reduction was thorough with appropriate statistical tests and clinical justification for derivation of the final clinical index. Internal validity was assessed with a split-sample technique. The REMS has good sensibility with adequate content and face validity. It is easy to understand and feasible to be calculated within minutes of patient assessment.

Overall, the REMS has been shown to have the necessary measurement properties to be both a predictive and evaluative clinical index to measure prehospital severity of illness in medical, trauma, and surgical populations. Most studies have found adequate model calibration, with the exception of 1 study by Olsson et al⁷ that did not show adequate model fit using the Hosmer-Lemeshow statistic. The latter finding may be explained by the large sample size of the study. No studies have adequately addressed the intra- or interrater reliability of the score, which does jeopardize both its reliability and validity. This gap should be addressed in future studies. One study did show good test-retest reliability. The REMS showed good criterion validity, discrimination, and concurrent construct validity in medical, trauma, and surgical patient populations. In most studies, it performed as well or better than other illness severity scores in predicting mortality. There was also evidence for good longitudinal construct validity, with 3 studies showing a change in the REMS being associated with clinical deterioration or improvement.

There are multiple potential uses for the REMS in HEMS evaluation. From a triage perspective when multiple patients need to be transported by a single resource, the REMS could identify patients at highest risk of mortality to assist in decision making. Additionally, use of the REMS can evaluate the clinical response to HEMS treatment. This could be useful to provide feedback to individual providers

or assess protocols within an entire HEMS organization. In conclusion, there is evidence to support the use of the REMS as a predictive or evaluative instrument.

Supplementary materials

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.amj.2019.02.001>.

References

- Myers JB, Slovis CM, Eckstein M, et al. Evidence-based performance measures for emergency medical services systems: a model for expanded EMS benchmarking. *Prehosp Emerg Care*. 2008;12:141–151.
- Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9:591–597.
- Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–829.
- Olsson T, Lind L. Comparison of the rapid emergency medicine score and APACHE II in nonsurgical emergency department patients. *Acad Emerg Med*. 2003;10:1040–1048.
- Rhee KJ, Fisher CJ, Willitts NH. The rapid acute physiology score. *Am J Emerg Med*. 1987;5:278–282.
- Bohnen JM, Mustard RA, Oxholm SE, Schouten BD. APACHE II score and abdominal sepsis. A prospective study. *Arch Surg*. 1988;123:225–229.
- Olsson T, Terent A, Lind L. Rapid emergency medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J Intern Med*. 2004;255:579–587.
- Tang K, Beaton DE, Laccaille D, Gignac MAM, Bombardier C. Canadian Arthritis Network (CAN) Work Productivity Group. Sensibility of five at-work productivity measures was endorsed by patients with osteoarthritis or rheumatoid arthritis. *J Clin Epidemiol*. 2013;66:546–556.
- Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–138.
- Ghanem-Zoubi NO, Vardi M, Laor A, Weber G, Bitterman H. Assessment of disease-severity scoring systems for patients with sepsis in general internal medicine departments. *Crit Care*. 2011;15:R95.
- Moseson EM, Zhuo H, Chu J, et al. Intensive care unit scoring systems outperform emergency department scoring systems for mortality prediction in critically ill patients: a prospective cohort study. *J Intensive Care*. 2014;2:40.
- Yu S, Leung S, Heo M, et al. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study. *Crit Care*. 2014;18:R132.
- Ha DT, Dang TQ, Tran NV, Vo NY, Nguyen ND, Nguyen TV. Prognostic performance of the rapid emergency medicine score (REMS) and Worthing physiological scoring system (WPS) in emergency department. *Int J Emerg Med*. 2015;8:18.
- Nakhjavan-Shahraki B, Baikpour M, Yousefifard M, et al. Rapid acute physiology score versus rapid emergency medicine score in trauma outcome prediction; a comparative study. *Emerg (Tehran)*. 2017;5:e30.
- Seak C-J, Yen DH-T, Ng C-J, et al. Rapid emergency medicine score: a novel prognostic tool for predicting the outcomes of adult patients with hepatic portal venous gas in the emergency department. *PLoS One*. 2017;12:e0184813.
- Prasad K. The Glasgow Coma Scale: a critical appraisal of its clinimetric properties. *J Clin Epidemiol*. 1996;49:755–763.
- Carrillo B MJ, Urrutia S MT. [Risk assessment of critical patients requiring ambulance transfers]. *Rev Med Chil*. 2012;140:1297–1303.
- Bulut M, Cebicci H, Sigirli D, et al. The comparison of modified early warning score with rapid emergency medicine score: a prospective multicentre observational cohort study on medical and surgical patients presenting to emergency department. *Emerg Med J*. 2014;31:476–481.
- Alter SM, Infinger A, Swanson D, Studnek JR. Evaluating clinical care in the pre-hospital setting: is rapid emergency medicine score the missing metric of EMS? *Am J Emerg Med*. 2017;35:218–221.
- Olsson T, Terent A, Lind L. Rapid emergency medicine score can predict long-term mortality in nonsurgical emergency department patients. *Acad Emerg Med*. 2004;11:1008–1013.
- Olsson T, Terent A, Lind L. Charlson Comorbidity Index can add prognostic information to rapid emergency medicine score as a predictor of long-term mortality. *Eur J Emerg Med*. 2005;12:220–224.
- Goodacre S, Turner J, Nicholl J. Prediction of mortality among emergency medical admissions. *Emerg Med J*. 2006;23:372–375.
- Howell MD, Donnino MW, Talmor D, Clardy P, Ngo L, Shapiro NI. Performance of severity of illness scoring systems in emergency department patients with infection. *Acad Emerg Med*. 2007;14:709–714.
- Cattermole GN, Mak SKP, Liow CHE, et al. Derivation of a prognostic score for identifying critically ill patients in an emergency department resuscitation room. *Resuscitation*. 2009;80:1000–1005.
- Crowe CA, Kulstad EB, Mistry CD, Kulstad CE. Comparison of severity of illness scoring systems in the prediction of hospital mortality in severe sepsis and septic shock. *J Emerg Trauma Shock*. 2010;3:342–347.
- Kuo S-H, Tsai C-F, Li C-R, et al. Rapid emergency medicine score as a main predictor of mortality in vibrio vulnificus-related patients. *Am J Emerg Med*. 2013;31:1037–1041.
- Imhoff BF, Thompson NJ, Hastings MA, Nazir N, Moncure M, Cannon CM. Rapid emergency medicine score (REMS) in the trauma population: a retrospective study. *BMJ Open*. 2014;4:e004738.
- Hilderink MJ, Roest AA, Hermans M, Keulemans YC, Stehouwer CD, Stassen PM. Predictive accuracy and feasibility of risk stratification scores for 28-day mortality of patients with sepsis in an emergency department. *Eur J Emerg Med*. 2015;22:331–337.
- Polita JR, Gomez J, Friedman G, Ribeiro SP. Comparison of APACHE II and three abbreviated APACHE II scores for predicting outcome among emergency trauma patients. *Rev Assoc Med Bras (1992)*. 2014;60:381–386.
- El-Sarnagawy GN, Hafez AS. Comparison of different scores as predictors of mechanical ventilation in drug overdose patients. *Hum Exp Toxicol*. 2017;36:539–546.
- Miller RT, Nazir N, McDonald T, Cannon CM. The modified rapid emergency medicine score: a novel trauma triage tool to predict in-hospital mortality. *Injury*. 2017;48:1870–1877.
- Brabrand M, Hallas P, Hansen SN, Jensen KM, Madsen JL, Posth S. Using scores to identify patients at risk of short term mortality at arrival to the acute medical unit: a validation study of six existing scores. *Eur J Intern Med*. 2017;45:32–36.
- Hung S-K, Ng C-J, Kuo C-F, et al. Comparison of the mortality in emergency department sepsis score, modified early warning score, rapid emergency medicine score and rapid acute physiology score for predicting the outcomes of adult splenic abscess patients in the emergency department. *PLoS One*. 2017;12:e0187495.
- Park HO, Kim JW, Kim SH, et al. Usability verification of the emergency trauma score (EMTRAS) and rapid emergency medicine score (REMS) in patients with trauma: a retrospective cohort study. *Medicine (Baltimore)*. 2017;96:e8449.
- Doucet J, Bulger E, Sanddal N, Fallat M, Bromberg W, Gestring M. Appropriate use of helicopter emergency medical services for transport of trauma patients. *J Trauma Acute Care Surg*. 2013;75:734–741.
- Delgado MK, Staudenmayer KL, Wang NE, et al. Cost-effectiveness of helicopter versus ground emergency medical services for trauma scene transport in the United States. *Ann Emerg Med*. 2013;62:351–364.
- Taylor CB, Stevenson M, Jan S, Middleton PM, Fitzharris M, Myburgh JA. A systematic review of the costs and benefits of helicopter emergency medical services. *Injury*. 2010;41:10–20.
- Bledsoe BE, Smith MG. Medical helicopter accidents in the United States: a 10-year review. *J Trauma*. 2004;56:1325–1328.