



Published in final edited form as:

Resuscitation. 2016 May ; 102: 1–5. doi:10.1016/j.resuscitation.2016.02.005.

The value of vital sign trends for detecting clinical deterioration on the wards

Matthew M Churpek, MD, MPH, PhD^{1,*}, Richa Adhikari, MPH¹, and Dana P Edelson, MD, MS¹

¹Department of Medicine, University of Chicago, Chicago, IL

Abstract

Aim—Early detection of clinical deterioration on the wards may improve outcomes, and most early warning scores only utilize a patient’s current vital signs. The added value of vital sign trends over time is poorly characterized. We investigated whether adding trends improves accuracy and which methods are optimal for modelling trends.

Methods—Patients admitted to five hospitals over a five-year period were included in this observational cohort study, with 60% of the data used for model derivation and 40% for validation. Vital signs were utilized to predict the combined outcome of cardiac arrest, intensive care unit transfer, and death. The accuracy of models utilizing both the current value and different trend methods were compared using the area under the receiver operating characteristic curve (AUC).

Results—A total of 269,999 patient admissions were included, which resulted in 16,452 outcomes. Overall, trends increased accuracy compared to a model containing only current vital signs (AUC 0.78 vs. 0.74; $p < 0.001$). The methods that resulted in the greatest average increase in accuracy were the vital sign slope (AUC improvement 0.013) and minimum value (AUC improvement 0.012), while the change from the previous value resulted in an average worsening of the AUC (change in AUC -0.002). The AUC increased most for systolic blood pressure when trends were added (AUC improvement 0.05).

Conclusion—Vital sign trends increased the accuracy of models designed to detect critical illness on the wards. Our findings have important implications for clinicians at the bedside and for the development of early warning scores.

*Corresponding author and requests for reprints (Matthew M Churpek), Mail: University of Chicago Medical Center, Section of Pulmonary and Critical Care Medicine, 5841 South Maryland Avenue, MC 6076, Chicago, IL 60637, matthew.churpek@uchospitals.edu; Office: (773) 702-1092; Fax: (773) 702-6500.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Contributions: Study concept and design: M.C., D.P.E.; acquisition of data: D.P.E.; analysis and interpretation of data: all authors; first drafting of the manuscript: M.C.; critical revision of the manuscript for important intellectual content: all authors; statistical analysis: R.A, M.C.; obtained funding: M.C., D.P.E.; administrative, technical, and material support: all authors; study supervision: M.C., D.P.E. Dr. Churpek had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Keywords

vital signs; heart arrest; hospital rapid response team; physiologic monitoring; quality improvement; early diagnosis; decision support techniques

INTRODUCTION

Early detection of critical illness is key to achieving timely transfer to the intensive care unit (ICU) and decreasing the rate of preventable in-hospital cardiac arrest. Vital signs have been shown to be the most accurate predictors of clinical deterioration.¹ Early warning scores consisting of vital sign severity thresholds have been implemented across the United States and around the world in order to accurately detect high-risk ward patients.^{2, 3} These scores typically utilize only the current vital sign values and rarely include trends of vital signs over time.^{3, 4} Although clinicians often include the trend in a patient's condition over time when assessing a patient, the additional value of vital sign trends to risk scores containing a patient's current values is poorly characterized, but has the potential to increase accuracy and decrease false alarms.

Although the idea of including vital sign trends in early warning scores sounds intuitive and straightforward, the low frequency of monitoring (e.g., every four hours), interventions provided to patients, and manual assessment of some of the variables add additional complexity. For example, treatments are often administered in an attempt to "normalize" vital signs, such as acetaminophen for fever and fluid boluses for hypotension. In addition, vital signs may be collected soon after a patient was ambulatory, which may cause a patient to meet the systemic inflammatory response criteria, or may not be accurately quantified, such as always inputting a respiratory rate of 18.^{5, 6} Therefore, simply including the change of a vital sign since last collection may not adequately capture a patient's true physiologic trajectory and additional methods, such as including vital sign variability, the most deranged previous values, and even smoothing the trajectory, may prove to be more accurate.

The aim of this study was to utilize a large, multicentre dataset to compare the accuracy of different methods of modelling vital sign trends for detecting clinical deterioration on the wards.

MATERIALS AND METHODS

Study Population and Data Sources

The study population and data sources have been described previously.^{1, 7} Briefly, we included all ward patients at the University of Chicago and four NorthShore University HealthSystem hospitals between November 2008 and January 2013. Patient vital sign data, which were both time- and location-stamped, were obtained from the Electronic Data Warehouse at NorthShore and the electronic health record (EPIC, Verona, WI) at the University of Chicago. Demographic information was obtained from administrative databases and cardiac arrest data were collected from quality improvement databases and manually checked for accuracy. Based on general impracticability and minimal harm,

waivers of consent were granted by the University of Chicago Institutional Review Board (IRB #16995A) and NorthShore University HealthSystem (IRB #EH11-258).

Primary Outcome

The primary outcome of interest was the development of critical illness on the wards, defined as the composite outcome of a ward cardiac arrest, ward to ICU transfer, or death. If a patient experienced multiple events during the same ward stay (e.g., a cardiac arrest followed immediately by a ward to ICU transfer), the time of the first event was used for the composite outcome. Patients with multiple ward stays during the same admission (e.g. ward to ICU transfers who were later transferred back out to the wards) had each ward stay analysed separately.

Predictor variables

The predictor variables utilized in this study were commonly collected vital signs and their trends over time (i.e., temperature, heart rate, respiratory rate, oxygen saturation, diastolic blood pressure, and systolic blood pressure). The following trend variables were investigated in this study: change in current value from the previous value (delta), mean of the previous six values (mean), standard deviation of the previous six values (SD), slope of the previous six values (slope), minimum value prior to current value (minimum), maximum value prior to current value (maximum), and an exponential smoothing method (smoothed): ($s_0 = x_0$, $s_t = \alpha x_t + (1 - \alpha)s_{t-1}$). The smoothed method involves taking a weighted average of the current and prior values, with a weight of α assigned to the current values and a weight of $1 - \alpha$ for the previous values. Thus, a weight of 1 would include only the current vital sign value and a weight of 0 would include only previous values for the smoothed variable. We chose to use the previous six values for the mean, SD, and slope variables because vital signs are typically collected every four hours so this would utilize approximately 24 hours of data and to standardize the amount of data used for each time point for each patient.

Statistical analysis

We divided the cohort into two subsets in order to develop the models in the training set (60% of the data) and estimate accuracy in the validation set (40% of the data). Because vital signs change over time during a patient's ward stay, discrete time survival analysis was utilized to model these data.^{1, 8} Based on the fact that vital signs were collected every four hours on average in this dataset, four-hour time intervals were chosen for the discrete-time model. Thus, variable values at the beginning of each time block were used to predict whether an event occurred during that four-hour time block. We have previously utilized discrete-time models to develop early warning scores, and its advantages include the ability to model time-varying predictors and to remove the bias that may occur if sicker patients receive more frequent vital signs.¹ All models were fit in the training cohort only and then accuracy was tested in the validation cohort. Ten-fold cross-validation was used in the training cohort to choose the smoothing factor (α) for each vital sign's smoothed trend variable based on the α that maximized the model AUC.

For each variable investigated in the study (i.e., all current and trend variables for each vital sign), univariate models were fit using that predictor variable alone, bivariable models were

fit using that variable plus the current value, and a full model was fit for each vital sign that utilized the current value and all trend variables. Finally, a full model was fit that included all current and trend variables for all the vital signs, and this full model's accuracy was compared to a model fit using only current vital sign values. All models in the study used restricted cubic splines with three knots, with knot placement as recommended by Harrell, for all continuous variables.⁹ This flexible method allows the probability of the event to increase at both low and high values of each variable. If any individual vital signs were missing for model estimation then the previous value was pulled forward. If no prior values were available then a median value was imputed, similar to prior work in this area.^{1, 10} For trend values, if fewer than six data points were available then the trend variables were calculated using all available data points, and if no prior values were available then the current value was imputed, except for the delta, SD, slope variables where a median value was imputed.

Model accuracy was compared in the validation cohort by calculating predicted probabilities from each model and then calculating the area under the receiver operating characteristic curve (AUC) based on outcomes occurring within 24 hours of each vital sign observation time.¹¹ This metric was used because it is a standard way of comparing early warning scores in the literature.^{1, 12, 13} All analyses were performed using Stata version 13.1 (StataCorps; College Station, Texas), and two-sided p-values <0.05 were considered statistically significant.

RESULTS

A total of 269,999 patient admissions were included in the study, which resulted in 16,452 outcomes (424 ward cardiac arrests, 13,188 ICU transfers, and 2,840 deaths on the ward) occurring during the study period. Our study population was 60% female, 52% white, and had an average age of 60 years. Additional details have been described elsewhere.^{1, 14}

During univariate analysis, respiratory rate was the most accurate vital sign when using the current value (AUC 0.70 (95% CI 0.70–0.70), and the trend values were more accurate than the current value for the variability in respiratory rate (AUC 0.71 (95% CI 0.71–0.71) for SD), smoothed heart rate (AUC 0.64 (95% CI 0.64–0.65) vs. 0.63 (95% CI 0.63–0.64) for the current value), diastolic blood pressure slope (AUC 0.61 (95% CI 0.61–0.61) vs. 0.60 (95% CI 0.59–0.60) for the current value), and minimum oxygen saturation (AUC 0.60 (95% CI 0.60–0.60) vs. 0.59 (95% CI 0.59–0.59) for the current value). The results from the bivariate models, which include both the current value and the trend value, are shown in Figures 1,2,3,4,5, 6. As shown in Supplemental Figure 1, the methods that resulted in the greatest average increase in accuracy were the vital sign slope (AUC improvement 0.013), minimum value (AUC improvement 0.012), and SD (AUC improvement 0.01), while the change from the previous value (delta) resulted in an average worsening of the AUC when added to the current value (change in AUC –0.002).

When comparing a model that utilized all trend variables compared to a model only utilizing the current value (Supplemental Figure 2), systolic blood pressure had the most improvement in accuracy (AUC increase of 0.05), followed by oxygen saturation and

respiratory rate (AUC increase of 0.04 for both models). Finally, the model containing all the trend variables of all vital signs had a higher AUC than a model containing only current values of the vital signs (AUC 0.78 vs. 0.74; $p < 0.001$). This increase in accuracy by adding trends was similar across the individual outcomes (0.77 (95% CI 0.76–0.78) vs. 0.74 (95% CI 0.73–0.75) for cardiac arrest, 0.77 (95% CI 0.77–0.77) vs. 0.73 (95% CI 0.73–0.73) for ICU transfer, and 0.90 (95% CI 0.89–0.90) vs. 0.87 (95% CI 0.87–0.87) for death).

DISCUSSION

In this large, multicentre study evaluating the value of vital sign trends, we found that trajectories of these variables significantly improved the accuracy of detecting clinical deterioration compared to the current vital sign values alone. The optimal method of modelling trend varied across the different vital signs. Importantly the simplest method, taking the difference from the previous value, was the least accurate method of modelling trends of the different techniques we studied. Methods such as the vital sign slope, vital sign variability, and the most deranged values since admission were more accurate for most of the vital signs studied. These findings have important implications for clinicians interpreting vital sign trends at the bedside, as well as for the development of early warning scores. Accuracy is paramount with these scores in order to get critical care resources to the bedside while avoiding alarm fatigue, and our study shows that trends in physiology are important.

There are currently over 100 different published early warning scores and there are likely many more in use in hospitals across the country.^{2–4} Most scores, such as the commonly cited Modified Early Warning Score (MEWS),¹⁵ only utilize a patient's current vital sign values when calculating a score.^{3, 4} The few scores that include trends over time typically utilize the change since the last vital sign observation. The fact that that this method was never the best way to model trends for any of the vital signs in our study has important implications for these scores and suggests that different methods to incorporate trends are needed to improve accuracy. Although calculating trends over time and vital sign variability would be error-prone to do by hand, electronic health records are commonplace in the United States and could provide a means to calculate these variables automatically.^{16–19}

Our finding that trends of vital signs are independent predictors of critical illness in ward patients is consistent with other studies.^{17, 20} For example, Escobar and colleagues developed a prediction model for ICU transfer and death on the wards using vital sign, laboratory, demographic, and additional patient data.¹⁷ They also found that trends in vital signs, such as the variability of respiratory rate and minimum oxygen saturation, were independent predictors of clinical deterioration in addition to the most recent vital sign values. Their final model, which also contained patient comorbidity and laboratory data, had an AUC of 0.78 for their combined outcome in the validation dataset. In addition, Mao et al. developed a model in a single-centre study to predict ward to ICU transfer by utilizing both current and previous vital signs and laboratory values.²⁰ The highest weighted variables in their final model included the maximum respiratory rate and the lowest oxygen saturation, and they also used exponential smoothing, as we did in our study, to improve the final predictions of their model. Groups in other areas have also studied trends in the ICU, in the pre-hospital setting, and in various disease states with mixed results.^{21–26} Our study extends

these findings by directly comparing multiple methods for modelling trends for predicting clinical deterioration in ward patients. Because no single method for modelling trends was best for all vital signs, careful consideration is needed when incorporating these trends in early warnings scores.

Although we studied many different methods for modelling vital sign trend data, including those proposed by other groups, there are many more methods available. In particular, there are many techniques that are useful when data are more frequently updated than ward vital signs. For example, Hravnak and colleagues have used continuous vital sign data to study the accuracy of an integrated monitoring system in a step-down unit, which detected deterioration six hours earlier than the Modified Early Warning Score.²⁷ In addition, groups have published from the MIMIC II dataset, which includes frequent vital signs in ICU patients, and have shown that vital sign trends can accurately predict clinical instability.²⁸ In our study, data was only updated approximately every four hours, much less frequently than in the studies noted above, so we did not pursue these other methods.

Our study has several limitations. First, we only investigated vital signs that were collected intermittently, and other time series methods may be more accurate for vital signs collected at a higher rate than our study. Second, our outcome of interest was a composite outcome of ICU transfer, cardiac arrest, and death, and it is possible that the optimal method for capturing vital sign trends may differ for other outcomes. Third, we did not have access to code status or to whether particular ICU transfers were elective vs. non-elective. Accounting for these factors may alter the accuracy measures in the study. Fourth, this study utilized data from five hospitals in Illinois, and these findings need to be externally validated in other hospital settings and countries. Finally, the trend metrics are complex and would require automated calculation if implemented in real-time to detect clinical deterioration.

CONCLUSIONS

In this large, multicentre study, we found that adding trends of vital signs significantly increased the accuracy of models designed to detect critical illness on the wards. Our findings have important implications for clinicians interpreting vital sign trends at the bedside, as well as for the development of early warning scores. Accuracy is paramount with these scores in order to get the right people to the bedside while avoiding alarm fatigue, and our study shows that trends in physiology are important.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Conflicts of Interest/ Sources of Funding: This research was funded in part by an institutional Clinical and Translational Science Award grant (UL1 RR024999; PI: Solway). Dr. Churpek is supported by a career development award from the National Heart, Lung, and Blood Institute (K08 HL121080). Drs. Churpek and Edelson have a patent pending (ARCD. P0535US.P2) for risk stratification algorithms for hospitalized patients. In addition, Dr. Edelson has received research support and honoraria from Philips Healthcare (Andover, MA), research support from the American Heart Association (Dallas, TX) and Laerdal Medical (Stavanger, Norway), and an

honorarium from Early Sense (Tel Aviv, Israel). She has ownership interest in Quant HC (Chicago, IL), which develops products for risk stratification of hospitalized patients.

We would like to thank Timothy Holper, Justin Lakeman, and Contessa Hsu for assistance with data extraction and technical support, Poome Chamnankit, MS, CNP, Kelly Bhatia, MSN, ACNP, and Audrey Seitman, MSN, ACNP for performing manual chart review of cardiac arrest patients, Nicole Twu, MS for administrative support, Christopher Winslow, MD and Ari Robicsek, MD, and Robert Gibbons, PhD for introducing us to discrete-time survival analysis.

REFERENCES

1. 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; 190:649–655. [PubMed: 25089847]
2. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. *Chest*. 2013; 143:1758–1765. [PubMed: 23732586]
3. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation*. 2008; 77:170–179. [PubMed: 18249483]
4. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI, Higgins B. A review, and performance evaluation, of single-parameter "track and trigger" systems. *Resuscitation*. 2008; 79:11–21. [PubMed: 18620794]
5. Semler MW, Stover DG, Copland AP, et al. Flash mob research: A single-day, multicenter, resident-directed study of respiratory rate. *Chest*. 2013; 143:1740–1744. [PubMed: 23197319]
6. Vincent JL. Dear sirs, i'm sorry to say that i don't like you. *Crit Care Med*. 1997; 25:372–374. [PubMed: 9034279]
7. Churpek MM, Yuen TC, Winslow C, Hall J, Edelson DP. Differences in vital signs between elderly and nonelderly patients prior to ward cardiac arrest. *Crit Care Med*. 2015; 43:816–822. [PubMed: 25559439]
8. Singer JD, Willett JB. Its about time - using discrete-time survival analysis to study duration and the timing of events. *J Educ Stat*. 1993; 18:155–195.
9. Harrell, FE. Regression modeling strategies : With applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
10. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991; 100:1619–1636. [PubMed: 1959406]
11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*. 1988; 44:837–845. [PubMed: 3203132]
12. 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; 84:465–470. [PubMed: 23295778]
13. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. Views--towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation*. 2010; 81:932–937. [PubMed: 20637974]
14. Churpek MM, Zadavec FJ, Winslow C, Howell MD, Edelson DP. Incidence and prognostic value of the systemic inflammatory response syndrome and organ dysfunctions in ward patients. *Am J Respir Crit Care Med*. 2015; 192:958–964. [PubMed: 26158402]
15. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning score in medical admissions. *QJM : monthly journal of the Association of Physicians*. 2001; 94:521–526. [PubMed: 11588210]
16. Prytherch DR, Smith GB, Schmidt P, et al. Calculating early warning scores--a classroom comparison of pen and paper and hand-held computer methods. *Resuscitation*. 2006; 70:173–178. [PubMed: 16806641]

17. 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; 7:388–395. [PubMed: 22447632]
18. Sawyer AM, Deal EN, Labelle AJ, et al. Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. *Crit Care Med.* 2011; 39:469–473. [PubMed: 21169824]
19. Bailey TC, Chen Y, Mao Y, et al. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. *J Hosp Med.* 2013; 8:236–242. [PubMed: 23440923]
20. Mao, Y.; Chen, Y.; Hackmann, G., et al. Medical data mining for early deterioration warning in general hospital wards; Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on; 2011. p. 1042-1049.
21. Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the sofa score to predict outcome in critically ill patients. *JAMA.* 2001; 286:1754–1758. [PubMed: 11594901]
22. Cabre L, Mancebo J, Solsona JF, et al. Multicenter study of the multiple organ dysfunction syndrome in intensive care units: The usefulness of sequential organ failure assessment scores in decision making. *Intensive Care Med.* 2005; 31:927–933. [PubMed: 15856171]
23. Wu BU, Bakker OJ, Papachristou GI, et al. Blood urea nitrogen in the early assessment of acute pancreatitis: An international validation study. *Arch Intern Med.* 2011; 171:669–676. [PubMed: 21482842]
24. Chen L, Reisner AT, Gribok A, Reifman J. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. *Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors.* 2009; 13:286–294.
25. Wong J, Taljaard M, Forster AJ, van Walraven C. Does adding risk-trends to survival models improve in-hospital mortality predictions? A cohort study. *BMC health services research.* 2011; 11:171. [PubMed: 21777460]
26. Masson S, Latini R, Anand IS, et al. Prognostic value of changes in n-terminal pro-brain natriuretic peptide in Val-HeFT (Valsartan Heart Failure Trial). *J Am Coll Cardiol.* 2008; 52:997–1003. [PubMed: 18786480]
27. Hravnak M, Edwards L, Clontz A, Valenta C, Devita MA, Pinsky MR. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. *Arch Intern Med.* 2008; 168:1300–1308. [PubMed: 18574087]
28. Cao, H.; Eshelman, L.; Chbat, N.; Nielsen, L.; Gross, B.; Saeed, M. Predicting ICU hemodynamic instability using continuous multiparameter trends; Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. *IEEE Engineering in Medicine and Biology Society;* 2008. p. 3803-3806.

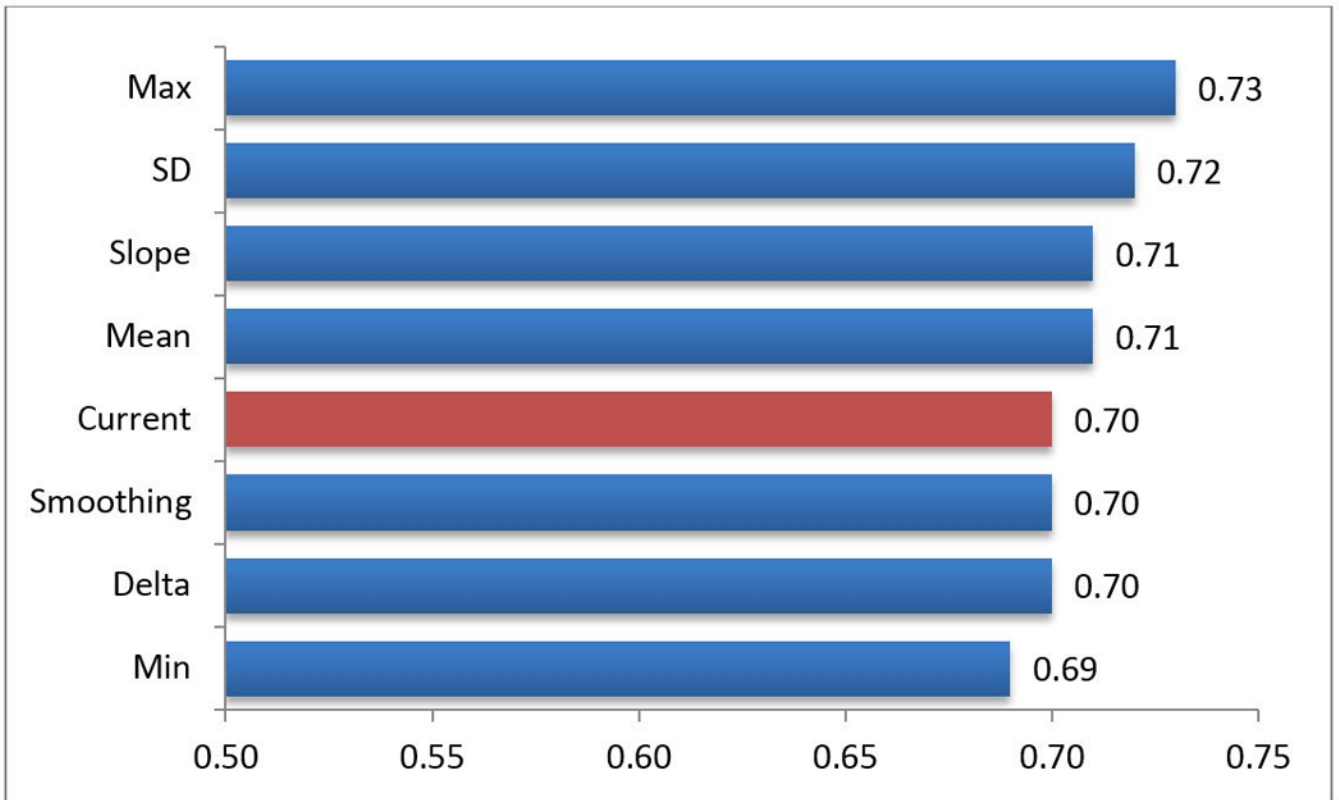


Figure 1. Respiratory rate: AUCs for the current value and the bivariable trend models

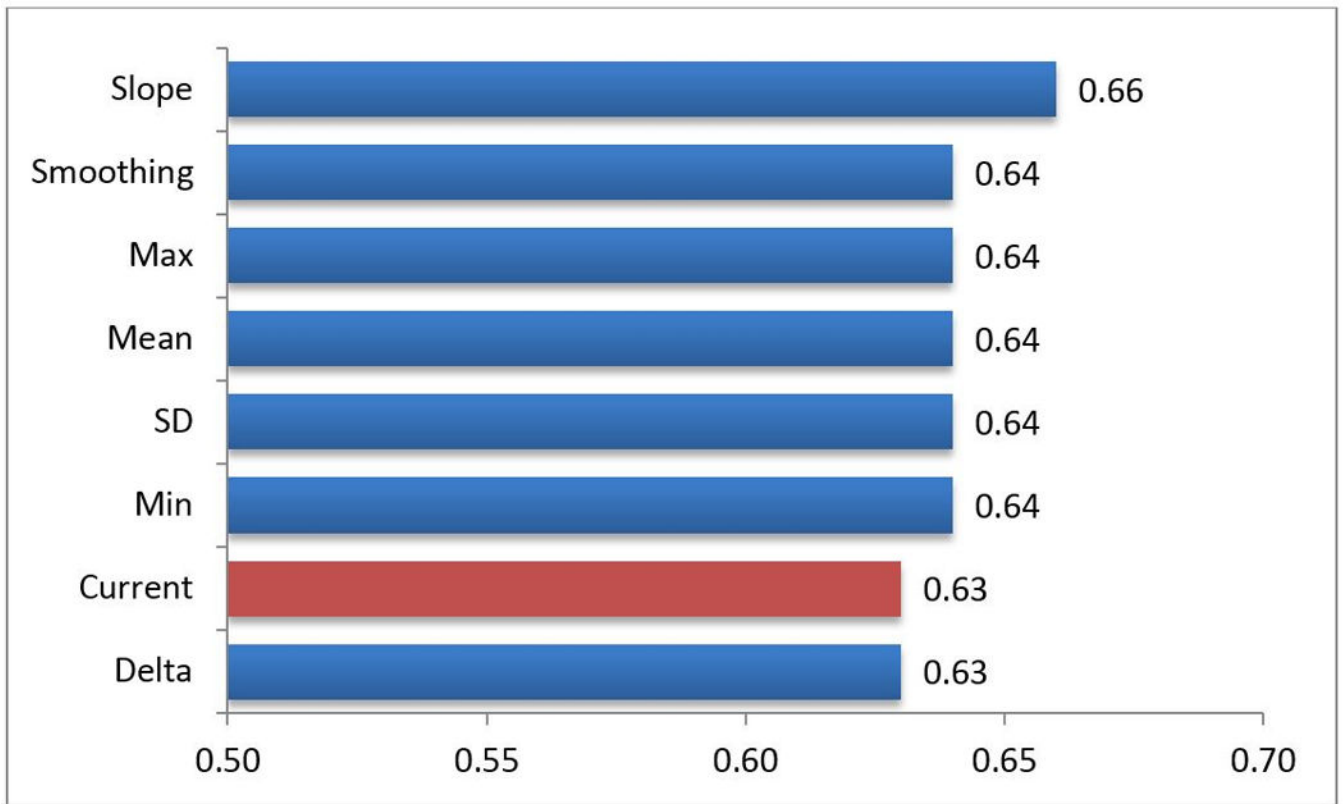


Figure 2. Heart rate: AUCs for the current vital sign value and the bivariable trend models

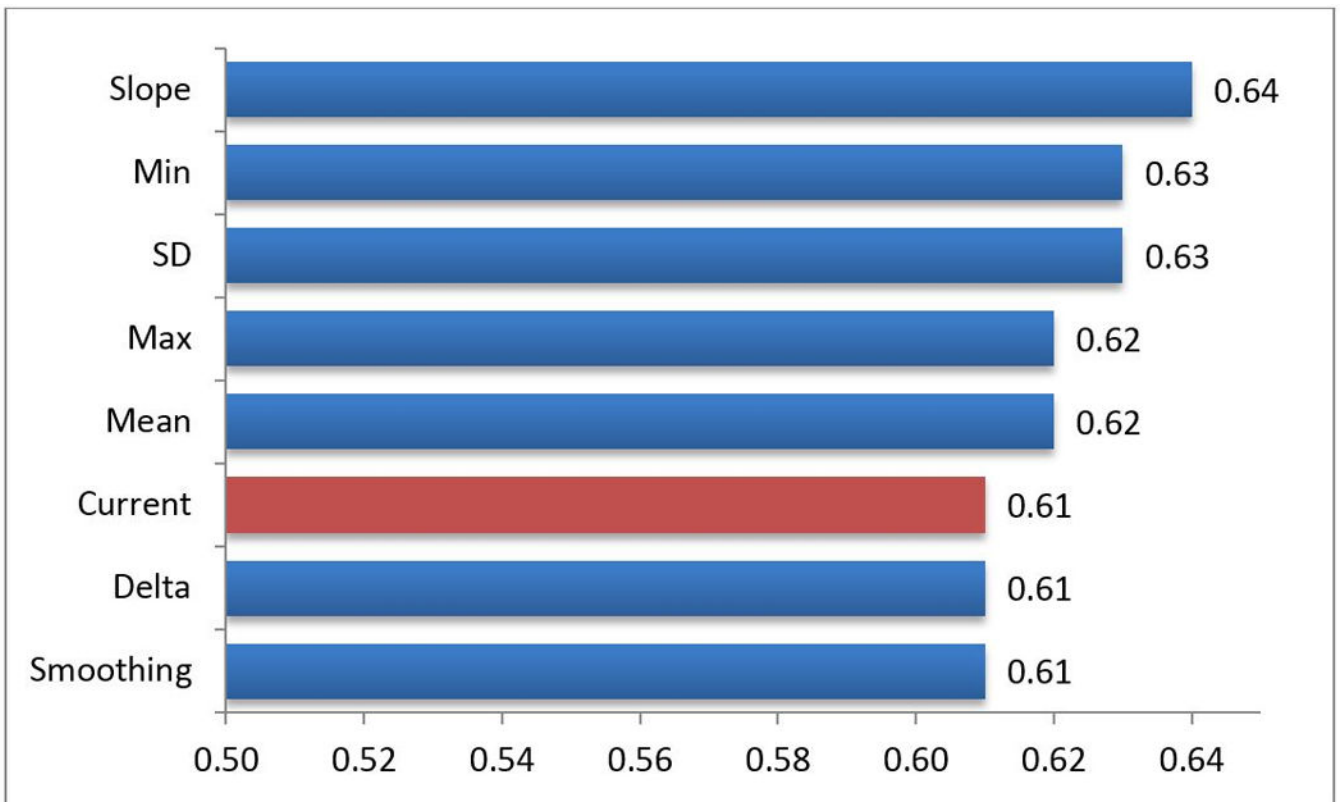


Figure 3. Systolic blood pressure: AUCs for the current vital sign value and the bivariable trend models

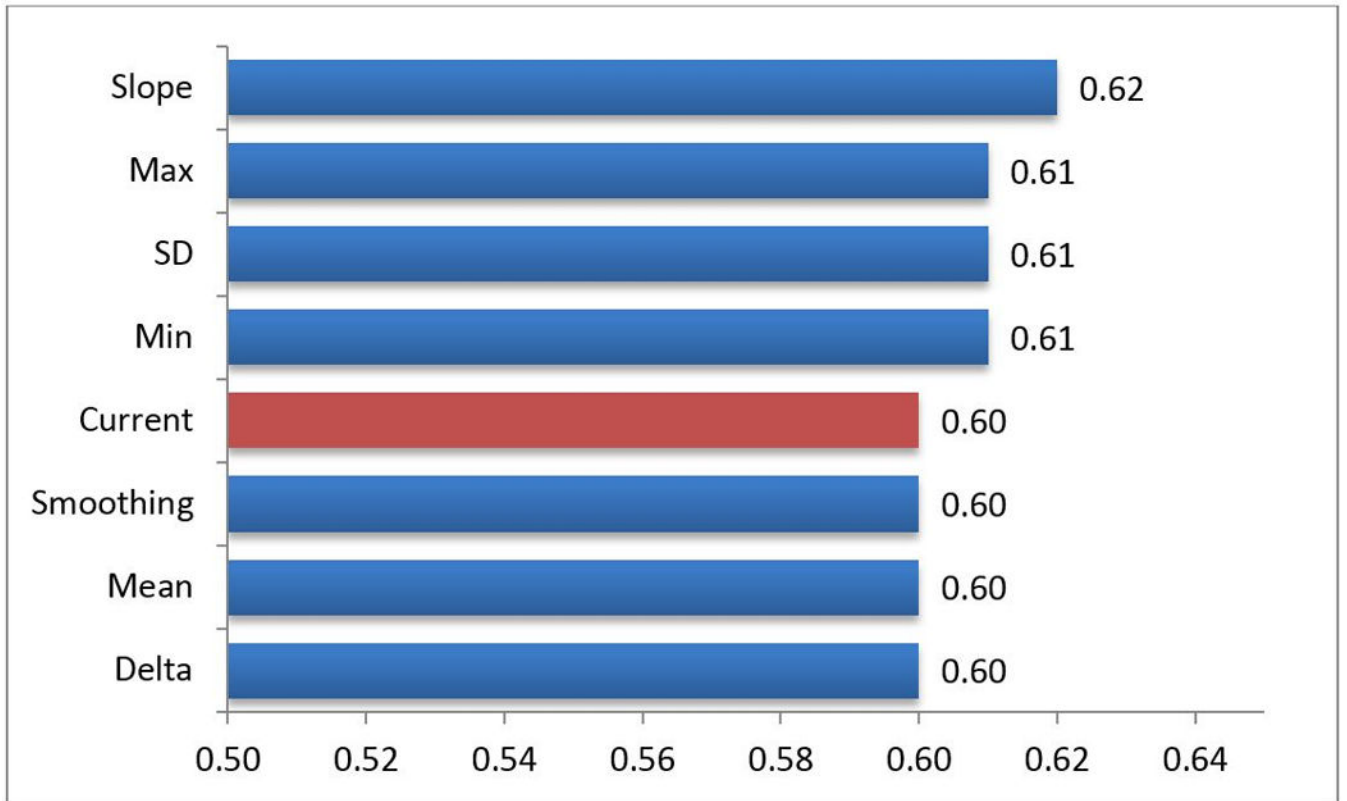


Figure 4. Diastolic blood pressure: AUCs for the current vital sign value and the bivariable trend models

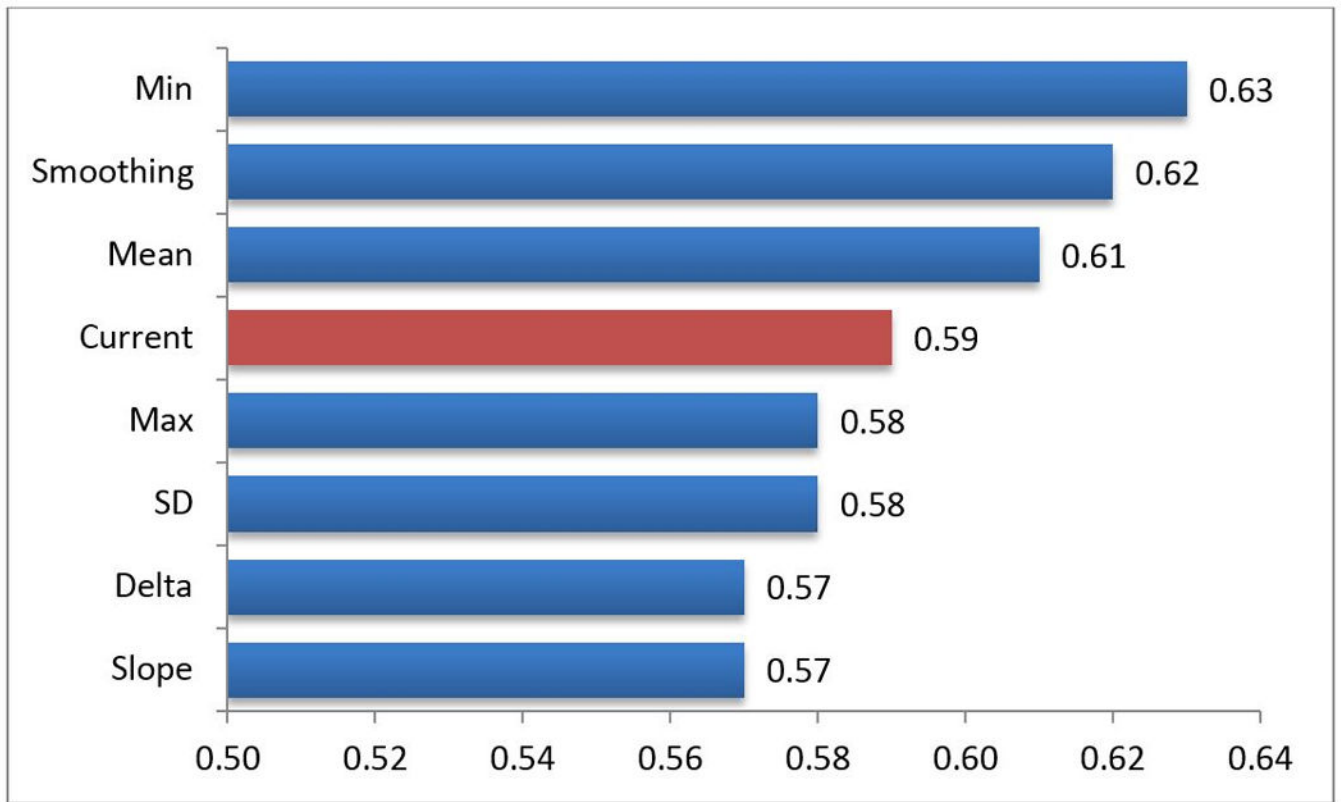


Figure 5. Oxygen saturation: AUCs for the current vital sign value and the bivariable trend models

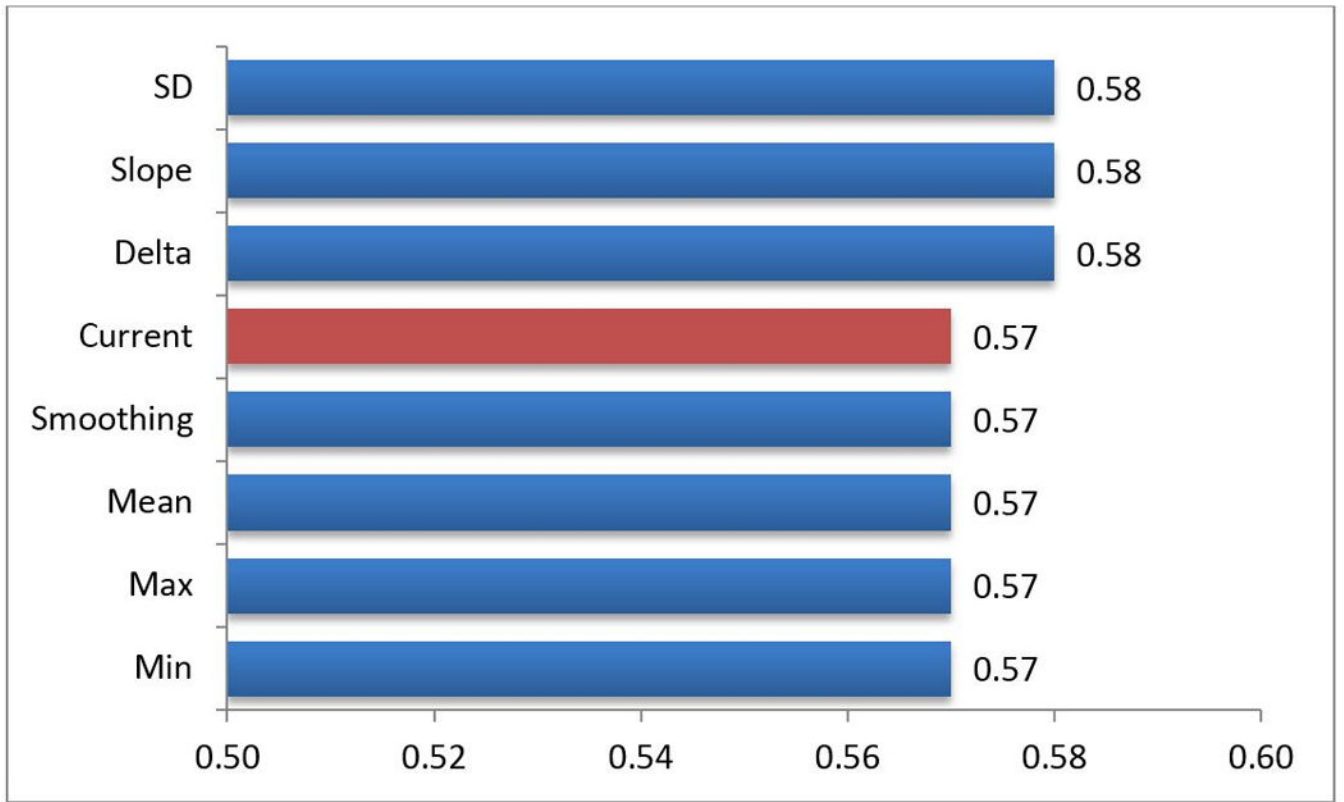


Figure 6. Temperature: AUCs for the current vital sign value and the bivariable trend models