

The SEX-SHOCK score—the emperor’s new clothes?

Karl-Patrik Kresoja ^{1,2}, Maria Rubini Giménez ^{3,4}, and Holger Thiele ^{1,*}

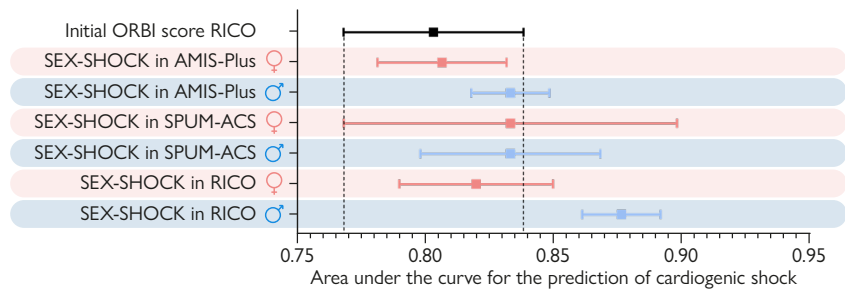
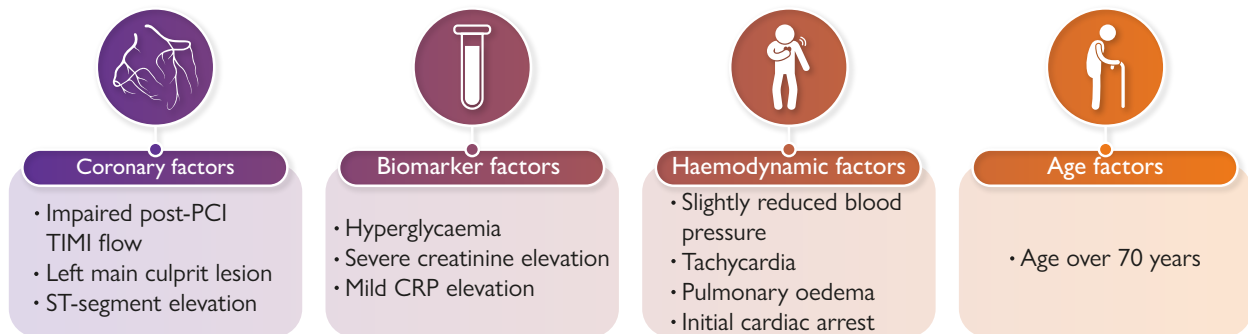
¹Department of Internal Medicine/Cardiology, Heart Center Leipzig at Leipzig University, Strümpellstr. 39, 04289 Leipzig, Germany; ²Department of Cardiology, University Medical Center, Johannes Gutenberg University, Mainz, Germany; ³Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain; and ⁴Department of Cardiology, Imed Hospitales, Valencia, Spain

Online publish-ahead-of-print 1 September 2024

This editorial refers to ‘Sex-specific prediction of cardiogenic shock after acute coronary syndromes: the SEX-SHOCK score’, by Y. Wang *et al.*, <https://doi.org/10.1093/eurheartj/ehae593>.

Graphical Abstract

Predisposing factors for cardiogenic shock in acute coronary syndrome



Risk factors for cardiogenic shock development and added statistical benefit of SEX-SHOCK score. TIMI, thrombolysis in myocardial infarction; PCI, percutaneous coronary intervention; CRP, C-reactive protein.

The opinions expressed in this article are not necessarily those of the Editors of the *European Heart Journal* or of the European Society of Cardiology.

* Corresponding author. Email: holger.thiele@medizin.uni-leipzig.de

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Cardiology. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Infarct-related cardiogenic shock (CS) constitutes one of the deadliest conditions with mortality rates up to 50% within the first 30 days.¹ Therapeutic options are limited and therapies have shown a lack of success in this context.^{1,2} Fortunately, owing to advances in medical and interventional therapy, only a small fraction of patients with acute coronary syndrome (ACS) develops CS.³ Since CS is recognized as a continuum, it is possible that life-threatening changes requiring earlier intervention may already be necessary, even if the clinical manifestation of shock has not yet occurred.⁴ The Observatoire Régional Breton sur l'Infarctus—ORBI—score has been developed as a first step in this endeavour by deriving and successfully validating a score for the prediction of CS in patients with ACS with an impressive area under the curve (AUC) of 0.80 (95% confidence interval [CI] 0.77 to 0.84).⁵ Wang *et al.* wanted to further address this problem by postulating that sex-specific performances of the ORBI score might differ and that adjusting for this difference might improve its predictive capability.⁶

In an impressive undertaking they analysed the data of 53 537 patients presenting with ACS without initial CS. A model was developed and validated internally in the Acute Myocardial Infarction in Switzerland Plus study (AMIS-Plus, $n = 35\,650$, 24% females) and externally in the Special Programme University Medicine Acute Coronary Syndrome (SPUM-ACS, $n = 4186$, 20% females) and the observatoire des Infarctus de Côte-d'Or (RICO, $n = 13\,701$, 26% females) thus formally allowing high-quality validation of a newly derived score. The RICO cohort already served as the external validation cohort for the initial ORBI score. The authors' first finding was that the ORBI score had a lower AUC for the prediction of CS in females (AUC 0.78, 95% CI 0.76 to 0.81) compared with males (AUC 0.81, 95% CI 0.79 to 0.83), with a significant $P = .048$, both AUCs staying within the predictive CIs of the initial ORBI validation ascertaining its validity. The authors adjusted the ORBI model subsequently by identifying potential variables that might have varying importance according to patients' sex. For this the cohort was split according to sex and three different statistical methods were applied to identify an optimized risk prediction for CS onset—logistic regression, a model that models the odds of an event in a linear way; random forest, a decision tree method; and a multilayer perception model, a neural network modelling method. These models were used to identify optimal prediction variables and then the model with optimal prediction capability was chosen for further analysis.

In comparison to the ORBI score the variables prior stroke/transient ischaemic attack, anterior ST-segment elevation myocardial infarction, first medical contact-to-PCI delay, and Killip-class II were replaced by creatinine, C-reactive protein (CRP), left ventricular ejection fraction (LVEF), and ST-segment elevation. It is interesting to note how the different modelling approaches weighted the relevance of variables. Irrespective of sex the two most important features in the logistic regression were LVEF and creatinine. This is not surprising as the logistic regression model works in a linear fashion where there is direct proportionality between dependent and independent variables—because the occurrence of CS in the absence of reduced LVEF is highly unlikely, this becomes the most important variable.⁷ On the contrary, the two machine learning models put more weight on variables like CRP, which were almost neglected by logistic regression ranging at rank 10 for females and rank 7 for males. It is important to keep in mind that machine learning models have a different way of handling data in comparison to conventional statistical models and can even draw information from missing variables or their interference with different variables, making their interpretation challenging.⁸ Given a large number of missing variables for CRP (32%) the model might have learned that missing CRP is of some prognostic utility in the prediction of CS, one possible explanation being that CRP was only measured in the sickest patients. Therefore, the mere presence of a

CRP might predict CS irrespective of any linear CRP increase or whether CRP was not measurable in patients who died beforehand. It is interesting to consider the theoretical patterns that machine learning might derive from the data, but it is also important to keep in mind that this might come at the cost of an inherent bias not visible at first sight.

After the identification of variables, the logistic regression approach was identified as the best way to model the new SEX-SHOCK score. The SEX-SHOCK score showed a good performance in internal validation (AUC for females 0.81, 95% CI 0.78 to 0.83; for males 0.83, 95% CI 0.82 to 0.85), but notably there was still an overlap within the 95% CI with initial validation of the ORBI score. On external validation the score confirmed its good predictive performance and showed a higher predictive capability in males compared with females in the RICO cohort (AUC for females 0.82, 95% CI 0.79 to 0.85; for males 0.88, 95% CI 0.86 to 0.89) but showed no improved sex-specific performance in the SPUM-ACS cohort (AUC for females 0.83, 95% CI 0.77 to 0.90; for males 0.83, 95% CI 0.80 to 0.87). Notably, from all AUCs calculated only one—a male patient in the RICO cohort—showed point estimates of the AUC that were not included in the CI of the original validation of the ORBI score (*Graphical Abstract*).

Did the authors succeed in their endeavour to identify a sex-specific score for the prediction of CS in patients with ACS? Well, there are two sides to this question, one statistical and the other clinical. From a statistical point of view the SEX-SHOCK score showed better results compared with the ORBI score—in one out of six analyses—in five out of six analyses the observed performance was within the expected range of the ORBI score, irrespective of sex. In line with this, it is important to mention that the authors state that they have used the same number of variables as the ORBI score—12 in total. Yet, this is not entirely true. While the SEX-SHOCK model does indeed only use 12 variables, there are actually two SEX-SHOCK models that account for the 13th variable—sex—which modifies the weighting of the other variables allowing for a smoother risk prediction fit. From a clinical point of view, even an AUC of 0.77, considering the worst lower CI, can be considered as a good predictive model. However, this comes at the cost of score calculation, which is time- and resource-demanding. In fact, there are more than 250 000 clinical risk prediction scores published of which little is known to most physicians and even less applied in clinical practice.⁹ Whether a score is worthwhile to calculate and apply in clinical practice depends on many factors, including but not limited to predictive performance. While the authors present impressive AUCs, which certainly suggest clinical relevance, these alone do not justify the score's application. It is also essential to weigh the implications of predicting an event against the potential consequences of acting on that prediction. In other words, the decision to use a predictive score should consider both the accuracy of the prediction and the impact of any interventions based on that prediction. This is counterbalanced by the time that any score calculation takes; the authors provide additional help by implementing an online calculation tool. One important aspect, which may be of greater significance to the medical community, is not yet another score, but the information on which patients are prone to develop CS. With the use of a logistic regression model the authors make interpretation of variables easy due to their direct proportionality and even in the absence of score calculation clinicians can use this knowledge to estimate the risk of CS (*Graphical Abstract*).

As clinicians, we often find ourselves in a position not unlike the tale of the Emperor's new clothes. In Hans Christian Andersen's story, the Emperor parades through the streets in what he believes to be a magnificent outfit, only to be revealed as wearing nothing at all when a child claims the truth. In the same way, the development of predictive models,

such as the SEX-SHOCK score, can sometimes lead us to place undue confidence in the complex and impressive algorithms they employ.

However, just as the Emperor's clothes were an illusion, the true value of any clinical score lies not in its apparent sophistication or statistical impressiveness, but in its real-world applicability and impact on patient outcomes. The SEX-SHOCK score, despite its statistical rigour and the promise of sex-specific predictions, ultimately reminds us of the limitations inherent in any model. A score that seems powerful may not always translate into meaningful improvements in clinical practice, beyond clinical judgement.

As we strive to advance our tools and techniques, we must remain vigilant, always questioning whether the elegance of a model is matched by its utility. Just as the Emperor's subjects learned to see through the illusion, we too must ensure that our clinical decisions are guided by clear-sighted assessment, balancing the potential benefits of prediction against the practicalities of application. In the end, the true measure of a score's worth lies not in the complexity of its construction, but in its capacity to genuinely improve patient care.

Declarations

Disclosure of Interest

K.-P.K. is a consultant to Edwards Lifesciences and Recor Medical. M.R. has none for this work. H.T. has none.

References

1. Thiele H, Zeymer U, Akin I, Behnes M, Rassaf T, Mahabadi AA, et al. Extracorporeal life support in infarct-related cardiogenic shock. *N Engl J Med* 2023;**389**:1286–97. <https://doi.org/10.1056/NEJMoa2307227>
2. Møller JE, Engstrøm T, Jensen LO, Eiskjær H, Mangner N, Polzin A, et al. Microaxial flow pump or standard care in infarct-related cardiogenic shock. *N Engl J Med* 2024;**390**:1382–93. <https://doi.org/10.1056/NEJMoa2312572>
3. Thiele H, Ohman EM, de Waha-Thiele S, Zeymer U, Desch S. Management of cardiogenic shock complicating myocardial infarction: an update 2019. *Eur Heart J* 2019;**40**:2671–83. <https://doi.org/10.1093/eurheartj/ehz363>
4. Burgos LM, Baro Vila RC, Botto F, Diez M. SCAI cardiogenic shock classification for predicting in-hospital and long-term mortality in acute heart failure. *J Soc Cardiovasc Angiogr Interv* 2022;**1**:100496. <https://doi.org/10.1016/j.jscai.2022.100496>
5. Auffret V, Cottin Y, Leurent G, Gilard M, Beer J-C, Zabalawi A, et al. Predicting the development of in-hospital cardiogenic shock in patients with ST-segment elevation myocardial infarction treated by primary percutaneous coronary intervention: the ORBI risk score. *Eur Heart J* 2018;**39**:2090–102. <https://doi.org/10.1093/eurheartj/ehy127>
6. Wang Y, Zeller M, Auffret V, Georgiopoulos G, Räber L, Roffi M, et al. Sex-specific prediction of cardiogenic shock after acute coronary syndromes: the SEX-SHOCK score. *Eur Heart J* 2024;**45**:4564–78. <https://doi.org/10.1093/eurheartj/ehae593>
7. Reynolds HR, Hochman JS. Cardiogenic shock: current concepts and improving outcomes. *Circulation* 2008;**117**:686–97. <https://doi.org/10.1161/CIRCULATIONAHA.106.613596>
8. Kresoja K-P, Unterhuber M, Wachter R, Thiele H, Lurz P. A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction. *Basic Res Cardiol* 2023;**118**:10. <https://doi.org/10.1007/s00395-023-00982-7>
9. Dambha-Miller H, Everitt H, Little P. Clinical scores in primary care. *Br J Gen Pract* 2020;**70**:163. <https://doi.org/10.3399/bjgp20X708941>