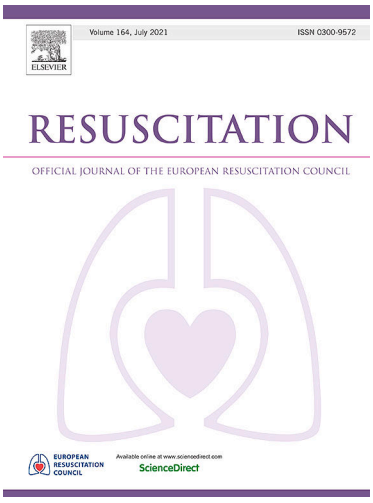
Journal Pre-proofs



Clinical paper

Comparison of four clinical risk scores in comatose patients after out-of-hospital cardiac arrest

Simon Schmidbauer, Christian Rylander, Alain Cariou, Matt P. Wise, Matthew Thomas, Thomas R. Keeble, David Erlinge, Matthias Haenggi, Pedro D. Wendel-Garcia, Jan Bělohlávek, Anders Morten Grejs, Niklas Nielsen, Hans Friberg, Josef Dankiewicz

PII: S0300-9572(23)00263-0
DOI: https://doi.org/10.1016/j.resuscitation.2023.109949
Reference: RESUS 109949

To appear in: *Resuscitation*

Received Date: 31 May 2023
Revised Date: 14 August 2023
Accepted Date: 16 August 2023

Please cite this article as: S. Schmidbauer, C. Rylander, A. Cariou, M.P. Wise, M. Thomas, T.R. Keeble, D. Erlinge, M. Haenggi, P.D. Wendel-Garcia, J. Bělohlávek, A. Morten Grejs, N. Nielsen, H. Friberg, J. Dankiewicz, Comparison of four clinical risk scores in comatose patients after out-of-hospital cardiac arrest, *Resuscitation* (2023), doi: https://doi.org/10.1016/j.resuscitation.2023.109949

COMPARISON OF FOUR CLINICAL RISK SCORES IN COMATOSE PATIENTS AFTER OUT-OF-HOSPITAL CARDIAC ARREST

Simon Schmidbauer[1*], Christian Rylander[2], Alain Cariou[3,4], Matt P. Wise[5], Matthew Thomas[6], Thomas R. Keeble[7,8], David Erlinge[9], Matthias Haenggi[10], Pedro D. Wendel-Garcia[11], Jan Bělohlávek[12], Anders Morten Grejs[13], Niklas Nielsen[14], Hans Friberg[1], Josef Dankiewicz[9]

[1]Department of Clinical Sciences, Anaesthesia and Intensive Care, Lund University, Skåne University Hospital, Malmö, Sweden.
[2]Anaesthesia and Intensive Care, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden
[3]Cochin University Hospital (APHP), Paris, France
[4]University Paris Cité (Medical School), Paris, France
[5]Adult Critical Care, University Hospital of Wales, Cardiff, UK
[6]Department of Intensive Care, University Hospitals Bristol and Weston, Bristol, UK
[7]Essex Cardiothoracic Centre, MSE, Basildon, Essex, United Kingdom
[8]MTRC, Anglia Ruskin School of Medicine, Chelmsford, Essex, United Kingdom
[9]Department of Clinical Sciences, Cardiology, Lund University, Skåne University Hospital, Lund, Sweden
[10]Department of Intensive Care Medicine, Bern University Hospital, University of Bern, Bern, Switzerland
[11]Institute of Intensive Care Medicine, University Hospital Zurich, Zurich, Switzerland
[12]2nd Department of Medicine – Department of Cardiovascular Medicine, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic
[13]Department of Intensive Care Medicine and Department of Clinical Medicine, Aarhus University Hospital and Aarhus University, Aarhus, Denmark
[14]Department of Clinical Sciences, Anaesthesia and Intensive Care, Lund University, Helsingborg Hospital, Helsingborg, Sweden

*Corresponding author (e-mail: simon.schmidbauer@med.lu.se)

ABSTRACT

BACKGROUND AND AIMS

Several different scoring systems for early risk stratification after out-of-hospital cardiac arrest have been developed, but few have been validated in large datasets. The aim of the present study was to compare the well-validated Out-of-hospital Cardiac Arrest (OHCA) and Cardiac Arrest Hospital Prognosis (CAHP)-scores to the less complex MIRACLE2- and Target Temperature Management (TTM)-scores.

METHODS

This was a post-hoc analysis of the Targeted Hypothermia versus Targeted Normothermia after Out-of-Hospital Cardiac Arrest (TTM2) trial. Missing data were handled by multiple imputation. The primary outcome was discriminatory performance assessed as the area under the receiver operating characteristics-

curve (AUROC), with the outcome of interest being poor functional outcome or death (modified Rankin Scale 4-6) at 6 months after OHCA.

RESULTS

Data on functional outcome at 6 months were available for 1829 cases, which constituted the study population. The pooled AUROC for the MIRACLE2-score was 0.810 (95% CI 0.790 - 0.828), 0.835 (95% CI 0.816 - 0.852) for the TTM-score, 0.820 (95% CI 0.800 - 0.839) for the CAHP-score and 0.770 (95% CI 0.748 - 0.791) for the OHCA-score. At the cut-offs needed to achieve specificities >95%, sensitivities were <40 % for all four scoring systems.

CONCLUSIONS

The TTM-, MIRACLE2- and CAHP-scores are all capable of providing objective risk estimates accurate enough to be used as part of a holistic patient assessment after OHCA of a suspected cardiac origin. Due to its simplicity, the MIRACLE2-score could be a practical solution for both clinical application and risk stratification within trials.

INTRODUCTION

While neurological prognostication in patients who remain unconscious after out-of-hospital cardiac arrest (OHCA) is not recommended until 72 hours after arrest or later,[1,2] clinicians might face situations where an early assessment of the risk of a poor outcome can be of value. Examples include decisions on escalating care, assessing potential benefit of more advanced therapies, communicating about prognosis with relatives, and stratification in clinical trials.

Several risk factors for poor outcome following OHCA are readily available in the early phase after return of spontaneous circulation (ROSC).[3] A number of risk scores with various combinations of these and additional data have been developed but only a few have been externally validated more than once.[4]

The Cardiac Arrest Hospital Prognosis (CAHP)[5] and Out of Hospital Cardiac Arrest (OHCA)-scores[6] stand out in this aspect with multiple external validations.[4] Neither of these scores can be calculated by hand and might therefore be of limited use in clinical situations. Additionally, with several early predictors being subject to varying degrees of uncertainty,[7,8] the use of complex scoring systems could conceivably mask errors caused by inaccurate input data.

We identified two summation-only scores with promising performance in their development cohorts: The MIRACLE2-score[9] by Pareek et al. and the TTM-score[10] developed in the Targeted Temperature Management-cohort.[11] The original MIRACLE2-publication included two external validation cohorts, and the TTM-score has been validated both within the MIRACLE2-study and in a Korean registry.[12]

In this study we sought to evaluate the usefulness of these two scores from a clinical standpoint with data from the TTM2-trial[13] and to compare them to the well-established OHCA- and CAHP-scores.

METHODS

OUTCOME

The primary outcome was discriminatory performance assessed as the area under the receiver operating characteristics-curve (AUROC). The outcome of interest was a poor functional outcome or death at 6 months after OHCA, assessed by a trained outcome assessor using the modified Rankin Scale (mRS) and defined as a score of 4-6 (moderately severe disability, severe disability or death). Details on outcome assessment are available in the trial protocol for outcome reporting and follow-up.[14]

STUDY POPULATION

This was a *post-hoc* analysis of the Targeted Hypothermia versus Targeted Normothermia after Out-of-Hospital Cardiac Arrest (TTM2) trial – an international, multicentre trial in which 1861 adult patients resuscitated after OHCA of presumed cardiac or unknown cause were randomised to either mild hypothermia or controlled normothermia.[13] The main exclusion criteria were an interval from return of spontaneous circulation to screening of more than 180 minutes, unwitnessed cardiac arrest with asystole as the initial rhythm, and limitations in care. All trial participants should be actively treated for at least 96 hours, after which withdrawal of life supporting therapies (WLST) was allowed only in cases fulfilling the TTM2-criteria for a likely poor neurological prognosis.[14] The main outcome, all-cause mortality at 6 months after arrest, did not differ between temperature groups. Functional outcome at 6 months was also similar. All participants in the intention-to-treat population were pooled to form the study population of the present study. Further details on the TTM2-trial are available in the original publication and its protocol.[13,15]

DATA

The TTM2-trial collected background, prehospital-, admission-, ICU-level and follow-up data through an electronic case report form.

Prehospital- and admission data were registered in accordance with the Utstein template.[16] Blood gas analyses, pH, glucose and lactate were registered from the first arterial blood gas after ROSC, whereas the first available value was recorded for creatinine. Neurological assessment at hospital admission was performed and recorded using the components (eye responses, motor responses, brainstem reflexes, and respiration pattern) of the Full Outline of UnResponsiveness (FOUR)-score, which allows assessment of the level of consciousness in intubated patients.[17]

All risk scores assessed in the present study except for MIRACLE2 include both the time from collapse to start of any cardiopulmonary resuscitation (the "no-flow"-time) and the time from start of resuscitation to ROSC (the "low-flow" time). The TTM2 trial, however, recorded the time to advanced life support (ALS) and time to ROSC as per the latest update of the Utstein-template.[16] Similarly, data on rhythm evolution over the course of resuscitation were not recorded.

This necessitated assumptions for variables that could not be derived from available data, i.e. no-flow/low-flow distinction for cases with bystander CPR and the MIRACLE2 item "changing intra-arrest rhythms". For cases with bystander CPR, a no-flow time of 1 minute (median time to BLS among patients receiving bystander CPR in both the TTM-trial[11] and other predominantly European studies[18–20]) was assumed. Participants with a non-shockable initial rhythm who were defibrillated were assumed to have changing intra-arrest rhythms. Details on data assumptions are outlined in table S2.

STATISTICAL ANALYSES

Score variables were presented as absolute and relative frequency, median and inter-quartile range (IQR) or mean and standard deviation (SD). Their association with a poor outcome was presented as odds ratios (OR) rather than relative risks to preserve comparability to the respective development studies.

Missing data was assumed to be missing at random (MAR) and handled using mice: Multivariate Imputation by Chained Equations in R,[21] with all available background, prehospital, admission-level and outcome variables of the TTM2-dataset considered for inclusion in the multiple imputation process. Details on variable selection, multiple imputation and diagnostics thereof are available in the supplementary appendix. A total of 50 datasets including all variables of all evaluated scoring systems were imputed. The handling and reporting of multiple imputation in this manuscript adheres to guidelines by Sterne et al.[22]

Scores were estimated in each of the imputed datasets using the criteria defined in the original publications of the OHCA-, TTM-[10] and MIRACLE2[9]-scores. For the CAHP-score, which in its original publication was presented only as a nomogram, parameters for estimation were retrieved from a later publication.[23] Score items are presented in relation to the evaluated scoring systems in figure 1, and details on their estimations are available in figure S1.

Overall discriminatory performance was evaluated using ROC-curves, with the area under the ROC-curve (AUROC) being estimated for each of the imputed datasets and then pooled (after log-transformation) using Rubin's rules. Diagnostic measures of sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) at the proposed cut-offs of each score were also calculated in all imputed datasets, with results presented as the between-imputations observed minimum-maximum value of each measure. The first data point yielding a specificity above 95 % was identified with the datasets ordered after each respective score. Pooled confidence intervals of these estimates are available in the supplementary appendix (table S3).

The full regression models have not been made publicly available for any of the evaluated scoring systems. With clinical applicability nonetheless being the aim of this study, no attempt to collect original data from the respective authors was made and only the scoring systems themselves were evaluated. As a result, calibration could only be evaluated by indirect measures. The relationship between score and the observed outcome was assessed graphically as the proportion of a poor outcome per level (for discrete scores) or interval (for continuous scores) of the score. Intervals of continuous scores were chosen to harmonise with the cut-off values proposed in the original publications.

All analyses were performed with the use of R: A Language and Environment for Statistical Computing, version 4.1.2.[24]

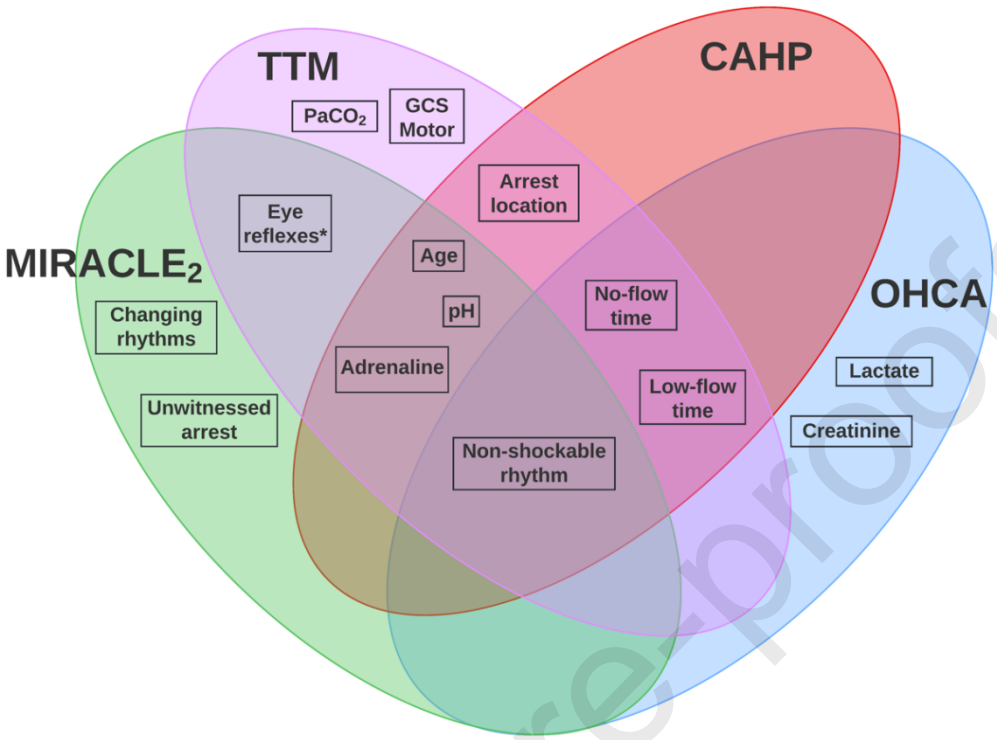


*Figure 1.Venn diagram of items included in each score. GCS = Glasgow coma scale. *Eye reflexes is a composite of "no pupillary or corneal reflexes" (TTM-score) and "no pupillary reflexes" (MIRACLE2).*

RESULTS

Of the 1861 participants included in the intention to treat-population of the TTM2 trial, functional outcome data at 6 months after cardiac arrest were available for 1829 (98.3 %), which constituted the study population. The mean age was 63.9 (SD 13.6) years, 1447 (79.1 %) patients were male and 988 (54.0 %) had a poor outcome (mRS 4-6) at 6 months.

Data needed for score estimations were available for > 95 % cases with the exception of information on corneal reflexes at admission (valid n = 1031, 56.4 %) and pupillary reflexes at admission (valid n = 1510, 82.6 %). All variables were significantly correlated with outcome except for unwitnessed arrest (OR 1.30, 95 % CI 0.93 – 1.82 for a poor outcome) and PaCO2 < 4.5 kPa at admission (OR 1.34, 95 % CI 0.90 – 2.02 for a poor outcome). A complete presentation of all score variables and their association with neurological outcome at 6 months post-arrest is available in table 1. The single strongest predictors in univariable analysis were age, initial rhythm, adrenaline administration and pupillary reflexes with ORs for poor outcome being 8.84 (95 % CI 5.81 – 13.9) for age >80 years, 6.54 (95 % CI 5.05 – 8.55) for non-shockable rhythm, 6.12 (95 % CI 4.91 – 7.66) for any adrenaline dose and 3.32 (95 % CI 2.61 – 4.23) for bilaterally absent pupillary reflexes, respectively.

*Table 1. Score items and their association with a poor outcome (n = 1829). Data presented as absolute (relative) frequency or median [IQR]. N denotes number of patients with valid data. CI = Confidence interval. IQR = Interquartile range. Ref. denotes reference category. * Per 10 unit change in creatinine.*

	N	Frequency/ distribution	OR [95% CI]	p
Age (years)	1829			
≤ 60		663 (36.2%)	Ref.	Ref.
61 - 80		992 (54.2%)	2.58 [2.11 - 3.16]	<0.001
> 80		174 (9.5%)	8.84 [5.81 - 13.9]	<0.001
Cardiac arrest at home	1829	968 (52.9%)	2.18 [1.80 - 2.63]	<0.001
Unwitnessed arrest	1829	157 (8.6%)	1.30 [0.93 - 1.82]	0.124
Non-shockable initial rhythm	1829	483 (26.4%)	6.54 [5.05 - 8.55]	<0.001
Changing intra-arrest rhythms	1814	161 (8.9%)	3.27 [2.25 - 4.86]	<0.001
Any adrenaline administration	1829	1247 (68.2%)	6.12 [4.91 - 7.66]	<0.001
No-flow time (min)	1829	1.00 [1.00 - 1.00]	1.08 [1.05 - 1.11]	<0.001
Low-flow time (min)	1829	24.0 [14.0 - 37.0]	1.03 [1.02 - 1.03]	<0.001
GCS motor 1 at admission	1667	1437 (86.2%)	2.77 [2.07 - 3.73]	<0.001
No corneal reflex at admission	1031	666 (64.6%)	2.66 [2.05 - 3.47]	<0.001
No pupillary reflex at admission	1510	453 (30.0%)	3.32 [2.61 - 4.23]	<0.001
Creatinine at admission (µmol/L)	1813	105 [87.0 - 127]	1.06 [1.03 - 1.08]*	<0.001
pH at admission	1796			
< 6.90		103 (5.7%)	6.54 [3.78 - 11.7]	<0.001
6.90 - 7.04		213 (11.9%)	5.13 [3.34 - 7.97]	<0.001
7.05 - 7.19		478 (26.6%)	2.99 [2.10 - 4.30]	<0.001
7.20 - 7.34		823 (45.8%)	1.44 [1.03 - 2.02]	0.032
> 7.35		179 (9.97%)	Ref.	Ref.
Lactate at admission (mmol/L)	1751	4.90 [2.60 - 8.10]	1.17 [1.14 - 1.20]	<0.001
PaCO2 < 4.5 kPa at admission	1798	107 (5.6%)	1.34 [0.90 - 2.02]	0.147

Overall discriminatory performance of each score for a poor outcome at 6 months is presented in figure 2 and table 2, with the TTM-score having the largest area under the ROC-curve of 0.835 (95 % CI 0.816 – 0.852). The 95 % confidence intervals of all AUROC-estimates overlapped each other, except for that of the OHCA-score (0.770, 95% CI 0.748 – 0.791), which was lower than those of both the TTM- and the CAHP-scores (0.820, 95 % CI 0.800 – 0.839 for the latter).

Table 2. Area under the ROC-curve. Pooled estimates from 50 imputations. AUROC = Area under the receiver operating characteristics curve. CI = Confidence interval.

	AUROC (95% CI)
CAHP-score	0.820 (0.800 - 0.839)
MIRACLE2-score	0.810 (0.790 - 0.828)
OHCA-score	0.770 (0.748 - 0.791)
TTM-score	0.835 (0.816 - 0.852)

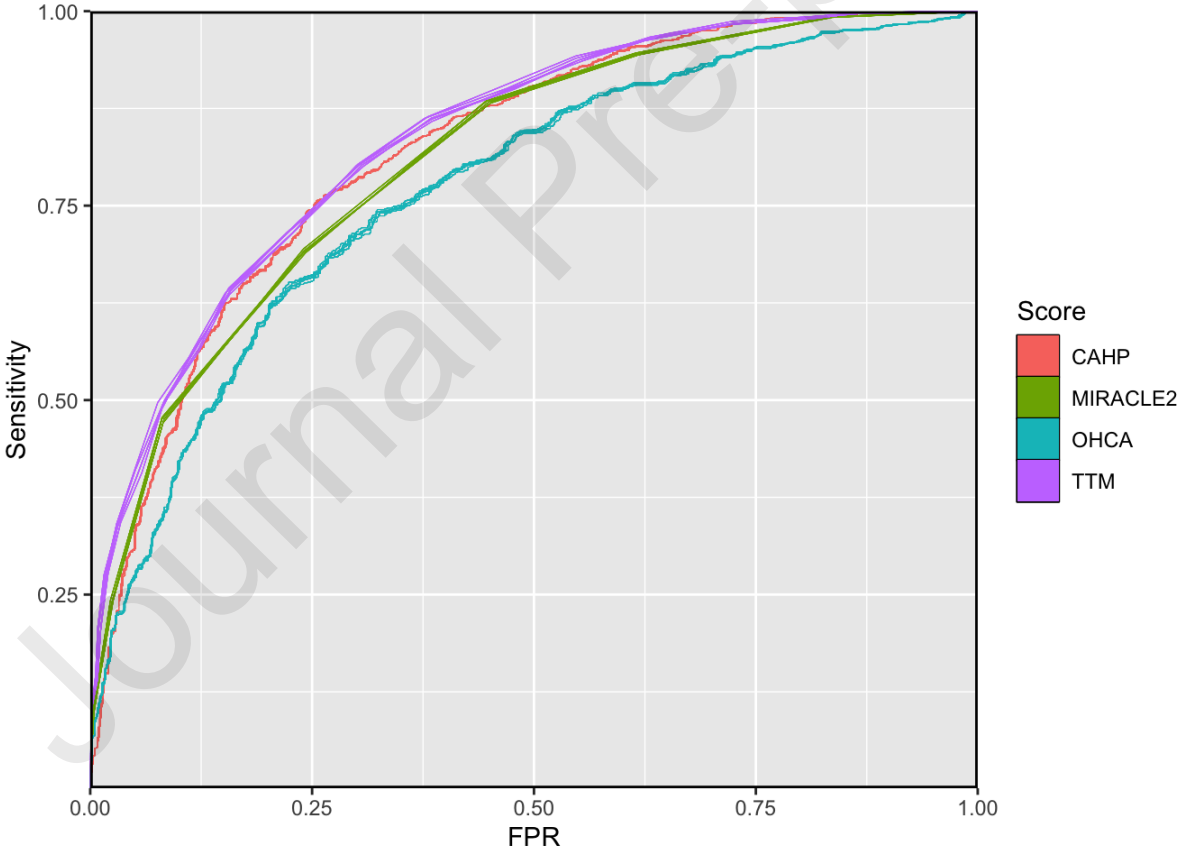


Figure 2. Receiver-operating characteristics curves for the respective scores. One line per imputation (n=50) per score. The resulting line thickness thus reflects imputation variance. FPR = False positive rate (1 - specificity).

Sensitivities, specificities, negative predictive values and positive predictive values at the cut-offs proposed in the original publications, as well as the cut-off needed to achieve > 95 % specificity in the present study population, are presented for all scores in table 3 with intervals representing the between-imputations minimum-maximum values. Additionally, these measures are reported with their corresponding pooled 95%

CIs in table S3. The high-risk categories of the TTM- and MIRACLE2-scores had similar diagnostic properties with both categorizing roughly a third of all patients as high-risk (30.1 – 30.8 % for the TTM-score, 29.2 – 29.5 % for MIRACLE2) in which the specificity for a poor outcome was similar for both scores (91.7 – 92.4 for the TTM-score and 91.7 – 91.9 for MIRACLE2). In contrast, the CAHP-score identified a smaller group comprising 15.5 – 15.7 % of all patients with a specificity for a poor outcome of 96.4 – 96.6 % at its highest cut-off of 200. At the lowest possible cut-off needed to achieve > 95 % specificity, sensitivities were 33.7 – 41.1 % for the TTM-score, 23.8 – 24.4 % for the MIRACLE2-score, 27.1 – 27.5 % for the OHCA-score and 30.9 – 33.3 % for the CAHP-score.

Table 3. Diagnostic performance by risk group. Point estimates of discriminatory performance by risk group. Cut-offs according to the original publication of each score. Only two risk groups were defined in the CAHP- and MIRACLE2-publications. Number of patients (n) and estimates are from 50 imputed datasets, with values representing min-max between imputations. NPV = Negative predictive value, PPV = Positive predictive value. A specificity of >95% corresponds to a false-positive rate (FPR) of <5%.

Risk group	**Intermediate**	**Intermediate-high**	**High**	**At >95% Specificity**
TTM-score	**>10**	**>13**	**>16**	**>18-19**
N (%)	1283 (70.1) - 1291 (70.6)	890 (48.7) - 900 (49.2)	551 (30.1) - 563 (30.8)	359 (19.6) - 447 (24.4)
Sensitivity (%)	89.8 - 90.4	71.4 - 71.9	48.9 - 49.9	33.7 - 41.0
Specificity (%)	52.2 - 53.5	77.2 - 78.0	91.7 - 92.4	95.0 - 96.9
NPV (%)	78.7 - 79.5	66.3 - 66.9	57.4 - 57.8	53.5 - 55.6
PPV (%)	68.9 - 69.5	78.6 - 79.2	87.5 - 88.5	90.6 - 92.4
MIRACLE2-score	**>2**	**N/A**	**>4**	**>5**
N (%)	1249 (68.3) - 1254 (68.6)		534 (29.2) - 540 (29.5)	255 (13.9) - 263 (14.4)
Sensitivity (%)	88.2 - 88.5		47.1 - 47.8	23.8 - 24.4
Specificity (%)	54.7 - 55.4		91.7 - 91.9	97.4 - 97.7

NPV (%)	67.5 - 67.9		52.2 - 52.3	48.5 - 48.7
PPV (%)	69.6 - 70.0		87.0 - 87.4	91.6 - 92.6
OHCA-score	**>2.0**	**>17.4**	**>32.5**	**>38.2-38.3**
N (%)	1607 (87.9) - 1613 (88.2)	1075 (58.8) - 1083 (59.2)	488 (26.7) - 497 (27.2)	310 (16.9) - 314 (17.2)
Sensitivity (%)	95.7 - 96.0	76.4 - 77.2	41.0 - 41.7	27.1 - 27.5
Specificity (%)	20.9 - 21.4	61.8 - 62.1	89.9 - 90.1	95.0 - 95.0
NPV (%)	81.2 - 81.7	69.1 - 69.9	56.5 - 56.7	52.6 - 52.7
PPV (%)	58.8 - 58.9	70.2 - 70.5	82.9 - 83.1	86.5 - 86.6
CAHP-score	**>150**	**N/A**	**>200**	**>191-195**
N (%)	910 (49.8) - 914 (50.0)		284 (15.5) - 287 (15.7)	347 (19.0) - 371 (20.3)
Sensitivity (%)	71.9 - 72.2		25.8 - 26.1	30.9 - 33.3
Specificity (%)	76.1 - 76.2		96.4 - 96.6	95.0 - 95.0
NPV (%)	69.8 - 70.0		52.5 - 52.6	53.9 - 54.8
PPV (%)	78.0 - 78.0		89.5 - 89.9	87.9 - 88.7

Sensitivity analyses restricted to patients with valid data only, a retrospectively identified non-cardiac cause of arrest, with no-flow times omitted and dichotomisations by sex and temperature target did not reveal any significant differences from the main results (figures S3-9).

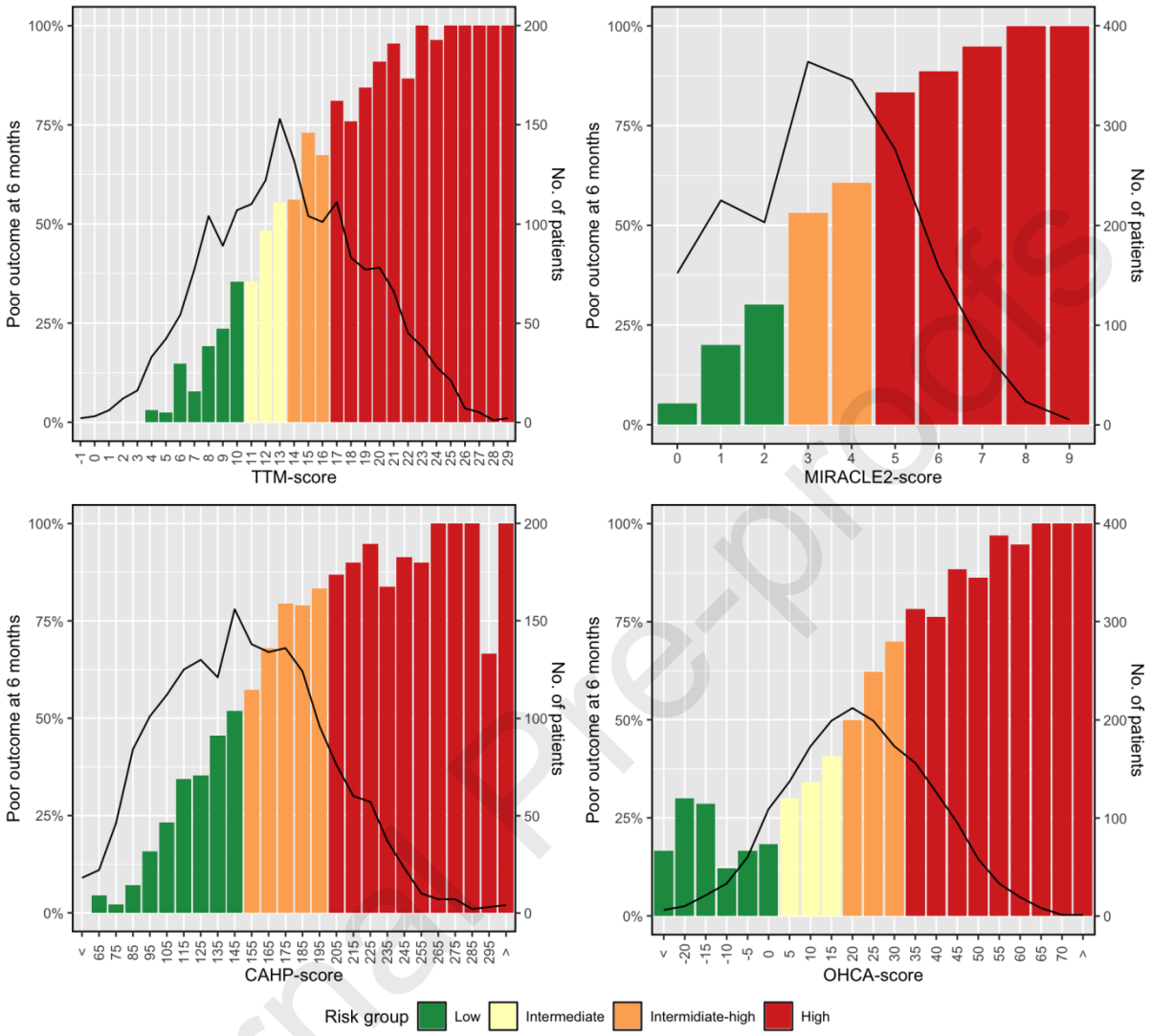


Figure 3. Score vs. observed outcome. Data from first imputed dataset. Black line shows number of patients per bar (right Y-axis). The continuous CAHP- and OHCA-scores have been categorised using equal-width binning with numbers on X-axis denoting the midpoint of each bin. The boundaries of the first and last bins (marked as "<" and ">") are open to negative and positive infinity, respectively.

DISCUSSION

In this study of different risk scores for early outcome prediction after OHCA applied to the large international TTM2 trial we demonstrated that differences between well-validated scoring systems and newer alternatives were minor. The TTM-score showed a slightly higher AUROC point estimate than the others, but only the OHCA-score showed a meaningful difference in its 95 % confidence interval not overlapping those of the TTM- and CAHP-scores.

As per the TRIPOD-statement,[25] both development and validation studies are recommended to report calibration of its prediction models. For reasons outlined in the methods section, traditional direct approaches were not possible in the present study. Perhaps the most useful surrogate of overall calibration, however, is the distribution of patients across the different risk categories defined for each respective score in comparison with their development cohorts. The OHCA-score and to some extent also the CAHP-score tended to be overly specific with correspondingly low sensitivity levels in the present

study when compared to their respective development cohorts. This likely reflects differences in the overall event rate (prevalence of a poor outcome 54 % in the TTM2-trial vs. 79 % in the OHCA-[6] and 74 % in the CAHP development cohorts,[5] respectively) and could be improved by simply adjusting the cut-offs. In contrast, the MIRACLE2-score showed similar specificity but lower sensitivity across all cut-offs compared to its development cohort, as well as a lower AUROC-estimate in the present study (0.81 vs 0.90), indicating diminished discriminative performance rather than simple calibration issues arising from baseline risk differences.

A somewhat lower AUROC is to be expected upon external validation, but one possible factor contributing to this could be the large effect of adrenaline administration in both the CAPH- and MIRACLE2-scores (univariate OR for a poor outcome 15.40, 95% CI 8.79 - 28.25 in the MIRACLE2 development cohort), which was roughly double that seen in our cohort (univariate OR 6.13, 95% CI 4.92 - 7.67). The TTM-score also included adrenaline administration, but with a smaller effect size (age-adjusted OR 4.62, 95% CI 3.29 - 6.48) that was comparable to the findings in our cohort (age-adjusted OR 6.31, 95% CI 5.01 - 8.00). With no randomised trials having shown adrenaline administration to be deleterious in its own right,[26,27] the prognostic value of adrenaline administration evident in observational studies more likely reflects prolonged resuscitation attempts and an initial rhythm that is not immediately shockable. Differences between treatment algorithms among emergency services and in patient populations might, however, affect the size of the effect.

Despite these constraints, all evaluated scoring systems showed smooth, roughly linear or sigmoid-shaped relationships between score and the observed outcome (figure 3). The proportion of a poor outcome was, however, higher than expected for some of the lowest values of the OHCA-score.[6] Exploratory analyses of cases with an OHCA-score of less than 0 (data not shown) revealed no single reason for this discrepancy, but patients with a poor outcome despite low OHCA-scores often had low-normal levels of creatinine. Given the design of the OHCA-score, low creatinine-levels can have a very large effect on the final score due to the continuous nature of this item. Abnormally low creatinine levels can theoretically result in infinitely negative scores which might effectively outweigh the prognostic value of all other score items combined. This highlights the susceptibility to outlier bias that is inherent to risk prediction scores reliant on continuous data.

Whether these limitations present a real problem depends on the intended usage of the scoring systems. Any of the evaluated scores could be used as-is for relative risk stratification within trials, as the cut-offs in this scenario can be adjusted as needed. For clinical application, however, our findings highlight the importance of calibrating at least the cut-offs in the intended usage population in order to adjust for any differences in baseline outcome rates.

Concerns have been raised that early risk prediction might be perceived as a means to guide early withdrawal of life sustaining therapy (WLST) and thus risk inducing self-fulfilling nihilism.[28] Acceptable false-positive rates for guiding life-and-death decisions are probably impossible to define, but proposed levels range from 0.1 % to 5 %.[29,30] While our analyses suggest that such rates are achievable at high cut-offs for all evaluated scores, it is also clear that these cut-offs vary across different populations and might be affected by inaccurate input data. Delayed multimodal neuroprognostication at least 72 hours post-arrest must therefore remain the norm before WLST can be considered for patients who continue to be comatose after OHCA.

Data from the original TTM-trial, however, suggest that about 15 % of OHCA-patients might die earlier than the timepoint for delayed neuroprognostication – primarily due to non-cerebral causes.[31] Within this group of severely ill patients, advanced treatment options such as coronary revascularization, mechanical circulatory support and ECMO might be necessary to avoid imminent death. At the same time, it is critical to avoid exposing dying patients and their relatives to highly invasive and resource-intensive interventions when the prognosis is dismal. In these situations, risk scores could provide clinicians with an early and objective measure of baseline risk, that should be considered in a holistic assessment of an OHCA patient.

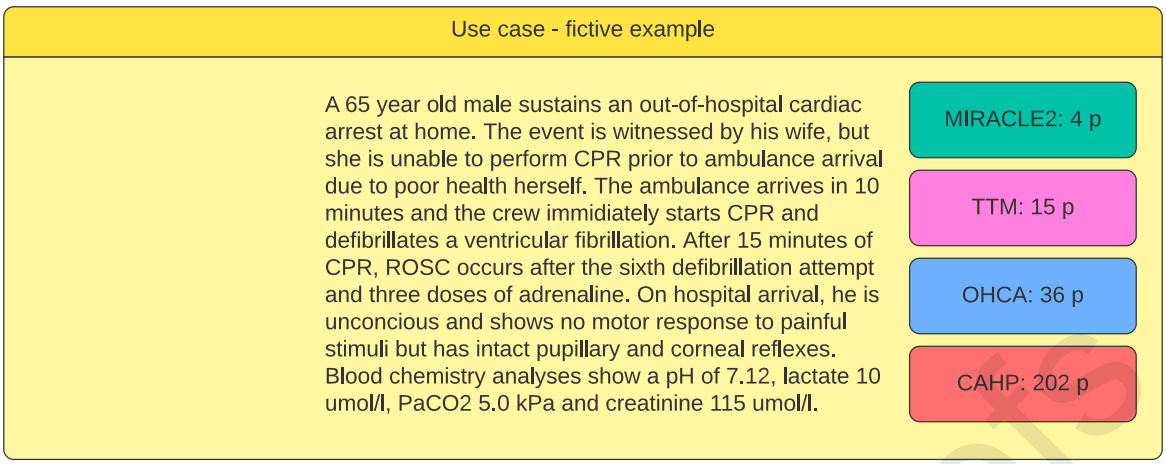


Figure 4. Application of all evaluated scoring systems on a fictive clinical scenario.

Our results add to those from several previous development and validation studies that outcome prediction utilising factors readily available at hospital admission is associated with similar degrees of uncertainty, almost regardless of the specific scoring system used.[4] This is likely a reflection of both inaccuracies in input data[7,8] and insufficient prognostic value of the predictors themselves.[32] With clinical judgement being subject to the very same constraints, we would nonetheless advocate that implementation of the simplest-to-use scoring system matching available data in a particular setting could improve objectivity in early outcome assessments, especially since a recent study indicated that clinician perception of prognosis might have a substantial impact on resuscitation outcomes that could plausibly extend into the post-resuscitation period.[33] Aside from ease-of-use, less complex scoring could also have an added effect on safety; the simpler the score, the easier it will be for the clinician to spot discrepancies that might lead to spurious results. In this aspect, the 7-item, 10-point MIRACLE2-score may offer an advantage over the other evaluated scores which either include a greater number of items (the TTM-score), requires a nomogram (the CAPH-score) or a calculator/web-based tool (the OHCA-score). Referring to the example case in figure 4, had the wife attempted CPR but it was unclear whether this was effective, a user of the TTM-score could easily judge the impact this would have on the final score (+/- 2 points in this case). For the CAHP- and OHCA-scores, where this item is on a continuous scale, such an estimation is arguably less intuitive. The MIRACLE2-score, on the other hand, has a strength in its omission of this often unclear interval altogether. A limited number of variables is also a clear benefit if a score were to be used for risk stratification within clinical trials on interventions targeting hypoxic-ischemic encephalopathy, where exclusion of patients with the highest probabilities of both a good and poor outcome, might increase power.

The main strength of the present study is the use of high-quality data from the international TTM2-trial. With its strict criteria for WLST, bias introduced from otherwise using the negative prognostic factors of the evaluated scoring systems to guide WLST is minimised. The slightly higher AUROC point estimate in this validation of the TTM-score may be attributable to the fact that the case-mix was very similar to its original development cohort. Furthermore, as the TTM2-trial included patients with OHCA of a presumed cardiac cause, excluding patients with an unwitnessed arrest and an initial rhythm of asystole, all of the present validation results are limited to a similar population.

The lack of a recorded time from arrest to BLS is a limitation, as all risk scores evaluated in the present study (except for MIRACLE2) included both the no-flow and low-flow time intervals. Nevertheless, it is also a realistic clinical scenario as both timing and quality of bystander CPR are difficult or impossible to estimate. Imputation of a low value for cases with bystander CPR, as done in the present study, retains the prognostic value of the more precise no-flow interval in cases not receiving bystander CPR while assuming a conservative "best-case"-scenario for patients receiving bystander-CPR. Completely omitting the no-flow interval has also been suggested and both the OHCA- and CAHP-scores have been validated without their

no-flow-item.[34] Another limitation is the lack of data on changing rhythms other than from non-shockable to shockable, although it is likely that this rhythm-change carries most of the prognostic value.[9,35]

Conclusion

The TTM-, MIRACLE2- and CAHP-scores are all capable of providing objective risk estimates accurate enough to be used as part of a holistic patient assessment after OHCA of a suspected cardiac origin. Due to its simplicity, the MIRACLE2-score could be a practical solution for both clinical application and risk stratification within trials.

Declarations

Ethical approval

The TTM2 trial was approved by the ethics committee in each participating country. Informed consent was obtained from all participants who regained mental capacity, and was otherwise waived, deferred or obtained from a legal surrogate.

Authors' contributions

SS, CR, HF, NN and JD conceptualised the study. SS, CR and JD wrote the first manuscript draft. SS and JD performed all data analyses. AC, MPW, MT, TRK, DE, MH, PDWG, JB, AMG, NN and HF provided input on the analyses and their interpretation, and revised the manuscript. All authors read and approved the final manuscript.

Funding

The TTM2-trial was supported by independent research grants from nonprofit or governmental agencies (the Swedish Research Council [Vetenskapsrådet], Swedish Heart–Lung Foundation, Stig and Ragna Gorthon Foundation, Knutsson Foundation, Laerdal Foundation, Hans-Gabriel and Alice Trolle-Wachtmeister foundation for medical Research, and regional research Support in Region Skåne) and by governmental funding of clinical research within the Swedish National Health Service. Authors S.S and H.F received regional research support from Region Skåne and funding from the Laerdal Foundation, respectively, for the present work.

Declaration of interest

Authors AC and HF are members of the editorial board of *Resuscitation*.

Data availability statement

The analyses described in this manuscript derive from data obtained as part of the TTM2-trial and we therefore refer to the data sharing statement of its original publication.[13]

Supplementary files

A supplementary appendix is available with the online version of this article.

REFERENCES

1. Nolan JP, Sandroni C, Böttiger BW, Cariou A, Cronberg T, Friberg H, et al. European Resuscitation Council and European Society of Intensive Care Medicine Guidelines 2021: Post-resuscitation care. Resuscitation. 2021 Apr;161:220–69.

2. Panchal AR, Bartos JA, Cabañas JG, Donnino MW, Drennan IR, Hirsch KG, et al. Part 3: Adult Basic and Advanced Life Support: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation [Internet]. 2020 Oct 20 [cited 2022 Jan 20];142(16_suppl_2). Available from: https://www.ahajournals.org/doi/10.1161/CIR.0000000000000916

3. Dyson K, Brown SP, May S, Smith K, Koster RW, Beesems SG, et al. International variation in survival after out-of-hospital cardiac arrest: A validation study of the Utstein template. Resuscitation. 2019 May 1;138:168–81.

4. Carrick RT, Park JG, McGinnes HL, Lundquist C, Brown KD, Janes WA, et al. Clinical Predictive Models of Sudden Cardiac Arrest: A Survey of the Current Science and Analysis of Model Performances. J Am Heart Assoc. 2020 Aug 18;9(16):e017625.

5. Maupain C, Bougouin W, Lamhaut L, Deye N, Diehl JL, Geri G, et al. The CAHP (Cardiac Arrest Hospital Prognosis) score: a tool for risk stratification after out-of-hospital cardiac arrest. Eur Heart J. 2016 Nov 7;37(42):3222–8.

6. Adrie C, Cariou A, Mourvillier B, Laurent I, Dabbane H, Hantala F, et al. Predicting survival with good neurological recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: the OHCA score. Eur Heart J. 2006 Dec 1;27(23):2840–5.

7. Frisch A, Reynolds JC, Condle J, Gruen D, Callaway CW. Documentation discrepancies of time-dependent critical events in out of hospital cardiac arrest. Resuscitation. 2014 Aug;85(8):1111–4.

8. Sundermann ML, Salcido DD, Koller AC, Menegazzi JJ. Inaccuracy of patient care reports for identification of critical resuscitation events during out-of-hospital cardiac arrest. Am J Emerg Med. 2015 Jan;33(1):95–9.

9. Pareek N, Kordis P, Beckley-Hoelscher N, Pimenta D, Kocjancic ST, Jazbec A, et al. A practical risk score for early prediction of neurological outcome after out-of-hospital cardiac arrest: MIRACLE2. Eur Heart J. 2020 Dec 14;41(47):4508–17.

10. Martinell L, Nielsen N, Herlitz J, Karlsson T, Horn J, Wise MP, et al. Early predictors of poor outcome after out-of-hospital cardiac arrest. Crit Care. 2017 Apr 13;21(1):96.

11. Nielsen N, Wetterslev J, Cronberg T, Erlinge D, Gasche Y, Hassager C, et al. Targeted Temperature Management at 33°C versus 36°C after Cardiac Arrest. N Engl J Med. 2013 Dec 5;369(23):2197–206.

12. Heo WY, Jung YH, Lee HY, Jeung KW, Lee BK, Youn CS, et al. External validation of cardiac arrest-specific prognostication scores developed for early prognosis estimation after out-of-hospital cardiac arrest in a Korean multicenter cohort. PLOS ONE. 2022 Apr 1;17(4):e0265275.

13. Dankiewicz J, Cronberg T, Lilja G, Jakobsen JC, Levin H, Ullén S, et al. Hypothermia versus Normothermia after Out-of-Hospital Cardiac Arrest. N Engl J Med. 2021 Jun 17;384(24):2283–94.

14. Lilja G, Nielsen N, Ullén S, Blennow Nordstrom E, Dankiewicz J, Friberg H, et al. Protocol for outcome reporting and follow-up in the Targeted Hypothermia versus Targeted Normothermia after Out-of-Hospital Cardiac Arrest trial (TTM2). Resuscitation. 2020 May;150:104–12.

15. Dankiewicz J, Cronberg T, Lilja G, Jakobsen JC, Bělohlávek J, Callaway C, et al. Targeted hypothermia versus targeted Normothermia after out-of-hospital cardiac arrest (TTM2): A randomized clinical trial—Rationale and design. Am Heart J. 2019 Nov;217:23–31.

16. Perkins GD, Jacobs IG, Nadkarni VM, Berg RA, Bhanji F, Biarent D, et al. Cardiac Arrest and Cardiopulmonary Resuscitation Outcome Reports: Update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest. Resuscitation. 2015 Nov 1;96:328–40.

17. Iyer VN, Mandrekar JN, Danielson RD, Zubkov AY, Elmer JL, Wijdicks EFM. Validity of the FOUR Score Coma Scale in the Medical Intensive Care Unit. Mayo Clin Proc. 2009 Aug;84(8):694–701.

18. Zeiner A, Holzer M, Sterz F, Schörkhuber W, Eisenburger P, Havel C, et al. Hyperthermia After Cardiac Arrest Is Associated With an Unfavorable Neurologic Outcome. Arch Intern Med. 2001 Sep 10;161(16):2007–12.

19. Kirkegaard H, Søreide E, de Haas I, Pettilä V, Taccone FS, Arus U, et al. Targeted Temperature Management for 48 vs 24 Hours and Neurologic Outcome After Out-of-Hospital Cardiac Arrest: A Randomized Clinical Trial. JAMA. 2017 Jul 25;318(4):341–50.

20. Rawshani A, Herlitz J. Svenska Hjärt-lungräddningsregistret årsrapport för år 2020 [Swedish CPR-registry annual report 2020] [Internet]. [cited 2022 Jun 13]. Available from: https://arsrapporter.registercentrum.se/shlr/20210929/

21. Buuren S van, Groothuis-Oudshoorn K. **mice** : Multivariate Imputation by Chained Equations in *R*. J Stat Softw [Internet]. 2011 [cited 2022 Sep 5];45(3). Available from: http://www.jstatsoft.org/v45/i03/

22. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Sep 1;338(jun29 1):b2393–b2393.

23. Bougouin W, Dumas F, Karam N, Maupain C, Marijon E, Lamhaut L, et al. Should We Perform an Immediate Coronary Angiogram in All Patients After Cardiac Arrest?: Insights From a Large French Registry. JACC Cardiovasc Interv. 2018 Feb 12;11(3):249–56.

24. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

25. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015;13(1):1.

26. Perkins GD, Ji C, Deakin CD, Quinn T, Nolan JP, Scomparin C, et al. A Randomized Trial of Epinephrine in Out-of-Hospital Cardiac Arrest. N Engl J Med. 2018 Aug 23;379(8):711–21.

27. Lin S, Callaway CW, Shah PS, Wagner JD, Beyene J, Ziegler CP, et al. Adrenaline for out-of-hospital cardiac arrest resuscitation: A systematic review and meta-analysis of randomized controlled trials. Resuscitation. 2014 Jun 1;85(6):732–40.

28. Williams A, Vlachos S, Maharaj R. Prognostication after out-of-hospital cardiac arrest: biases and caveats. Eur Heart J. 2021 Mar 14;42(11):1113–4.

29. Steinberg A, Callaway CW, Arnold RM, Cronberg T, Naito H, Dadon K, et al. Prognostication after cardiac arrest: Results of an international, multi-professional survey. Resuscitation. 2019 May;138:190–7.

30. Sandroni C, Cariou A, Cavallaro F, Cronberg T, Friberg H, Hoedemaekers C, et al. Prognostication in comatose survivors of cardiac arrest: An advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. Intensive Care Med. 2014 Dec;40(12):1816–31.

31. Dragancea I, Wise MP, Al-Subaie N, Cranshaw J, Friberg H, Glover G, et al. Protocol-driven neurological prognostication and withdrawal of life-sustaining therapy after cardiac arrest and targeted temperature management. Resuscitation. 2017 Aug;117:50–7.

32. Rea TD, Cook AJ, Stiell IG, Powell J, Bigham B, Callaway CW, et al. Predicting Survival After Out-of-Hospital Cardiac Arrest: Role of the Utstein Data Elements. Ann Emerg Med. 2010 Mar 1;55(3):249–57.

33. Laurenceau T, Marcou Q, Agostinucci JM, Martineau L, Metzger J, Nadiras P, et al. Quantifying physician's bias to terminate resuscitation. The TERMINATOR study. Resuscitation. 2023 Jul 1;188:109818.

34. Wang CH, Huang CH, Chang WT, Tsai MS, Yu PH, Wu YW, et al. Prognostic performance of simplified out-of-hospital cardiac arrest (OHCA) and cardiac arrest hospital prognosis (CAHP) scores in an East Asian population: A prospective cohort study. Resuscitation. 2019 Apr;137:133–9.

35. Meaney PA, Nadkarni VM, Kern KB, Indik JH, Halperin HR, Berg RA. Rhythms and outcomes of adult in-hospital cardiac arrest*: Crit Care Med. 2010 Jan;38(1):101–8.

CRediT AUTHOR STATEMENT

Manuscript title: COMPARISON OF FOUR CLINICAL RISK SCORES IN COMATOSE PATIENTS AFTER OUT-OF-HOSPITAL CARDIAC ARREST

Schmidbauer S: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Visualisation

Rylander C: Conceptualisation, Methodology, Investigation, Writing – Original Draft

Cariou A: Writing - Review & Editing

Wise MP: Writing - Review & Editing

Thomas M: Writing - Review & Editing

Keeble TR: Writing - Review & Editing

Erlinge D: Writing - Review & Editing

Haenggi M: Writing - Review & Editing

Wendel-Garcia PD: Writing - Review & Editing

Bělohlávek J: Writing - Review & Editing

Grejs AM: Writing - Review & Editing

Nielsen N: Conceptualisation, Writing - Review & Editing, Project administration

Friberg H: Conceptualisation, Writing - Review & Editing, Supervision, Project administration

Dankiewicz J: Conceptualisation, Methodology, Software, Formal analysis, Supervision, Project administration

DECLARATIONS

ETHICAL APPROVAL

The TTM2 trial was approved by the ethics committee in each participating country. Informed consent was obtained from all participants who regained mental capacity, and was otherwise waived, deferred or obtained from a legal surrogate.

AUTHORS' CONTRIBUTIONS

SS, CR, HF, NN and JD conceptualised the study. SS, CR and JD wrote the first manuscript draft. SS and JD performed all data analyses. AC, MPW, MT, TRK, DE, MH, PDWG, JB, AMG, NN and HF provided input on the analyses and their interpretation, and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

The TTM2-trial was supported by independent research grants from nonprofit or governmental agencies (the Swedish Research Council [Vetenskapsrådet], Swedish Heart–Lung Foundation, Stig and Ragna Gorthon Foundation, Knutsson Foundation, Laerdal Foundation, Hans-Gabriel and Alice Trolle-Wachtmeister foundation for medical Research, and regional research Support in Region Skåne) and by governmental funding of clinical research within the Swedish National Health Service. Authors S.S and H.F received regional research support from Region Skåne and funding from the Laerdal Foundation, respectively, for the present work.

DECLARATION OF INTEREST

Authors AC and HF are members of the editorial board of *Resuscitation*.

DATA AVAILABILITY STATEMENT

The analyses described in this manuscript derive from data obtained as part of the TTM2-trial and we therefore refer to the data sharing statement of its original publication.