Clinical paper

External validation and insights about the calibration of the return of spontaneous circulation after cardiac arrest (RACA) score

Lorenzo Gamberinia, Chiara Natalia Tartivitaa, Martina Guarneraa, Davide Allegrib, Simone Baroncini, Tommaso Scquizzatoc, Marco Tartaglionea, Carlo Alberto Mazzoli, Valentina Chiarinia, Cosimo Picocca, Carlo Coniglioa, Federico Semeraro, Giovanni Gordinia

a Department of Anesthesia, Intensive Care and Prehospital Emergency, Ospedale Maggiore Carlo Alberto Pizzardi, Bologna, Italy
b Department of Clinical Governance and Quality, Bologna Local Healthcare Authority, Bologna, Italy
c Department of Anesthesia and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy

Abstract

Background: The return of spontaneous circulation (ROSC) after cardiac arrest (RACA) score was developed as a tool to predict ROSC probability (pROSC) based on easily available information and it could be useful to compare the performances of different EMS agencies or the effects of eventual interventions.

We performed an external validation of the RACA score in a cohort of out of hospital cardiac arrest (OHCA) patients managed by the EMS of the metropolitan city of Bologna, Italy.

Methods: We analyzed data from 2,310 OHCA events prospectively collected between January 2009 and June 2021. Discrimination was assessed with the area under the ROC curve (AUROC), while the calibration belts were used for the comparison of observed versus expected ROSC rates. The AUROCs from our cohort and other validation cohorts were compared using a studentized range test.

Results: The AUROC for the study population was 0.691, comparable to that described by previous validation studies. Despite an acceptable overall calibration, we found a poor calibration for asystole and low pROSC ranges in PEA and shockable rhythms. The model showed a good calibration for patients aged over 80, while no differences in performance were found when evaluating events before and after the implementation of 2015 ERC guidelines.

Conclusions: Despite AUROC values being similar in different validation studies for RACA score, we suggest separating the different rhythms when assessing ROSC probability with the RACA score, especially for asystole.

Keywords: Cardiopulmonary Resuscitation, Heart Arrest, Emergency Medical Services, RACA score

Introduction

Out of hospital cardiac arrest (OHCA) is still a global health burden, with an estimated annual incidence between 67 and 170 cases per 100,000 inhabitants in Europe. Resuscitation of victims suffering OHCA is attempted or continued by the emergency medical services (EMS) personnel in about 50–60% of cases.1

OHCA management is complex and involves multiple specialities from the initial resuscitation to hospital discharge, and the final patients’ outcome could be influenced by multiple aspects in this chain.2
The local EMS are firstly involved, and play a key role in determining patient survival to the next phases of assistance. However, there is considerable variability in terms of prehospital assistance among Europe and reported prehospital ROSC range between 8% and 42%. Clearly, there is a need to build useful instruments to benchmark the performance of different EMS systems and to evaluate the effects of eventual interventions within a system.

In 2011, Gräsner et al. developed and validated the return of spontaneous circulation after cardiac arrest (RACA) score, a predictive tool aiming at estimating the probability of ROSC on the field for patients undergoing OHCA. The RACA score is based on eight variables (sex, age, aetiology, witnessed OHCA, location, initial ECG, bystander CPR and EMS arrival time) easily accessible from cardiac arrest registries and represents an effective solution for predicting ROSC probability.

The external validations of the RACA score performed in recent years in Europe and Asia gave similar results concerning the discrimination capacity of the model. On the other hand, the Finnish external validation study reported a better calibration of the model for the different contexts, especially focusing on the calibration of the model in different settings, including patients with shockable versus non-shockable rhythms, and it also resulted poorly calibrated when applied to the Pan-Asian Resuscitation Outcomes Study (PAROS) registry, which required an adjustment of the original constant coefficient. Additional validation studies, especially focusing on the calibration of the model in different contexts, could help to understand these differences.

We particularly focused on the calibration of the model for the different initial rhythms.

The secondary objective of the study was to assess the calibration of the model based on patients’ age and year of the OHCA event, to account for the recent revision of the European Resuscitation Council guidelines on resuscitation in 2015, and the different potential implications of age on resuscitation-related decisions.

| Table 1 – General characteristics of the overall study population and its subgroups divided by the first observed rhythms with regard to the RACA score variables. |
|-------------------------------|----------------|----------------|----------------|--------------|---|
| Age – years – median (IQR) | Overall (n = 2310) | VT/VF (n = 730) | PEA (n = 905) | Asystole (n = 675) | p  |
| Age > 80 years – n (%) | 609 (26.4%) | 131 (17.9%) | 328 (36.2%) | 150 (22.2%) | <0.001 |
| Sex – male – n (%) | 1563 (67.7%) | 587 (80.4%) | 545 (60.2%) | 431 (63.9%) | <0.001 |
| Aetiology – n (%) | 52.3% | 41.1% | 39.7% | 30.3% | 0.018 |
| – Trauma | 71 (4.2%) | 2 (0.3%) | 64 (7.1%) | 31 (4.6%) | 0.001* |
| – Hypoxia | 223 (9.7%) | 7 (1.0%) | 116 (12.8%) | 100 (14.8%) | 0.001* |
| – Intoxication | 48 (2.1%) | 5 (0.7%) | 18 (2%) | 25 (3.7%) | 0.001* |
| – Other | 1942 (84.1%) | 716 (98.1%) | 707 (78.1%) | 519 (76.9%) | 0.001* |
| Witnessed – n (%) | 1483 (64.2%) | 560 (76.7%) | 518 (57.2%) | 405 (60%) | 0.001* |
| – Lay people | 541 (23.4%) | 120 (16.4%) | 324 (35.8%) | 97 (14.4%) | 0.001* |
| – Professional | 286 (12.4%) | 50 (6.8%) | 63 (7.2%) | 173 (25.6%) | 0.001* |
| Location – n (%) | 19 (0.8%) | 6 (0.8%) | 8 (0.9%) | 5 (0.7%) | 0.001* |
| – Doctor’s office | 1 (0.04%) | 0 (0%) | 1 (0.1%) | 0 (0%) | 0.001* |
| – Public place | 293 (12.7%) | 141 (19.3%) | 82 (9.1%) | 70 (10.4%) | 0.001* |
| – Medical institution | 40 (1.7%) | 19 (2.6%) | 14 (1.5%) | 7 (1%) | 0.001* |
| – Other | 1957 (84.7%) | 564 (77.3%) | 800 (98.4%) | 593 (87.9%) | 0.001* |
| Bystander CPR – n (%) | 1437 (62.2%) | 507 (69.5%) | 497 (54.9%) | 433 (64.1%) | 0.001* |
| EMS arrival time – min – median (IQR) | 9 (7–12) | 8 (6–11) | 9 (7–12) | 9 (7–12) | 0.001* |
| Predicted ROSC – % – median (IQR) | 44% (34.6–58.9) | 62.2% (57.4–65.9) | 41.1% (34.8–47.5) | 33.7% (27.0–39.2) | <0.001 |
| Observed ROSC – % – median (IQR) | 97% (42.3%) | 455 (62.3%) | 395 (43.6%) | 128 (19.0%) | <0.001 |

Abbreviations: VF/VT – ventricular fibrillation/ventricular tachycardia; PEA – pulseless electrical activity; IQR – interquartile range; CPR – cardiopulmonary resuscitation; EMS – emergency medical services; ROSC – return of spontaneous circulation.

*p value calculated with the Chi square test on the whole contingency table referred to the main variable.
... evidence about ethics based on scoping reviews of 22 research questions, and expert opinions from different countries. The variation of ethics approach worldwide is very wide and the culture and legislation of the single country affect deeply the behaviour of healthcare professionals in the management of cardiac arrest. Concerning OHCA, in Italy, only physicians are allowed to withhold, withdraw, or terminate resuscitation, therefore, only at ALS team arrival, basic life support can be interrupted if ROSC hasn’t been already achieved. Termination of resuscitation (ToR) follows the contemporary ERC guidelines.

The study was approved by the local Ethics committee.

Inclusion and exclusion criteria
We included all the patients suffering OHCA assisted by the metropolitan EMS in whom a full resuscitation was attempted by the ALS healthcare personnel attending the scene. The exclusion criteria were incomplete data and immediate resuscitation withhold or withdrawal at the arrival of an ALS unit.

Outcome measures
The primary outcome measure was ROSC, defined as a palpable pulse at any point during resuscitation for over 30 seconds, patients transported to the emergency department (ED) with ongoing CPR were classified as no-ROSC.

The RACA score was calculated as originally described by Gräsner et al., please see supplement Table 1 for details.

The expected probability of ROSC for every patient was calculated based on the RACA score as follows: pROSC = 1/(1 + e^−x), where x is the obtained value for the RACA score.

Statistical analysis
Data were collected in an electronic case report form (FileMaker Pro 16; FileMaker, Inc, Santa Clara, California, USA) and analyzed using Stata/CI 16 (College Station, Texas, USA) and Microsoft Excel 2019 (Microsoft Corporation, Redmond, Washington, USA).

Continuous variables were expressed as a median and interquartile range [IQR], while categorical variables were expressed as numbers and percentages. Comparisons between continuous variables were performed with the Mann-Whitney U test, while the Chi-square test or, if appropriate, the Fisher’s exact test were used for categorical variables.

Discrimination was examined by calculation of the area under the receiver operating characteristic curve (AUROC). The observed AUROC for the study population, together with those reported in the other external validation studies available from the literature, and the internal validation cohort of the original study were compared using a studentized range test.

Calibration was assessed using the calibration belts, a graphical approach designed to evaluate the goodness of fit of binary outcome models examining the relationship between estimated probabilities and observed outcome rates. This approach allows the creation of confidence belts for the calibration curve by fitting the observed data with a general calibration function and plotting the corresponding curve.

The resulting plot contains a bisector, indicating the expected probabilities from the model, while the calibration belts represent the confidence intervals for observed probabilities, therefore allowing to finely discriminate the overall calibration and the ranges in which the model miscalibrates, in addition to indicating the direction of this phenomenon.

The graphical approach is paired to a statistical test, also available for external validation, thus synthesizing the calibration assess-
ment in a standard hypothesis testing framework. This method thus offers a more analytical view in the assessment of calibration of dichotomous models, compared to other approaches such as the Hosmer and Lemeshow goodness of fit method, and has already been used in evaluating the calibration for the predictive scores.

Results

Demographics and main characteristics

During the study period, 12,919 OHCA cases were managed by the metropolitan EMS of Bologna, and resuscitation was attempted or continued by the ALS units in 2,322 of them (18.0%) that were considered for inclusion in this study, supplement Fig. 1 shows the yearly distribution of the OHCA events and resuscitation attempts. Twelve cases had missing data about the first observed rhythm and were subsequently excluded from the analyses, therefore, the final study population in all the analyses included 2,310 out of 2,322 patients (99.48%), supplement Fig. 2 shows the flow of patients throughout the study.

Globally, patients were mostly men (n = 1,563, 67.7%), with a median age of 71 years. The most frequent first observed rhythm was PEA (n = 905, 39.2%), followed by VF/VT (n = 730, 31.6%) and asystole (n = 675, 29.2%), and the most frequently suspected etiology, following the categorization used in the RACA score, was medical (n = 1,942, 84.1%), followed by hypoxia (n = 223, 9.7%).

In the majority of cases, the OHCA event was witnessed (n = 1,084, 87.6%), and resuscitation was started by the bystanders in 1,437 out of 2,310 patients (62.2%), median EMS arrival time was 9 minutes.

The median expected ROSC rate following the RACA score for the whole population was 44% (IQR 34.6–58.9%), and the observed ROSC rate was 42.3%. Table 1 shows the main characteristics of the population concerning RACA score variables and the first observed rhythm, while Fig. 1 shows the distribution of the different rhythms within the annual number of OHCA and the deciles of ROSC probability (pROSC) calculated from the RACA score. As expected, based on the RACA scoring system, shockable rhythms mostly fell in the upper range of pROSC, PEA was mostly present in the fifth decile (pROSC 0.4–0.5), while asystole was mostly expressed in the lower deciles (pROSC 0.1–0.4).

Discrimination and calibration

Fig. 2 shows the ROC curve of the model, the calculated AUROC was 0.691 (95% CI: 0.669–0.713), while Fig. 3 displays the multiple AUROC comparisons between the study population and the different populations from available literature, expressed in the graph with their 95% confidence intervals. The studentized range test showed no significant differences among the AUROCs (p = 1.000).

Fig. 4 shows the calibration belts for the model considering the whole population and the different subpopulations based on the first observed rhythm, age and year of enrollment.

Concerning the overall population, despite the test statistic being significant, the 95% CI belt ran over the bisector, demonstrating an observed ROSC rate higher than expected, only for pROSC 0.01–0.02.

The calibration belts performed on the different subgroups based on the first observed rhythm were also significant and showed larger incongruences between the predicted and observed ROSC rates, especially for lower values of pROSC in PEA (<0.2) and shockable rhythms (<0.28), where observed ROSC rates were higher than predicted. Finally, for asystole, the observed 95% CI belt ran under the bisector, showing observed ROSC rates lower than expected, for most of the probability ranges (0.15–0.71).

Fig. 3 – 95% Confidence interval distributions of the internal and external validation cohorts of the RACA score. Note: studentized range test: p = 1.000.
Globally, the model showed an overall good calibration for patients aged over 80, while a similar profile to the global population was observed for the population under 80 years old. Finally, event year-based calibration belts showed a very sharp belt, but mostly under the bisector with observed ROSC lower than expected for the lower range of pROSC (0.1–0.46) for OHCAs happened up to 2015, and a wider belt running both over and under the bisector for events after 2015.

Fig. 4 – Calibration belts for the RACA score considering the whole population, first observed rhythms different age intervals and years of the event. Abbreviations: ROSC, return of spontaneous circulation.
Discussion

Up to a quarter of the urban EMS missions are related to OHCA, and the achievement of ROSC represents the first step in the care process for these patients. Moreover, ROSC could be considered a quality indicator for the EMS taking into account both organizational and technical aspects of the delivered resuscitation. Therefore, the development and validation of a predictive score for ROSC such as the RACA score could be useful in benchmarking different Emergency services or evaluating the effects of interventions.

In this external validation study of the RACA score applied to an Italian metropolitan EMS, we found moderate discrimination (AUROC: 0.691), that was comparable to those reported in other external validation studies. Calibration was deemed acceptable for the overall population despite a significant statistic test for the calibration belt due to the curve running over the bisector for the lowest 1% of expected ROSC probability (Fig. 4). Since the number of patients in this decile was very low, this aspect should be considered less relevant (see Fig. 1).

However, when the different first observed rhythms were separately analysed, suboptimal calibration was observed for the shockable rhythms and PEA, showing higher than expected ROSC rates, especially for lower pROSC. Moreover, the observed ROSC rates for asystole were almost constantly under the bisector, therefore constantly lower than expected.

Differences between expected and observed ROSC rates for non-shockable rhythms were already reported in previous validation studies, and they were attributed to the different protocols in regards to the cessation of resuscitation attempts.

In particular, Kupari et al. showed similar results with a good overall calibration for shockable rhythms but lower than expected ROSC rates for low pROSC, in particular for PEA and Asystole. On the other hand, Caputo et al. showed a good calibration of the model for all the pROSC ranges explored.

The populations evaluated in these calibration studies were notably different, in particular for age, with a proportion of patients aged over 80 ranging from 15% (Kupari et al.) to 39% (Caputo et al.). Also, the rate of bystander CPR and the first observed rhythms showed notable differences.

In fact, bystander CPR prevalence ranged from 14.6% (Gräsnér et al.) to 62.2% (reported in our population), moreover, our population and Finnish population had a lower relative frequency asystole (29% and 22%, respectively) compared to both the Caputo et al. population (45%) and the score development study (46.4%). Table 2 summarises the population differences among the evaluated studies.

OHCA events encompass a wide range of underlying mechanisms and situations, moreover, policies about do-not-attempt-cardiopulmonary-resuscitation (DNACPR) records and orders, withdrawal and termination of resuscitation (ToR) may vary greatly across the different EMS systems even within the same nation.

| Table 2 – Population characteristics of the development and the published external validation cohorts of the RACA score. |
|---------------------------------------------------------------|
| **RACA validation cohort** | **Helsinki** | **Panasian registry** | **Pavia Canton Ticino** | **Bologna** |
| **n** = 2,218 | **n** = 680 | **n** = 63,069 | **n** = 2,041 | **n** = 2,310 |
| Sex (male) – n (%) | 67.8% | 70% | 59.5% | 62% | 67.7% |
| Age > 80 – n (%) | 20.2% | 15% | 37.9% | 39% | 26.4% |
| Initial rhythm- n (%) | | | | | |
| VT/VF | 28.2% | 41% | 7.2% | 20% | 31.6% |
| PEA | 11.4% | 36% | 15.5% | 45% | 39.2% |
| Asystole | 46.4% | 22% | 51.7% | 27% | 29.2% |
| Unknown | 13.9% | 1% | 25.6% | 7% | 0% |
| Aetiology – n (%) | | | | | |
| Cardial/Other | 86.8% | 85% | 82.1% | 88% | 84.1% |
| Trauma | 2.5% | 2% | 12.9% | 4% | 4.2% |
| Respiratory | 6.8% | 8% | 5.0% | 5% | 9.7% |
| Intoxication | 1.9% | 5% | NA | 3% | 2.1% |
| Witnessed – n (%) | | | | | |
| None | 40.6% | 15% | 57.4% | 27% | 12.4% |
| Lay people | 50.3% | 62% | 7.7% | 53% | 64.2% |
| Professionals | 9.1% | 23% | 34.9% | 20% | 23.4% |
| Location – n (%) | | | | | |
| At home/other | 72.4% | 55% | 70.8% | 78% | 84.7% |
| Nursing Home | 4.3% | 2% | 8.2% | 7% | 0.8% |
| Workplace | 2.1% | 3% | NA | 1% | NA |
| Doctor’s Office | 1.6% | 1% | NA | 1% | 0.04% |
| Public place | 17.8% | 38% | 15.0% | 13% | 12.7% |
| Medical Institution | 1.8% | 2% | 6.0% | NA | 1.7% |
| Bystander CPR – n (%) | 14.6% | 50% | 35.9% | 45% | 62.2% |
| Time to EMS arrival – min | NA | NA | 6.66 | 10.4 | 9 |

Abbreviations: NA – not assessable either because directly reported in the ROSC/no ROSC groups without information about the overall population or not evaluated.
In Italy, DNACPR records are not in place by law, and only physicians can withhold, withdraw or terminate the resuscitation attempts, therefore, basic life support may be started by non-medical health personnel and subsequently stopped once ALS vehicles arrive. In particular, once that resuscitation is started from the first arrived personnel, situational and emotional aspects could have an impact on the decision to provisionally pursue resuscitative attempts even in cases with perceived futility and eventually early stop the manoeuvres. This could partially explain the notable reduction in the observed versus expected ROSC rate for asystole. Moreover, in support of this hypothesis, the age-based subanalysis of calibration belts showed a better calibration of the model when assessing patients aged over 80, where asystole was less represented among the first observed rhythms (24.6% vs 30.9%), probably reflecting the higher attitude of the local EMS personnel to withdraw resuscitation in elderly people with rhythms associated to a low survival probability.

It could be hypothesized that the calibration of the RACA score in different EMS contexts could be influenced by the attitude of the EMS personnel to initiate and terminate resuscitation efforts, especially for the low pROSC intervals, involving mostly asystole as first observed rhythm. Probably, the influence of the propensity of EMS personnel to initiate or terminate resuscitation should be furtherly investigated in this field.

Criteria for withholding, withdrawing and terminating resuscitation (ToR) are still being debated, especially in the prehospital setting, therefore, different socio-cultural, personal and situational aspects could introduce variability in these decisions, finally influencing the access of patients to advanced resuscitation.

Moreover, in recent years, the resources available for advanced resuscitation became more complex with the advent of extracorporeal CPR, the advances in trauma-related resuscitative algorithms and the growing collective consciousness of the need for organ donors harvesting, that could have furtherly influenced the decisions based on the different capabilities of the local health system. These aspects could partially explain the reduced calibration of the model in the cohort of patients experiencing OHCA after 2015.

Finally, the described limitation of the RACA score calibration for low pROSC and its associated rhythms could be theoretically applicable to other scores evaluating the probability of ROSC since the inclusion of patients with low pROSC could be different in different populations. The progressive definition of guidelines addressing the clinical decisions concerning the beginning and continuation of resuscitative efforts should reduce this phenomenon.

The main limitation of this study relies upon its monocentric nature, which could have emphasized the differences in terms of resuscitation decisions due to the local clinicians’ behaviour, however, this aspect could also have underlined a potential role of the RACA score in fostering local audit and discussion about DNACPR, withdrawal and ToR policies.

**Conclusion**

In this external validation of the RACA score encompassing eleven years of activity of a single Italian EMS system, we found similar discrimination (AUROC: 0.691) to that reported in other contexts. Calibration was overall acceptable in the global population for the most represented deciles of expected ROSC, however, when the single first observed rhythms were analysed separately, we observed significant discrepancies between observed and expected ROSC rates at low baseline pROSC, in particular for asystole.

Therefore, we suggest that, when interpreting the observed versus expected ROSC with the RACA score in different contexts, separate analyses based on the first observed rhythms should be performed to avoid inclusion biases due to the different behaviours of the EMS in conditions of low pROSC, particularly when asystole is the first observed rhythm.

**Ethics approval and consent to participate**

The study was approved by the Institutional Review Board (IRB) of Maggiore Hospital, Bologna, Italy.

**Conflict of interest**

No relationship exists between any of the authors and any commercial entity or product mentioned in this manuscript that might represent a conflict of interest. No inducements have been made by any commercial entity to submit the manuscript for publication. All within 3 years of beginning the work submitted. LG is Scientific Committee member of the Italian Resuscitation Council. CC is member of Executive Committee of the Italian Resuscitation Council. FS is the Chair-Elect of the European Resuscitation Council, Chair of the ILCOR Social Media Working Group, ILCOR BLS Working Group members, Scientific Committee member of Italian Resuscitation Council. CNT, MG, DA, SB, TS, MT, CAM, VC, CP, and GG have no conflicts of interest.

**Authors contributions**

LG and FS concepted and designed the work. CNT, MG, SB and CP acquired and interpreted the data.

DA and LG performed the statistical analysis. LG, MT and VC drafted the article. TS, CC, FS, and GG substantively revised the article.

All the Authors have approved the submitted version and have agreed both to be personally accountable for the author’s own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved.

**Appendix A. Supplementary material**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.resplu.2022.100225.

**References**

1. Gräsnér J-T, Herlitz J, Tjølmland IBM, et al. European Resuscitation Council Guidelines 2021: Epidemiology of cardiac arrest in Europe 2021. https://doi.org/10.1016/j.resuscitation.2021.02.007.
2. Ørbo M, Aslaksen PM, Larsby K, et al. Determinants of cognitive outcome in survivors of out-of-hospital cardiac arrest. Resuscitation 2014;85:1462–8. https://doi.org/10.1016/j.resuscitation.2014.08.010.
