Original Article

Multicenter observational study on the reliability of the HEART score

Nicola Parenti[1], Giuseppe Lippi[2], Maria Letizia Bacchi Reggiani[3], Antonio Luciani[1], Mario Cavazza[3], Antonello Pietrangelo[4], Alberto Vegetti[4], Lucio Brugioni[1], Laura Bonfanti[5], Gianfranco Cervellin[5]

[1]Emergency Department, University Hospital of Modena, Modena, Italy
[2]Section of Clinical Biochemistry, University of Verona, Verona, Italy
[3]Emergency Department, University of Bologna, Bologna, Italy
[4]Internal Department, University Hospital of Modena, Modena, Italy
[5]Emergency Department, University Hospital of Parma, Parma, Italy

Objective To rapidly and safely identify the risk of developing acute coronary syndrome in patients with chest pain who present to the emergency department, the clinical use of the History, Electrocardiogram, Age, Risk Factors, and Troponin (HEART) scoring has recently been proposed. This study aimed to assess the inter-rater reliability of the HEART score calculated by a large number of Italian emergency physicians.

Methods The study was conducted in three academic emergency departments using clinical scenarios obtained from medical records of patients with chest pain. Twenty physicians, who took the HEART score course, independently assigned a score to different clinical scenarios, which were randomly administered to the participants, and data were collected and recorded in a spreadsheet by an independent investigator who was blinded to the study's aim.

Results After applying the exclusion criteria, 53 scenarios were finally included in the analysis. The general inter-rater reliability was good (kappa statistics [κ], 0.63; 95% confidence interval, 0.57 to 0.70), and a good inter-rater agreement for the high- and low-risk classes (HEART score, 7 to 10 and 0 to 3, respectively; κ, 0.60 to 0.73) was observed, whereas a moderate agreement was found for the intermediate-risk class (HEART score, 4 to 6; κ, 0.51). Among the different items of the HEART score, history and electrocardiogram had the worse agreement (κ, 0.37 and 0.42, respectively).

Conclusion The HEART score had good inter-rater reliability, particularly among the high- and low-risk classes. The modest agreement for history suggests that major improvements are needed for objectively assessing this component.

Keywords HEART score; HEART pathway; Chest pain; Acute coronary syndrome; Emergency service, hospital

Capsule Summary	**What is already known**
	The History, Electrocardiogram, Age, Risk Factors, and Troponin (HEART) score is useful for management of chest pain patients presenting to the emergency department because is simple, easy and rapid, and has also been validated to predict major adverse cardiac events in many studies conducted in the emergency department.
	What is new in the current study
	In this study we found that the HEART reliability is moderate-good but the parameter history showed a fair inter-rater-reliability for its arbitrary interpretation.

INTRODUCTION

Chest pain is one of the most frequent symptoms leading to emergency department (ED) admission, and it may be triggered by several causes, ranging from mostly harmless to immediate life-threatening disorders. Based on the perspective of an emergency physician (EP), rapid identification of high-risk patients and concomitant ruling out of low-risk conditions are important. According to previous evidence, acute coronary syndrome (ACS) may not be identified as an underlying cause in approximately 20% to 25% of patients with chest pain who visited the ED[1] and in early 45% of those admitted to a chest pain unit.[2] The leading aspects in the EPs' toolbox that can help identify the probability of ACS are patient history, electrocardiogram (ECG) findings, and cardiac troponin testing results, which are often combined with diagnostic algorithms designed for the rapid rule-in or rule-out of ACS.[3] However, a definitive and universally agreed strategy is still not identified and acknowledged.

Some official documents, endorsed by eminent scientific societies, have recently encouraged the use of clinical scores for evaluating patients with chest pain suggestive of ACS who present to the ED.[4,5] In particular, a recent systematic review comprehensively analyzed the leading clinical prediction rules for chest pain, including the Thrombolysis in Myocardial Infarction (TIMI) risk score, the History, ECG, Age, Risk Factors, and Troponin (HEART) score, and the Global Registry of Acute Coronary Events (GRACE) scores.[6] Among the aforementioned risk stratification tools, the HEART score was found to be useful for managing patients with chest pain who present to the ED because it is simple, easy, and quick to use and it has also been validated in several studies conducted in the ED.[7-12] Five main parameters contribute to the calculation of the final HEART score, which include clinical history, ECG findings, age, risk factors, and troponin testing results (Table 1). The assignment of points ranging from 0 and 2 to each of these five

items contributes in obtaining the final score, which will be between 0 and 10.[7,8]

According to the final HEART score, patients can be classified into three groups: low (score of 0 to 3), intermediate (score of 4 to 6), and high (score of 7 to 10) risk for major adverse cardiac event (MACE) within 6 weeks. Notably, the definition of MACE includes acute myocardial infarction, percutaneous coronary intervention, coronary artery bypass grafting, coronary angiography revealing significant stenosis, and death due to any cause. A different management is then advocated for patients at low, intermediate, and high risk who require discharge, admission, and early invasive strategies, respectively.

The different clinical scores should be accurately evaluated and clinically validated.[13] However, to the best of our knowledge, only a single study has assessed the reliability of the HEART score.[14] This clearly represents a major drawback since the interpretation

Table 1. Items of the HEART (History, Electrocardiogram, Age, Risk Factors, and Troponin) score

Parameter	Data	Score
History	Highly suspicious	2 Points
	Moderately suspicious	1 Point
	Slightly or non-suspicious	0 Points
Age	≥ 65 years	2 Points
	45–64 years	1 Point
	< 45 years	0 Points
Risk factors	≥ 3 risk factors or a history of CAD	2 Points
	1 or 2 risk factors	1 Point
	No risk factors	0 Points
ECG	Significant ST depression	2 Points
	Nonspecific reporalization	1 Point
	Normal	0 Points
Troponin	≥ 3 × normal limit	2 Points
	> 1– < 3 × normal limit	1 Point
	Normal limit	0 Points

of both history and ECG is arbitrary to some extent (Fig. 1), and as a result, the assignment of the final score may be biased by a substantial degree of heterogeneity. Therefore, this multicenter study aimed to evaluate the inter-rater reliability of the HEART score calculated by a large number of Italian EPs.

METHODS

Study design and setting
This multicenter study was conducted in three Italian academic EDs (university hospitals in Bologna, Parma, and Modena) between March 2017 and December 2017. All three EDs used a harmonized triage procedure and managed a high number of patients, including those with chest pain.

The study was approved by the ethical committee of the Azienda Ospedaliero-Universitaria Policlinico Modena, Italy (Dnr: 96/17; 1977) and was conducted according to the Declaration of Helsinki under the terms of the relevant local legislation.

Data collection
The method suggested by Rotondi and Donner[15] was used to calculate the minimum sample size needed for a reliable estimation of kappa statistics (κ) according to multiple raters and multinomial outcomes.

According to the preliminary calculation (i.e., κ, 0.80; estimate prevalence of 0.2, 0.4, and 0.4 for low-, intermediate-, and high-risk scores, respectively), we planned to collect at least 53 different clinical scenarios. Hence, paper scenarios were obtained from the medical records of patients with chest pain who were admitted to the ED of the university hospital in Modena during a 2-month period (i.e., from January 1, 2017 to February 28, 2017).

Information about 59 patients was collected (one randomly selected patient with chest pain per day): demographic and clinical characteristics, nurse triage category, discharge data, clinical setting of admission, history, previous diseases, vital signs, and pain score. Additional information included age, gender, ECG data, and results of the cardiac troponin I testing (Ortho Vitros ECi, Ortho-Clinical Diagnostics, Raritan, NJ, USA; 99th upper reference limit, < 34 ng/L). The exclusion criteria were as follows: (1) incomplete demographic and clinical data, (2) patients presenting with only dyspnea or palpitations, and (3) patients presenting with chest pain and significant ST segment elevation on ECG.

Study protocol
Twenty physicians who were recruited from the EDs, internal medicine wards, and postgraduate emergency medicine school were randomly assigned to 5-hour training for the utilization of the HEART score. These participants were selected by the directors of the wards according to their willingness to participate in the study.

After completing the training on HEART score, each physician independently assigned a score to the different clinical scenarios. To prevent intercommunication among participants, they were asked to calculate the score on the same day and in the presence of the principal investigators. The clinical scenarios were randomly administered to the participants, who had access to the HEART score rules, with a 2-hour limit for completing the scoring process. Data were collected and recorded in a spreadsheet by an independent investigator, who was blinded to the aim of the study.

Data analysis and outcome
A participant was then asked to rank the score (i.e., 0, 1, or 2) for each of the five demographic and clinical characteristics to obtain the final HEART score, which helped in the classification of patients who are at low (score of 0 to 3), intermediate (score of 4 to 6), and high (score of 7 to 10) risk for MACE.[7,8]

The main endpoint of this study was the estimation of inter-rater agreement in the calculation of the HEART score (κ value and 95% confidence interval [CI]) among physicians. Whether clinical experience could help improve the inter-rater reliability of calculating the HEART score was the secondary endpoint. This second aspect was assessed by comparing inter-rater reliability among expert EPs (i.e., those with more than 10 years of experience in emergency medicine) and students or physicians with no experience in emergency medicine.

According to the literature, poor, fair, moderate, good, and very good agreements were defined as a κ value between 0.00 and 0.20, 0.21 and 0.40, 0.41 and 0.60, 0.61 and 0.80, and 0.81 and 1.00, respectively. Statistical significance was set at a 0.05 alpha level. The Stata ver. 14.2 (StataCorp., College Station, TX, USA) was used for statistical analysis.

Table 2. Characteristics of the centers and raters

Characteristics	Modena	Bologna	Parma
No. of patients who visited the ED	63,808	71,994	94,858
Triage urgency code (%)			
Red (level 1)	1.5	2.4	2.7
Yellow (level 2)	17.1	24	18
Green (level 3)	69.9	55.3	73.3
White (level 4)	12.5	18.3	6
No. of raters	14	3	3
Years of experience in the ED (mean)	9	11	13

Triage urgency code: red (level 1), immediate response; yellow (level 2), green (level 3); and white (level 4), assessment within 20, 60, and 120 minutes, respectively.
ED, emergency department.

RESULTS

The three centers were all university hospitals, with similar characteristics in terms of patient volume and case mix (Table 2). The EPs recruited from the three centers also had similar experience in emergency medicine practice. Overall, 6 of the 59 clinical scenarios were excluded since they did not fulfill all our inclusion/exclusion criteria. Finally, 53 clinical scenarios were included in the analysis. The characteristics of the 53 clinical scenarios are shown in Table 3. The mean age of the patients was 56 (range, 16 to 92) years. Of the participants, 27 were men and 26 were women. Hypertension and smoking were the most frequent cardiovascular risk factors.

The final HEART score of each scenario was similar among all participants (Fig. 1). The distribution of the final HEART scores was similar to that observed in previous studies,[7,8] with 20% and 40% of clinical scenarios assigned to high-risk class as well as

intermediate- and low-risk classes, respectively. The general inter-rater reliability was good (κ, 0.63; 95% CI, 0.57 to 0.70) and was similar between senior physicians (κ, 0.65; 95% CI, 0.57 to 0.73) and junior physicians (κ, 0.60; 95% CI, 0.51 to 0.72) (Table 4).

Overall, the study participants also had a good inter-rater agreement for high- and low-risk classes (HEART scores of 7 to 10 and 0 to 3; κ, 0.70 and 0.72, respectively), whereas moderate agreement was observed for the intermediate-risk class (HEART score of 4 to 6; κ, 0.51) (Table 3).

Importantly, history was characterized by a worse agreement (κ, 0.37) among the different HEART score items, with an extremely modest reliability among all participants. Modest agreement

Table 3. Characteristics of the scenarios

Characteristics	Value
Number	53
Age	56 ± 16
Sex, male	27 (51)
Smokers	22 (42)
Diabetes mellitus	10 (19)
Obesity	12 (23)
Family history of CAD	9 (17)
Hypertension	24 (45)
Hypercholesterolemia	13 (24)
History of atherosclerotic disease	18 (34)
Troponin I (ng/L)	12 (12–34)

Values are presented as mean ± standard deviation, number (%), or median (interquartile range).
CAD, coronary artery disease.

Fig. 1. History, Electrocardiogram, Age, Risk Factors, and Troponin (HEART) score assignment. Seniors: physicians with more than 5 years of experience in emergency medicine. Juniors: physicians with less than 5 years of experience in emergency medicine. HEART risk class: according to the literature, we showed the assignment of HEART score among participants by using three groups of risk class for future myocardial ischemic accidents: low (score of 0 to 3), intermediate (score of 4 to 6), and high (score of 7 to 10) risk.

Table 4. Inter-rater agreement among the HEART scores

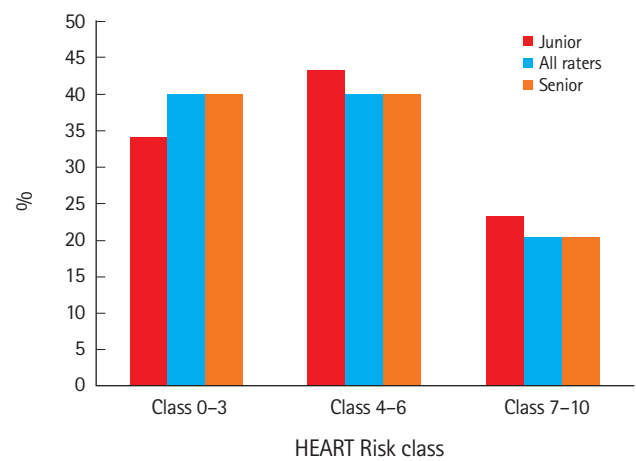|	All raters (n = 20)	Senior physicians (n = 12)	Juniors physicians (n = 8)
κ (95% CI)	0.63 (0.57–0.70)	0.65 (0.57–0.73)	0.60 (0.51–0.72)
Low risk (class 1)[a]	0.72	0.73	0.60
Medium risk (class 2)[a]	0.51	0.53	0.47
High risk (class 3)[a]	0.70	0.70	0.69
History (95% CI)	0.37 (0.27–0.44)	0.38 (0.31–0.43)	0.35 (0.27–0.43)
ECG findings (95% CI)	0.42 (0.35–0.48)	0.46 (0.35–0.57)	0.37 (0.29–0.47)
Risk factors (95% CI)	0.71 (0.60–0.76)	0.70 (0.59–0.82)	0.72 (0.65–0.78)
Troponin testing findings (95% CI)	0.92 (0.88–0.95)	0.90 (0.85–0.94)	0.96 (0.92–1)
Age (95% CI)	0.94 (0.92–0.97)	0.93 (0.90–0.97)	0.96 (0.93–1)

HEART, History, Electrocardiogram, Age, Risk Factors, and Troponin; CI, confidence interval; ECG, Electrocardiography.
[a]We tested the inter-rater agreement among the HEART score risk classes: low (score of 0–3), intermediate (score of 4–6), and high (score of 7–10) risk for major adverse cardiac event within 6 weeks. According to the literature, we considered a poor, fair, moderate, good, and very good agreement in κ values between 0.00 and 0.20, 0.21 and 0.40, 0.41 and 0.60, 0.61 and 0.80, and 0.81 and 1, respectively.

was also found for ECG score (κ, 0.37 to 0.46) (Table 3), whereas a significantly better concordance was observed for the remaining three parameters of the HEART score (i.e., risk factors, age, and troponin), as shown in Table 3.

DISCUSSION

Results showed that the calculation of the HEART score was similar among all participants, with comparable scores obtained by senior and junior physicians. In particular, a good inter-rater agreement was found for high- and low-risk classes, whereas the agreement was only modest for the intermediate-risk class.

The hypothesis that the subjective interpretation of history and ECG may influence the final calculation of the HEART score is supported by our data since larger heterogeneity was observed in scoring these two variables compared to risk factors, age, and troponin.

The HEART score has only been validated in a single center retrospective study[7] and in an ensuing multicenter study,[8] which both analyzed the predictive value of the score for the combined end point of acute myocardial infarction, percutaneous coronary intervention, coronary artery bypass grafting, or death (MACE) within 6 weeks after initial assessment.

More recently, another study has compared the performance of the HEART score with that of the GRACE and TIMI scores in predicting MACE in 1,748 patients with chest pain who were admitted to the ED. Results have shown that the HEART score outperformed the other two risk assessment tools and reliably and safely identified a larger group of low-risk patients.[16] The impact of the HEART score on health care resources and expenditure has also been assessed in another study,[11] which confirmed that the use of this score is safe in patients with chest pain, although a high non-compliance rate to management recommendations mitigates its otherwise favorable impact on the utilization of healthcare resources. Based on this evidence, the HEART score may have a good performance in the diagnosis and prognosis of patients with ACS in several clinical settings; hence, it may be reliable when used for estimating the risk of MACE in this category of patients.

In particular, Van Den Berg et al.[17] have conducted a recent systematic review and meta-analysis of the literature, which included 12 studies and 11,217 patients, and concluded that the HEART score identifies patients with a suspected diagnosis of ACS who have a low probability (1.6%) of developing MACEs and who could be safely discharged from the ED. The area under the curve and the pooled sensitivity of the HEART score for predicting MACEs were both excellent (i.e., 0.81 and 0.97, respectively), whereas the pooled specificity was modest (i.e., 0.47).[17]

To the best of our knowledge, only a single study about the in-

ter-rater reliability of the HEART score have been previously conducted.[14] Although the study design was similar to that of our investigation (i.e., retrospective observational study that used clinical scenarios), the conclusion was quite different. In particular, Wu et al.[14] have found a substantial disagreement in the assignment of the HEART score to 33 clinical scenarios between EPs and cardiologists. Unlike these findings, we found a good agreement in the assignment of the HEART score to clinical scenarios among all raters. According to our findings, history was the primary source of disagreement (κ, 0.13; 95% CI, -0.1 to 0.40). In addition, a better agreement was observed among EPs and cardiologists for risk factors, age, and troponin.

The findings of our study may have some practical implications for managing patients with chest pain in the ED. In fact, our data showed that the HEART score may be used by both senior and junior physicians, with good inter-rater agreement (at least for patient classification in high- and low-risk classes). Notably, a score assignment to history should be modified to allow a more objective interpretation and ultimately mitigation of the impact of subjectivity.

Since the HEART score was a reliable tool for classifying patients who are at low or high risk for MACE, it may be safely used in ruling out patients who are at low risk for ACS and encouraging additional investigations on high-risk patients. Nevertheless, the modest agreement found in classifying patients with an intermediate risk suggests its efficiency in identifying whether or not patients should be continuously monitored is uncertain. Indeed, further studies must be conducted to compare the reliability of other assessment tools (e.g., GRACE and TIMI) using a similar cluster set of clinical scenarios.

Interestingly, the good inter-rater reliability among all participants may allow an accurate communication among the users of the HEART score, promote a better standardization in health care, help obtain a more reliable information for benchmarking, enhance patient safety, and encourage larger support for clinical research for national surveillance.

The use of clinical scenarios rather than actual clinical settings may be considered as a drawback in our study. However, performing actual trials with patients in the ED remains challenging, and more importantly, the clinical scenario approach has been used and validated in other studies that aimed to estimate the inter-rater reliability of other clinical scores.[18] Second, another possible limitation of our study is the fact that the participants only had a relative short experience in using the HEART score. Finally, we did not compare the reliability of the HEART score with that of the other scores since this will be the focus of our next investigation in the future.

In summary, the HEART score had a good inter-rater reliability among a large number of Italian physicians, whereas a less satisfactory agreement was found in assigning the score to history. The experience of our participants did not substantially influence their scoring reliability. Overall, our participants had a good inter-rater agreement for high- and low-risk classes based on the HEART score. Meanwhile, the agreement was only modest for intermediate-risk class. In particular, the modest agreement for assigning the score to history suggests that additional efforts should be exerted in achieving a more objective assessment of this parameter.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Goodacre S, Cross E, Arnold J, Angelini K, Capewell S, Nicholl J. The health care burden of acute chest pain. Heart 2005;91: 229-30.
2. Conti A, Paladini B, Toccafondi S, et al. Effectiveness of a multidisciplinary chest pain unit for the assessment of coronary syndromes and risk stratification in the Florence area. Am Heart J 2002;144:630-5.
3. Cervellin G, Mattiuzzi C, Bovo C, Lippi G. Diagnostic algorithms for acute coronary syndrome-is one better than another? Ann Transl Med 2016;4:193.
4. Hamm CW, Bassand JP, Agewall S, et al. ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: the task force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2011;32:2999-3054.
5. Amsterdam EA, Kirk JD, Bluemke DA, et al. Testing of low-risk patients presenting to the emergency department with chest pain: a scientific statement from the American Heart Association. Circulation 2010;122:1756-76.
6. Fanaroff AC, Rymer JA, Goldstein SA, Simel DL, Newby LK. Does this patient with chest pain have acute coronary syndrome?: the rational clinical examination systematic review. JAMA 2015;314:1955-65.
7. Six AJ, Backus BE, Kelder JC. Chest pain in the emergency room: value of the HEART score. Neth Heart J 2008;16:191-6.
8. Backus BE, Six AJ, Kelder JC, et al. Chest pain in the emergency room: a multicenter validation of the HEART score. Crit Pathw Cardiol 2010;9:164-9.
9. Backus BE, Six AJ, Kelder JC, et al. A prospective validation of the HEART score for chest pain patients at the emergency department. Int J Cardiol 2013;168:2153-8.
10. Six AJ, Cullen L, Backus BE, et al. The HEART score for the assessment of patients with chest pain in the emergency department: a multinational validation study. Crit Pathw Cardiol 2013;12:121-6.
11. Poldervaart JM, Reitsma JB, Backus BE, et al. Effect of using the HEART score in patients with chest pain in the emergency department: a stepped-wedge, cluster randomized trial. Ann Intern Med 2017;166:689-97.
12. Hyams JM, Streitz MJ, Oliver JJ, et al. Impact of the HEART pathway on admission rates for emergency department patients with chest pain: an external clinical validation study. J Emerg Med 2018;54:549-57.
13. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. Ann Emerg Med 1999;33:437-47.
14. Wu WK, Yiadom MY, Collins SP, Self WH, Monahan K. Documentation of HEART score discordance between emergency physician and cardiologist evaluations of ED patients with chest pain. Am J Emerg Med 2017;35:132-5.
15. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. J Clin Epidemiol 2012;65: 778-84.
16. Poldervaart JM, Langedijk M, Backus BE, et al. Comparison of the GRACE, HEART and TIMI score to predict major adverse cardiac events in chest pain patients at the emergency department. Int J Cardiol 2017;227:656-61.
17. Van Den Berg P, Body R. The HEART score for early rule out of acute coronary syndromes in the emergency department: a systematic review and meta-analysis. Eur Heart J Acute Cardiovasc Care 2018;7:111-9.
18. Worster A, Sardo A, Eva K, Fernandes CM, Upadhye S. Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. J Emerg Nurs 2007;33:319-23.