

Evaluation of Provider Assessment of Clinical History When Using the HEART Score

Ravindra Gopaul¹, Robert A Waller¹, Ricci Kalayanamitra², Garrett Rucker³, Andrew Foy⁴

¹Department of Emergency Medicine, Penn State Hershey Medical Center, Hershey, PA, USA; ²Department of Emergency Medicine, Staten Island University Hospital, New York, NY, USA; ³Penn State Milton S. Hershey College of Medicine, Hershey, PA, USA; ⁴Department of Internal Medicine, Division of Cardiology, Penn State Hershey Medical Center, Hershey, PA, USA

Correspondence: Ravindra Gopaul, Email rgopaul@pennstatehealth.psu.edu

Objective: The HEART Score is a clinically validated risk stratification tool for patients with chest pain. Using five parameters (History, Electrocardiogram, Age, Risk factors, and Troponin), this instrument categorizes patients as low, moderate, or high risk for major adverse cardiac events within six weeks after evaluation. Of these parameters, History is the most subjective, as providers independently assign their level of clinical suspicion. Overestimation of history, and ultimately the HEART Score, can result in increased resource utilization, expense, and patient risk. We sought to evaluate bias in provider assessment of history when determining the HEART Score.

Methods: Emergency medicine (EM) and Cardiology providers received surveys with one of two versions of clinical vignettes randomized at the question level and were asked to estimate the history component of the HEART Score. Vignettes differed by age, risk factors, sex, and socioeconomic status (SES), but both versions should have received the same score for history. Statistical analysis was then used to assess differences in history assessment between vignettes.

Results: Of the 884 responses analyzed, most providers overestimated the historical portion of the HEART Score when assessing risk factors, patient distress, age, and lower SES. Many underestimated history with knowledge of a previous negative stress test. When controlling for specialty, the universal theme was overestimation by EM providers and underestimation by cardiologists. Despite the presence of hypertension, gender differences, and the appearance of mild distress, cardiologists were more likely to correctly estimate history compared to EM providers. SES consideration generally led to an underestimation of history by cardiologists. These findings were all statistically significant.

Conclusion: Our study demonstrates that both EM and cardiology providers overestimate history when considering prognosticators that are frequently viewed as concerning. Further education on proper usage of the HEART Score is needed for more appropriate scoring of history and improved resource allocation for hospital systems.

Keywords: HEART score, physician, provider, history

Introduction

Background

Chest pain is one of the most common chief complaints for which patients present to health-care providers. Differentiating between potentially lethal forms of chest pain, such as acute coronary syndrome (ACS), and other benign etiologies is paramount for all practitioners but especially for those in the Emergency Department (ED). To assist with this distinction, risk stratification tools such as TIMI, GRACE, and the HEART Score have been developed. Several studies have compared the HEART Score to TIMI and GRACE scores and found that the HEART Score better discriminated those with and without Major Adverse Cardiac Events (MACE: myocardial infarction, percutaneous coronary intervention [PCI], coronary artery bypass graft, or death) and was a better tool for identifying low-risk patients.^{1,2} The HEART Score utilizes five parameters (History, Electrocardiogram [ECG], Age, Risk factors, and Troponin) to categorize a patient's risk as low, moderate, or high for MACE within 6 weeks following presentation.³ The original HEART Score study by Backus et al found a 2.5% risk of MACE with an overall score of 0–3 (low risk),

a 20.3% risk with a score of 4–6 (moderate risk), and a 72.7% risk with a score ≥ 7 (high risk).³ In 2010, Backus et al performed a multicenter retrospective validation study showing a 0.99% risk of MACE with a HEART Score of 0–3, 11.6% risk with a score of 4–6, and 65.2% risk with a score ≥ 7 .⁴ A prospective multicenter validation study was done in 2013 with very similar results.⁴ Given these findings, Backus et al recommended that low-risk patients be discharged from the ED, moderate-risk patients be admitted for clinical observation, and high-risk patients be considered for early invasive strategies.³

There have since been several studies evaluating the interrater reliability of the HEART Score among clinicians with an overall strong level of reliability, though History and ECG had the lowest degree of reliability of the individual components of HEART Score.^{5,6} A multicenter observational study by Parenti et al found good interrater reliability among high- and low-risk groups, but only moderate agreement for the intermediate-risk group.⁶ History was noted to have the worse agreement with a kappa statistic of 0.37, suggesting that additional efforts should be made at establishing a more objective assessment of that parameter.⁶ Of the 5 HEART Score parameters, history determination is the most subjective and relies on the provider to assign their level of clinical suspicion for MACE as highly suspicious (2 points), moderately suspicious (1 point), or slightly suspicious (0 points). Based on the original HEART Score study, a highly suspicious history constitutes “specific” symptoms for ACS, such as exertional pain, diaphoresis, nausea, vomiting, and radiation. Moderately suspicious constitutes a mix of specific and non-specific factors (ie, chest pain that is sharp, pleuritic, reproducible, and positional). Slightly suspicious is assigned with only non-specific factors, as defined in the HEART Score paper by the “absence of specific elements in terms of pattern of the chest pain, onset and duration, relation with exercise, stress or cold, localisation, concomitant symptoms and the reaction to sublingual nitrates”.³

Importance

In the ED, physicians are often presented with tremendous amounts of information, both verbal and nonverbal, pertaining to patient appearance, demographics, cardiac risk factors, etc., in a relatively short period, making it difficult to objectively assess chest pain characteristics in the absence of other components of the HEART Score. Overestimation of patient history, and ultimately the HEART Score, can result in increased resource utilization (ie, stress tests, catheterization, or consults) and health-care costs. From a patient safety standpoint, overestimating the HEART Score may expose patients to potentially invasive tests that may not have been warranted. Consequently, accurate risk assessment is essential, and thus far, there have been no studies evaluating provider biases when assigning suspicion to patient history using the HEART Score.

Goals of This Investigation

As previously mentioned, multiple prior studies have shown low inter-rater reliability for the history portion, with kappa values between 0.13 and 0.73.⁷ Our goal is to determine how well emergency medicine (EM) providers (attending, residents, advanced practice providers) and cardiology attendings/fellows estimate the history component of the HEART Score using definitions from the original HEART Score study as well as assess bias from distracting risk factors.

The specific aims of this study are summarized in the following paragraph. Primarily, we sought to determine whether certain patient characteristics or risk factors would bias providers to either overestimate or underestimate the historical parameter of the HEART Score even though many of these factors are accounted for elsewhere in the tool. Variables that we hypothesized may result in provider overestimation include classic prognosticators for chest pain such as male gender, older age, concerning past medical history, and distressed patient appearance/behavior. At the same time, we anticipated that knowledge of a prior negative cardiac workup would bias providers to underestimate the history component of the HEART Score. We then evaluated differences in the HEART Score history assessment between EM providers and Cardiologists. Secondarily, we looked for variances in provider experience (ie, years of clinical practice), academic vs community setting, attendings vs learners, physicians vs non-physician providers, and males vs females in their assessment of history using the HEART Score.

Methodology

Study Design and Setting

Our primary purpose was to evaluate how well providers in Emergency Medicine and Cardiology utilize the HEART Score as it was originally intended to be used. This goal required assessment of a provider's ability to make a risk stratification based on chest pain description (ie, the history portion of the HEART Score) without being influenced by other elements (risk factors, patient age, patient sex, etc.). The standard for the risk assessment of the history portion that we used was that provided by the original HEART Score study, as alluded to above.³ A history containing only specific features of cardiac chest pain as suggested by "pattern of the chest pain, onset and duration, relation with exercise, stress or cold, localisation, concomitant symptoms and the reaction to sublingual nitrates" was given a score of 2 points (highly suspicious).³ A history completely lacking any specific elements of cardiac chest pain was given a score of zero (slightly suspicious). If the history contained a combination of specific and non-specific elements, a score of 1 point was given (moderately suspicious). As our primary intent was to evaluate proper usage of the HEART Score rather than to evaluate the HEART Score itself, we considered these definitions to be objective and the gold standard. With this in mind, a board-certified EM and cardiology attending developed a survey with a total of 21 questions (see [Supplemental Materials](#) for full survey). Of these, questions 1–8 dealt with demographic information of the participants (age, gender, specialty, academic vs community practice setting, number of years in practice, attending vs fellow/resident/advanced practice provider status, current year in training, and in which state the provider practices). Question 9–11 pertained to HEART Score utilization (Do you use it, is it useful for risk stratification, and which component of the HEART Score do you find most subjective?) to assist with data integrity, as those who do not regularly use the score were thought more likely to misuse it. Finally, questions 12–20 presented a series of clinical vignettes that were written to satisfy the HEART Score definition of slightly, moderately, and highly suspicious. Institutional Review Board (IRB) approval was obtained from the Penn State Health Human Subjects Protections Office [Study00014146]. Informed consent of participants was obtained in electronic surveys. Guidelines documented in the Declaration of Helsinki were followed.

Selection of Participants

Our distribution was based upon email lists that included EM attendings, residents, and advanced practice providers (APP), and cardiology attendings and fellows. We unfortunately did not have access to an email list for cardiology APPs. The survey recruitment email was sent to approximately 35,000 providers across both specialties. Institutional review board approval was sought and the study was deemed exempt.

Interventions and Exposures

All clinical scenarios were assigned a "correct" score (slightly, moderately, and highly suspicious) in alignment with definitions provided in the original HEART Score study prior to survey distribution. A board-certified emergency physician and cardiologist critically evaluated each scenario based on definitions for history estimation provided in the original HEART Score study and were in agreement. We then added distracting elements of differing ages, ethnicity, patient sex, risk factors, socioeconomic status, and clinical appearance of the patient to assess a provider's likelihood of deviating from the HEART Score standards for history estimation ([Figure 1](#)). These elements are either accounted for by other categories of the HEART Score (ie, age and risk factors) or not addressed at all (ie, sex, socioeconomic status, race). In either case, these factors were not intended to have an impact on the score assigned to clinical history, as it should be based solely on chest pain characteristics. Using these distractors, we created two different versions of the nine clinical vignette questions. EM and cardiology providers were randomized to receive one of the two clinical vignette questions for each of questions 12–20 in the survey. Randomization was performed using Survey Monkey and surveys were distributed using the same software.

Measurements and Outcomes

The primary outcome measure was a 3-point ordinal-scale value. If a participant responded with a risk assessment higher than our estimated gold standard based on the HEART Score definition, then they were said to have overestimated risk. If their response matched our gold standard estimate, they were said to have accurately assessed risk. If they provided

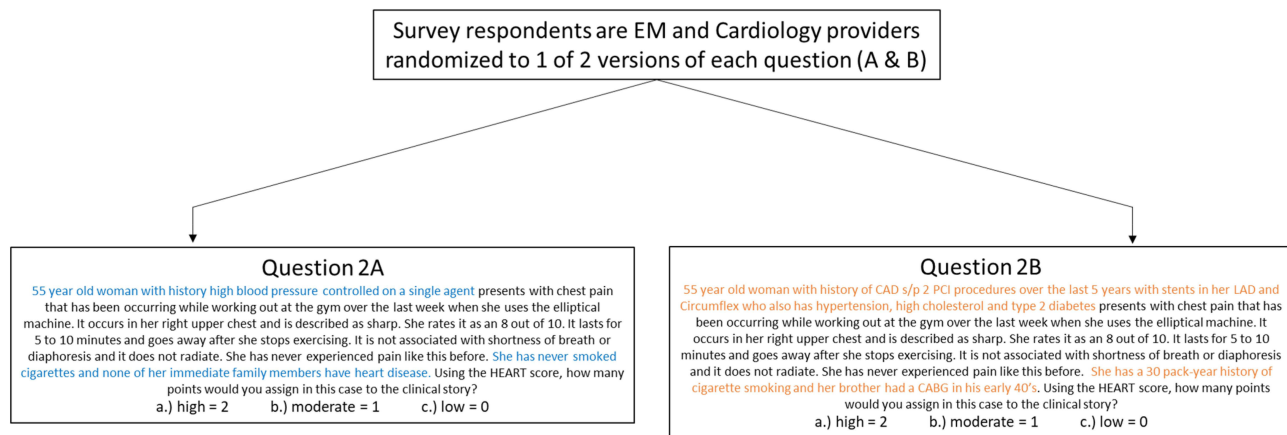


Figure 1 Patients were randomized to 1 of 2 clinical vignettes, which differed by risk factors, demographics, socioeconomic status, patient appearance, or past medical history but were otherwise unchanged with respect to chest pain characteristics.

a lower risk assessment, they were said to have underestimated. Thus, a participant's responses were graded against the gold standard estimate of history based on definitions provided in the original HEART Score study.

Analysis

The responses to the two surveys were compared using contingency table analysis. The Chi-square statistic was calculated to assess differences in responses between the two clinical scenarios. Subgroup analyses were performed on demographic variables such as age, sex, and years in practice. Differences between surveys were characterized using contingency table analysis, and significance levels were determined by chi-square statistic, Fisher's Exact Tests, and other non-parametric tests, as needed. Data were analyzed using SAS, version 9.4, by a dedicated team of statisticians.

Additional data was collected for analysis by asking participants to provide their age, sex, specialty, number of years in practice or training, state or country in which they practice, academic versus community practice setting, and which component of the HEART Score they find the most subjective. Due to the likelihood of misapplication of the HEART Score if not frequently utilized, providers were questioned whether they use this tool in clinical practice and if they think it is useful as a prognosticator.

Results






Of the 35,000 emails that were sent, approximately 35% reached non-working addresses. We received a total of 1330 responses from both EM and cardiology providers, with a response rate of approximately 6%. In order to maintain the integrity of our results, we did not analyze the responses of participants who stated that they do not regularly employ the HEART Score. As such, we excluded 446 responses and examined the remaining 884. Not all participants answered each question and therefore the number of individual responses varies per question. Of the respondents, most were EM providers, community physicians, attending physicians, males, and physicians, and the mean number of years in practice was 16 (standard deviation of 11 years) (Table 1). We utilized this mean of 16 years of clinical practice as a delineation between those with less experience (less than 16 years) and those with greater experience (more than 16 years).




Regarding the risk factors and variables assessed, most providers overestimated the historical portion of the HEART Score (Table 2). This tendency applied in considering a history of coronary artery disease (CAD) status post PCI as compared with a history of hypertension [30.03 (23.74, 36.31)], significantly distressed appearance vs mildly distressed [10.16 (4.71, 15.58)], older vs younger patients [35.16 (29.68, 40.62)], and lower socioeconomic status [5.74 (0.34, 11.16)]. Most practitioners also underestimated history if they knew that a patient had a previous negative stress test [10.04 (4.58, 15.49)]. Lastly, though not statistically significant, there were marginal trends for providers to overestimate for male vs female gender [6.72 (-3.91, 17.35)] and for Caucasian vs African American ethnicity [11.69 (-0.46, 23.84)].

Table 1 Demographics of Study Participants

Participant Demographics	Total Number	Percentage
Specialty		
Emergency Medicine	746	84.9%
Cardiology	133	15.1%
Sex		
Males	673	76.4%
Females	208	23.6%
Educational status		
Attendings	735	89.4%
Learners	87	10.6%
Years of Experience		
Less than 16 years	461	52.3%
Greater than 16 years	420	47.7%
Work setting		
Academic	381	43.4%
Community	497	56.6%
Physician vs nonphysician		
Physician	821	93.2%
Nonphysician	60	6.8%

Table 2 Summary of Performance of All Providers When Asked to Assign the Historical Portion of the HEART Score in the Presence of the Variables Listed Below Compared with the Correct Suspicion Score

Risk Factors Assessed	Tendency of Respondents	P value
CAD s/p PCI vs HTN		<0.0001
Significant vs Mild Distress		<0.0001
Older vs Younger		<0.0001
Lower vs Higher SES		0.0003
Pervious negative stress test vs No prior workup		<0.0001

Notes: : underestimated, : accurately estimated, : overestimated.

Abbreviations: HTN, hypertension; CAD s/p PCI, coronary artery disease status post percutaneous coronary intervention; SES, socioeconomic status.

When compared directly, EM tended to overestimate and cardiology was inclined towards underestimation of the historical portion of the HEART Score (Table 3). EM overestimated in the following areas: patients with a history of HTN [15.99 (5.59, 26.40)], differences in patient sex [male, 16.2 (1.93, 30.46); female, 18.95 (7.32, 30.58)], mildly distressed appearance [13.25 (1.09, 25.39)], and Caucasian ethnicity [11.69 (-0.46, 23.84)]. Cardiology underestimated

Table 3 Summary of Performance of Emergency Medicine Providers Vs Cardiologists When Evaluated for the Impact of Risk Factors on History Portion of the HEART Score

	Emergency Medicine	Cardiology	P value
HTN	↑	↓	<0.0001
Sex	↑	▬	<0.0008
Mildly distressed	↑	▬	<0.0001
Caucasian	↑	▬	0.0007
SES	▬	↓	<0.0001

Notes: ↓: underestimated, ▬: accurately estimated, ↑: overestimated.

Table 4 Comparison of the Performance of the Various Demographics in Assessment of the Historical Portion of the HEART Score

Groups Compared	Overestimated	Accurate	Underestimated
Community vs Academic	Community	Academic	
Attendings vs Learners	Attendings	+/-	Learners
Physicians vs Non-physicians		Physicians	Non-physicians
Clinical Experience	More experience	+/-	Less experience
Male vs female	Females	Males	

Note: +/- denotes similar performances between the compared groups.

in the following areas: patients with a history of HTN [12.51 (2.19, 22.83)] and regardless of socioeconomic status [higher SES, 27.93 (14.56, 41.30); lower SES, 26.45 (14.42, 38.47)].

Using the demographic data collected from the participants, we also identified several general trends among the five compared groups (Table 4). Firstly, academic providers were more accurate in utilizing the HEART Score while community providers tended to overestimate. Secondly, in examining attending physicians vs learners, we found that the former were more likely to overestimate while learners were more prone to underestimation. Thirdly, when evaluating physicians and non-physicians, the former were more accurate and the latter more likely to underestimate. Fourthly, regarding years in clinical experience, those with more years were more likely to overestimate, and those with less experience underestimated more frequently. Lastly, when male providers were compared with their female counterparts, the former trended toward greater accuracy while the latter was inclined to overestimate.

Discussion

The HEART Score has been shown to be a reliable risk stratification tool for chest pain in the ED but improper application could lead to unnecessary resource utilization and potential patient harm. We sought to assess proper usage of the HEART Score according to the definitions provided by the tool's developers and found that risk factors appear to bias responses.

Although of secondary interest in this paper, we would like to comment first on the findings identified using the participants' demographic information. When comparing academic vs community providers, it may be that the latter was more likely to overrate a patient's risk due to fewer resources being immediately available (ie, immediate cardiology consultation or extended stress testing accessibility) and a desire to err on the side of caution. Alternatively, it may be that academic providers are more likely to be aware of the details of the studies on which prognosticating tools are based,

given that they are involved in education and may need to provide such data to their learners. In evaluating attending physicians vs learners, the former were more likely to overestimate, which may be due to the liability that they assume as the physician of record or due to their increased experience compared to learners, who may not yet know what they do not yet know. This phenomenon may also account for the differences in physician and non-physician providers and the distinctions seen between practitioners with more experience than those with less. Variances among the latter comparison may also be related to comfort with using the HEART Score, as it would have likely been developed after many of the more experienced providers were practicing independently, while the new providers may have been trained to use it during their graduate education. Concerning the differences seen between male and female providers, there is no readily available explanation and further investigation is needed.

Another ancillary comparison of our project was that of the EM vs cardiology providers. As demonstrated, EM providers were more likely to overestimate history while cardiologists were more likely to underestimate or more accurately assess clinical scenarios. This phenomenon may be due to fundamental differences in their approaches to medicine. EM physicians are trained to assume the worst and to work in reverse order by ruling out potentially lethal diagnoses one at a time until the patient is deemed safe for discharge or determined to need admission. On the other hand, cardiologists have a background in internal medicine and may be able to approach the patient in a more measured and objective manner, presuming that fatal causes have already been worked up and therefore a more thorough, albeit time-consuming, investigation is in order. Alternatively, it may be that EM physicians do not want to accept sole liability, as EM is a general field of training, and therefore EM providers are more likely to overestimate when uncertain, while cardiologists are more accustomed to making the final diagnosis as specialists and therefore feel more comfortable assigning lower risk when justified. Additional investigation into the reasons behind these differences is necessary to avoid conjecture.

Our primary goal was to determine if providers were biased to either overrate or underestimate a patient's history based on variables that are already accounted for in other sections of the HEART Score, such as HTN or CAD s/p PCI, or that are commonly viewed as poor prognosticators, such as male sex or lower SES. As suspected, most providers, regardless of specialty training, were more likely to overestimate based on past cardiac history, concerning patient appearance, older age, and lower SES. Similarly anticipated was the finding that both EM practitioners and cardiologists underestimated history if they had knowledge of patients having had a previous negative stress test. Identification of these biases indicate the need for better education in the proper use of the HEART Score, as providers are at risk for increasing the suspicion assessment of the historical portion of the HEART Score and thereby elevate a patient's overall risk into a category that requires further resource utilization or testing (ie, stress testing, cardiology consultation, catheterizations, observation/admissions, etc.). Additional education into the original descriptions of low, moderate, and high suspicion chest pain defined in the original HEART Score study may be as simple as describing these on commonly used online or mobile applications. We hope that by having identified this area of weakness within the use of this nearly ubiquitous prognosticating tool, providers may reevaluate their methods and follow the guidelines of the original HEART Score study to more accurately assign risk to patients with chest pain.

Limitations

Our project aimed to identify provider bias, and overall, we feel that our project did demonstrate this. However, there were some limitations in subgroup analyses due to the small number of participants from certain categories, including cardiologists (which only accounted for 15% of the total study number), fellows/residents (10.6%), females (23.6%), and nonphysician providers (6.8%, see [Table 1](#)). Thus, while statistically significant differences were found, a better-balanced study group would be ideal to confirm that these variances are generalizable. Secondly, we had a low response rate for an online survey at 6%, as the average is between 35% and 40%.⁸ Greater participation would certainly provide more certainty as to any conclusions. In addition to sampling and response bias which is inherent to surveys, another limitation was the use of clinical vignettes in our study which may have varying interpretations by respondents. Lastly, our survey did not assess the implications of the bias identified. Therefore, we cannot conclude that an overestimation of the historical portion of the HEART Score indeed leads to any differences for the patient (ie, resource utilization, exposure to invasive testing). To date, no evaluations have been performed to determine how frequently elevated historical suspicion leads to different dispositions—although logic indicates that it likely occurs often—and further research in this area is indicated to prove such an association.

Conclusion

Our study demonstrates that both EM and cardiology providers overestimate the history component of the HEART Score when considering prognosticators that are frequently viewed as concerning (risk factors, patient distress, age, and lower SES). Further education into proper usage of the HEART Score is needed for more appropriate scoring of history and better resource allocation for hospital systems.

Acknowledgments

Virtually presented and selected “Best of the Best” at the Pennsylvania American College of Emergency Physicians (PACEP) William H. Spivey Resident Research Competition on April 9, 2021 at the virtual PACEP conference. Also virtually presented at SAEM 2021 during the Lightning Oral presentations on May 14, 2021. Presented at AAEM on June 20, 2021 in person in St. Louis, MO. Presented orally at ACEP21 Research Forum.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Penn State Milton S. Hershey Medical Center Department of Emergency Medicine and Junior Faculty Development Program, Penn State Health.

Disclosure

RAW, RK, GR, AF, and RG report no conflicts of interest in this work.

References

1. 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;227:656–661. doi:10.1016/j.ijcard.2016.10.080
2. Nieuwets A, Poldervaart JM, Reitsma JB, et al. Medical consumption compared for TIMI and HEART score in chest pain patients at the emergency department: a retrospective cost analysis. *BMJ Open.* 2016;6:e010694. doi:10.1136/bmjopen-2015-010694
3. Six AJ, Backus BE, Kelder JC. Chest pain in the emergency room: value of the HEART score. *Neth Heart J.* 2008;16(6):191–196. doi:10.1007/bf03086144
4. 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(3):2153–2158. doi:10.1016/j.ijcard.2013.01.255
5. Niven WGP, Wilson D, Goodacre S, Robertson A, Green SJ, Harris T. Do all HEART Scores beat the same: evaluating the interoperator reliability of the HEART Score. *Emerg Med J.* 2018;35(12):732–738. doi:10.1136/emered-2018-207540
6. Parenti N, Lippi G, Reggiani MLB, et al. Multicenter observational study on the reliability of the HEART score. *Clin Exp Emerg Med.* 2019;6(3):212–217. doi:10.15441/ceem.18.045
7. Green SM, Schriger DL. A methodological appraisal of the HEART score and its variants. *Ann Emerg Med.* 2021;78(2):253–266. PMID: 33933300. doi:10.1016/j.annemergmed.2021.02.007
8. Cunningham CT, Quan H, Hemmelgarn B, et al. Exploring physician specialist response rates to web-based surveys. *BMC Med Res Methodol.* 2015;15:32. doi:10.1186/s12874-015-0016-z

Open Access Emergency Medicine

Dovepress

Publish your work in this journal

The Open Access Emergency Medicine is an international, peer-reviewed, open access journal publishing original research, reports, editorials, reviews and commentaries on all aspects of emergency medicine. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/open-access-emergency-medicine-journal>