



HHS Public Access

Author manuscript

Acad Emerg Med. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:

Acad Emerg Med. 2019 May ; 26(5): 552–555. doi:10.1111/acem.13665.

Inter-rater Reliability of the HEART Score

Colin A. Gershon, MS, MPH¹, Annick N. Yagapen, MPH, CCRP¹, Amber Lin, MS¹, David Yanez, PhD¹, and Benjamin C. Sun, MD, MPP¹

¹Center for Policy Research-Emergency Medicine, Department of Emergency Medicine, Oregon Health and Sciences University, Portland OR

Abstract

Background—The HEART score is a risk stratification tool for suspected acute coronary syndrome and contains several subjective components. A single previous study found good inter-rater reliability. Our objective was to assess the inter-rater reliability of the HEART score in an external prospective cohort.

Methods—We prospectively collected paired, independent physician ratings of the HEART score for patients > 20 years of age presenting to the emergency department with chest pain for which an ECG and troponin were ordered. Two emergency physicians independently provided HEART scores for each unique patient. The primary outcome, the HEART score, was dichotomized by low risk (0–3) vs non- low risk (4–10). Additional outcomes included the HEART score across the entire scale (0–10) and subcomponents of the HEART score (e.g., history, electrocardiogram, risk factors; score of 0–2 for each). We calculated kappa statistics and percent agreement for all outcomes.

Results—We collected paired physician HEART score ratings on 311 patients from October 2017 to April 2018. The mean HEART score was 3.5 (SD 1.9). About half (49.2%) of our patients had a HEART score of ≤ 3, and 50.8% had a HEART score > 3. The kappa score for “low risk” (HEART ≤ 3) was 0.68 (95%CI: 0.60 – 0.77). There was 84.2% agreement between physicians on this variable.

Conclusions—Our study demonstrates there is substantial inter-rater reliability among emergency department physicians in identifying patients at low risk of acute coronary syndrome using the HEART score.

Keywords

HEART score; inter-rater reliability; chest pain; kappa

Corresponding Author: Colin A. Gershon, MS, MPH, gershon@ohsu.edu, Phone: 503-494-1193, Fax: 503-494-4640, Address: Mail Code CR114, 3181 SW Sam Jackson Park Rd, Portland, OR 97239.

Prior Presentations: None

Conflicts of Interest: None

Introduction

Background

Acute coronary syndrome (ACS) is the leading cause of worldwide mortality and morbidity.¹ The evaluation of suspected ACS typically occurs in an emergency department (ED) setting and accounts for over seven million annual ED visits in the United States.² Risk stratification with a careful history, examination, 12-lead electrocardiogram (ECG), and cardiac biomarkers is the cornerstone of the ED evaluation³, but can be challenging due to atypical presentations of ACS.

Published risk scores may help to supplement clinical judgment in the assessment of patients with chest pain. The HEART score, derived in undifferentiated ED chest pain patients, has been validated in multiple settings^{4,5} and appears to have superior test characteristics compared to other risk scores^{6,7}. Values of the HEART score range from 0–10, and scores less than or equal to 3 identify very low risk patients that potentially can be discharged safely without further immediate cardiac testing.⁷

Importance

Two of the HEART score components require physician interpretation: an assessment of how suggestive the patient's presentation is for ACS (the "history"), and an interpretation of the ECG. There may also be variability in assessing the number of ACS risk factors. The inter-rater reliability of the HEART score appears to have been assessed in only one study to date⁸, in which "substantial" ($0.6 < \text{kappa} \leq 0.8$) inter-rater reliability was found. Inter-rater reliability is important because variations in scoring may affect management, particularly across the lowest risk threshold (HEART ≤ 3).

Goal of this Study

We prospectively evaluated the inter-rater reliability of the overall HEART score as well as the HEART score components of history, ECG, risk factors, and troponin individually. Our primary hypothesis was that the HEART score is reliable ($\text{kappa} > 0.6$) for identifying low-risk (HEART score ≤ 3) adults presenting to the emergency department with chest pain.

Methods

Study Design

This was a prospective cohort study of pairs of physicians who independently evaluated patients and then independently calculated HEART scores on each patient. No patient identifying information was collected. A full waiver of consent and HIPAA authorization was granted by the university IRB.

Study Setting and Population

We identified pairs of attending and second- or third-year emergency medicine resident physicians who were working together in a single university ED. To mitigate against bias from limited clinical experience, we excluded pairs that included a first-year emergency medicine residents or rotating non-emergency medicine residents.

Study Protocol

The unit of analysis was paired HEART ratings. Eligible encounters were for adult patients (age >20 years) with a chief complaint of chest pain, for whom the treating team had ordered a troponin and 12-lead ECG, and who were seen by two emergency department physicians. Research assistants screened for eligible encounters between 7a-11p, 7 days per week.

Measures

For each eligible physician pair and patient encounter, we collected data on all elements of the HEART score. Research assistants abstracted objective data on age from the electronic medical record.

The attending and resident physicians independently completed evaluations of the other elements of the HEART score, including History, ECG interpretation, Risk factors, and Troponin value. The number of troponin tests for each patient was at the discretion of the clinical team but the HEART score was based on the first troponin. Each physician was blinded to the ratings of the other paired physician. The physicians rated the elements of the HEART score on a standardized form shortly after the patient encounter.

From these data, we calculated overall HEART scores and subscale scores for each physician rating.

Data Analysis

We calculated the kappa statistic and total agreement for dichotomized low- vs. high risk HEART rating (3) and for the ordinal sub-scale components (History, ECG, Risk factors and Troponin), we used weighted kappa and total agreement statistics. Kappa statistics can be used when calculating inter-rater agreement for a dichotomous variable and weighted kappa statistics extend this to ordinal variables with greater than 2 categories. We calculated weighted kappa statistics and intra-class correlations (ICC) across all possible values (0–10) of the HEART score.⁹ ICC was calculated as the ratio of variance components produced by a one-way random effects model and provides a way to quantify the degree of consistency between raters on the values of the HEART score by accounting for both the correlation and agreement¹⁰. As a sub-analysis, we compared agreement statistics between 2nd and 3rd year residents using a two-sample test of binomial proportions.

Our sample size calculation was based on a traditional Cohen's kappa. We assumed 60% of each raters' classifications would be for low risk scores (i.e., HEART 0–3), and a modest sized kappa, $\kappa = 0.5$, for a 0.05 alpha-level test. We required a sample size of N=300 patients to attain statistical power of 80 percent. Data entry and management were performed with RedCAP. Analyses were performed using SAS 9.4 (Cary, NC).

Results

We collected paired physician assessments on 311 patients from October 2017 through April 2018. Our participants were a mean age of 55.8 years old (SD 14.6) and 173 (55.6 %) were male. The mean HEART score was 3.5 (SD 1.9). Approximately half (49.2%) of our patients had a HEART score of 3.

We calculated a kappa score for “low risk” (HEART ≤ 3) of 0.68 (95% CI: 0.60 – 0.77). There was 84.2% agreement between physicians on this dichotomized scale. Weighted kappa statistics for the individual HEART subscales ranged from 0.46 to 0.83. The weighted kappa statistic for the full scale of the HEART score (0–10) was 0.68 (95% CI: 0.64 – 0.72). The intraclass correlation (ICC) was 0.86 (95% CI: 0.81–0.92) indicating good to excellent reliability.

Of the five components of the HEART score, only the history component was significantly different between the attending physicians & 2nd year residents vs. the attending physicians & 3rd year residents (p-value = < 0.01), where the attending & 2nd year resident pairs had greater inter-rater reliability than the attending & 3rd year resident pairs.

Discussion

The HEART pathway has been validated as effective in identifying patients with chest pain who can be safely discharged safely from the ED without further immediate cardiac testing.

Risk stratification tools such as the HEART score will have greater clinical utility if they can demonstrate robust inter-rater reliability. The only study known to these authors to have examined inter-rater reliability of the HEART score showed substantial inter-rater reliability ($k > 0.6$) for classifying low risk patients (HEART ≤ 3).⁸ Our study also found substantial ($k = 0.68$) inter-rater agreement on this variable.

Our study was the first to examine the inter-rater reliability of four subscales of the HEART score (History, ECG, Risk factors, and Troponin). We found that the History and ECG components had lower agreement ($k = 0.52$ and 0.46 , respectively) compared to Risk factors and Troponin components ($k = 0.67$ and 0.83 , respectively). The History component was originally scored using a list of pre-defined, specific chest pain characteristics that were categorized as either suggestive or not suggestive of ACS.^{4,5,8} In our study, physicians scored the History as they do in clinical practice – without reference or strict adherence to such a list. Some of the disagreement in the score for History, therefore, may be attributable to variability in physician conception of what constitutes chest pain suspicious for ACS.

The agreement on ECG interpretation was comparable to other studies assessing interrater reliability of ED physician ECG reads. The fact that the agreement for troponin (an objective number that should have been easily classified) was less than 100% suggests that the physicians did not consistently score the variables fully utilizing all of the tools (electronic health record, HEART score calculator) available to them.

Limitations

Our study had several limitations. First, while the physicians were blinded to each other’s HEART scores, they may have discussed some patients’ clinical picture prior to estimating their HEART score. This may lead to a bias toward higher levels of agreement.

Second, we utilized the kappa score to evaluate inter-rater reliability given strong precedent for this test in the health care literature, with kappa > 0.6 indicating substantial inter-rater

reliability. However, many clinical risk tools for chest pain in the emergency department setting aim for a negative predictive value of > 99% - effectively, a miss rate of < 1% for patients with a major cardiac event.

Third, our study examined patients who presented to the ED between 7 a.m. – 11 p.m. Patients who presented to the ED with chest pain between 11 p.m. and 7 a.m. may have different demographic and epidemiologic profile compared to patients at other times of day.

Finally, this analysis was performed at a tertiary care academic medical center ED. Our results may not be generalizable to the primary care, urgent care, or inpatient setting.

Conclusion

In conclusion, we found substantial inter-rater reliability of the HEART score to identify low-risk patients. Our findings support the adoption of the HEART score into emergency department chest pain management algorithms.⁸ Given the importance of accurate cardiovascular risk stratification in the ED, further studies should elucidate how higher concordance between providers might be achieved.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

CAG, AY, DY, and BCS designed the study. BCS obtained funding for this study. AY and AL were responsible for data management and data analysis. CAG drafted the manuscript. All authors contributed substantially to manuscript revisions. CAG takes responsibility for the paper as a whole. AY and AL had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors approved the final report for submission.

Funding Sources/Disclosures:

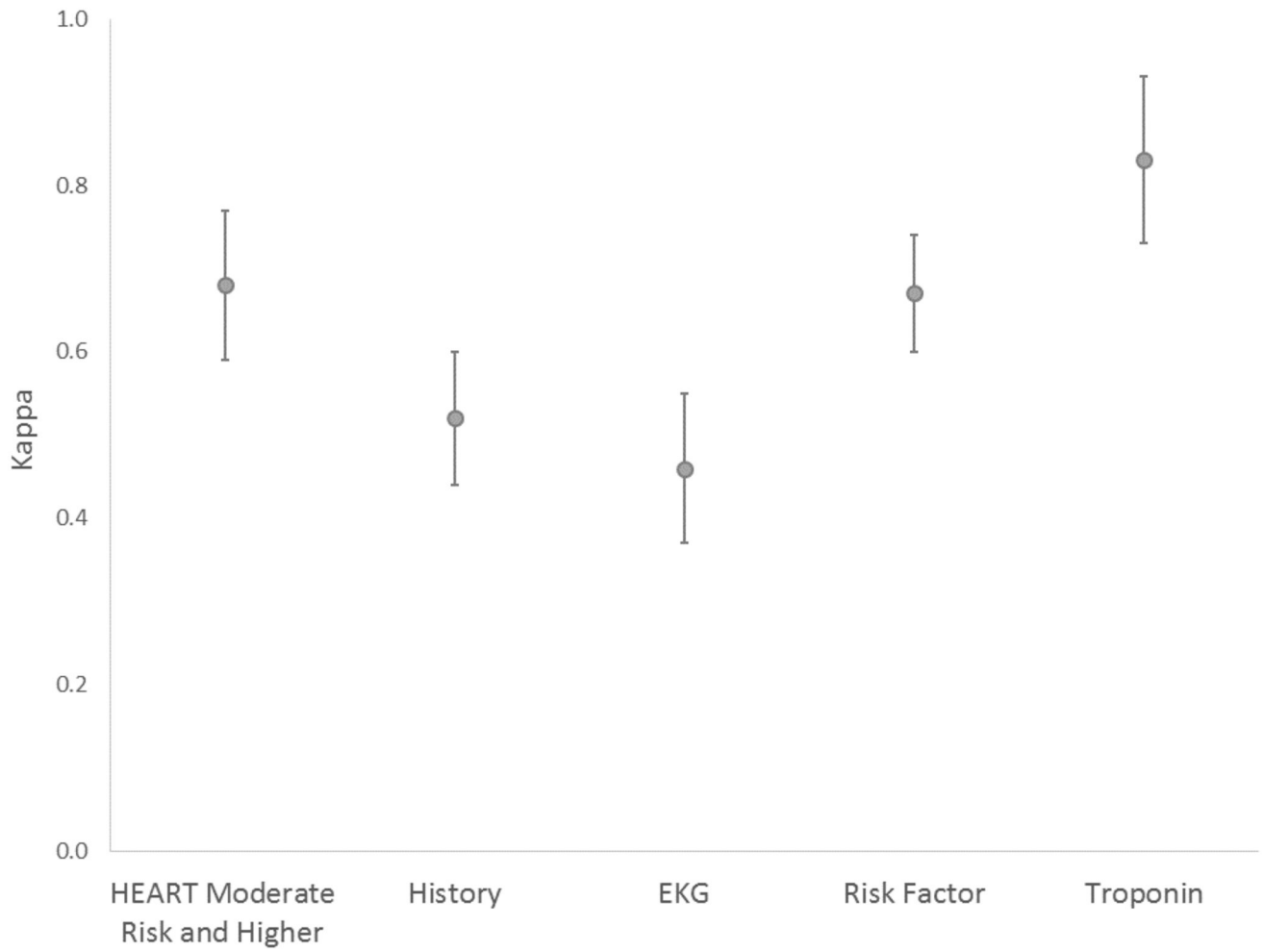
This study was supported by National Institutes of Health (NIH) grant R01HL111033.

The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. The contents do not necessarily represent the official views of the National Institutes of Health.

References

1. Vedanthan R, Seligman B, Fuster V. Global perspective on acute coronary syndrome: a burden on the young and poor. *Circ Res* 2014;114:1959–75. [PubMed: 24902978]
2. CDC. National Hospital Ambulatory Medical Care Survey: 2010 Emergency Department Summary Tables 2010.
3. Amsterdam EA, Wenger NK, Brindis RG, et al. 2014 AHA/ACC Guideline for the Management of Patients with Non-ST-Elevation Acute Coronary Syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014;64:e139–e228. [PubMed: 25260718]
4. 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. *Critical pathways in cardiology* 2013;12:121–6. [PubMed: 23892941]

5. Backus BE, Six AJ, Kelder JC, et al. A prospective validation of the HEART score for chest pain patients at the emergency department. *International journal of cardiology* 2013;168:2153–8. [PubMed: 23465250]
6. Sun BC, Laurie A, Fu R, et al. Comparison of the HEART and TIMI Risk Scores for Suspected Acute Coronary Syndrome in the Emergency Department. *Critical pathways in cardiology* 2016;15:1–5. [PubMed: 26881812]
7. 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. *International journal of cardiology* 2017;227:656–61. [PubMed: 27810290]
8. Mahler SA, Riley RF, Hiestand BC, et al. The HEART Pathway randomized trial: identifying emergency department patients with acute chest pain for early discharge. *Circulation Cardiovascular quality and outcomes* 2015;8:195–203. [PubMed: 25737484]
9. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–82. [PubMed: 23092060]
10. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 1979;86:420–8. [PubMed: 18839484]



Point estimates and 95% confidence intervals for agreement statistics: kappa statistic for HEART (moderate risk and higher, 4+) and weighted kappa statistics for history, EKG, risk factors and troponin

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript