Received: 5 January 2022 | Revised: 7 April 2022 | Accepted: 11 April 2022

DOI: 10.1002/emp2.12732

JACEP OPEN
JOURNAL OF THE AMERICAN COLLEGE OF EMERGENCY PHYSICIANS OPEN
A GLOBAL JOURNAL OF EMERGENCY MEDICINE
Open Access

WILEY

ORIGINAL RESEARCH

Cardiology

Interrater agreement of the HEART score history component: A chart review study

Alec J. Pawlukiewicz MD[1] | Matthew R. Geringer DO[2] | W. Tyler Davis MD[1] | Daniel R. Nassery DO[2] | Michael D. April MD, DPhil[1] | Matthew J. Streitz MD[1] | Jessica M. Hyams MD[1] | Alex W. Martin LPN[3] | Sadie A. Martin R.T.(R)[3] | Joshua J. Oliver MD[4]

[1]Department of Emergency Medicine, San Antonio Uniformed Services Health Education Consortium, San Antonio, Texas, USA

[2]Department of Internal Medicine, San Antonio Uniformed Services Health Education Consortium, San Antonio, Texas, USA

[3]University of Arizona, Tucson, Arizona, USA

[4]Leadership and Faculty Development Fellowship, Madigan Army Medical Center, 9040 Fitzsimmons Dr, Joint Base Lewis-McChord, Washington, USA

Correspondence
Joshua Oliver, MD, Madigan Army Medical Center, 9040 Fitzsimmons Dr, Joint Base Lewis-McChord, WA 98433, USA.
Email: joshua.j.oliver6.mil@mail.mil

Funding and support: By *JACEP* Open policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Abstract

Study objectives: This study investigated the interrater reliability of the history component of the HEART (history, electrocardiogram, age, risk, troponin) score between physicians in emergency medicine (EM) and internal medicine (IM) at 1 tertiary-care center.

Methods: We conducted a retrospective, secondary analysis of 60 encounters selected randomly from a database of 417 patients with chest pain presenting from January to June 2016 to an urban tertiary-care center. A total of 4 raters (1 EM attending, 1 EM resident, 1 IM attending, and 1 IM resident) scored the previously abstracted history data from these encounters.

The primary outcome was the interrater agreement of HEART score history components, as measured by kappa coefficient, between EM and IM attending physicians. Secondary outcomes included the agreement between attending and resident physicians, overall agreement, pairwise percent agreement, and differences in scores assigned.

Results: The kappa value for the EM attending physician and IM attending physician was 0.33 with 55% agreement. Interrater agreement of the other pairs was substantial between EM attending and resident but was otherwise fair to moderate. Percent agreement between the other pairs ranged from 48.3% to 80%. There was a significant difference in scores assigned and the subgroup in which there was disagreement between the raters demonstrated significantly higher scores by the EM attending and resident when compared to the IM attending.

Conclusion: This study demonstrates fair agreement between EM and IM attending physicians in the history component of the HEART score with significantly higher scores by the EM attending physician in cases of disagreement at 1 tertiary-care center.

Supervising Editor: Henry Wang, MD, MS.

KEYWORDS
agreement, emergency medicine, HEART score, history, internal medicine, interrater

1 | INTRODUCTION

1.1 | Background

The HEART (history, electrocardiogram, age, risk, troponin) score is a validated tool designed for risk stratification of patients presenting to the emergency department (ED) with chest pain (Appendix S1).[1] Previous analysis has shown significant variability of the HEART score, particularly the history component.[2] Prior studies have analyzed the interrater variability between emergency physicians and nurses and among attending physicians in emergency medicine (EM) and between attending and resident physicians in EM with agreement levels ranging from slight to substantial.[3–9] Further, comparisons of HEART scores between physicians and cardiologists in EM by Wu et al and physicians and researchers in EM by Soares et al also showed slight interrater reliability of the history component of the HEART score with kappa values of 0.14 and 0.10, respectively.[10,11]

1.2 | Importance

The disposition of patients with low-risk chest pain frequently involves interactions between physicians in EM and internal medicine (IM), yet no study has looked at the interrater variability among EM and IM physicians. Although most of the HEART score components are based upon simply defined objective criteria, the history component is subjective as it relies heavily on the patient's reported symptoms and the physician's interpretation of those symptoms.[12] A better understanding of the variability of the history component of the HEART score between EM and IM physicians is important to understand the impact of this subjectivity on patient disposition decisions.

1.3 | Goals of this investigation

Our study measured the interrater variability of the history component of the HEART score between physicians in EM and IM. We hypothesized that there would be marked variability between these 2 groups.

2 | METHODS

2.1 | Study design

We conducted a retrospective, secondary analysis of a previous single-center chart review study that produced a database of patients presenting with chest pain. The local institutional review board approved this study with the reference number C.2016.023d. We adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement in our research design, reporting, and analysis.[13]

2.2 | Setting and participant selection

The study setting was an urban tertiary-care medical and level I trauma center with an annual ED census of approximately 90,000 patient visits.[6] The patients selected for this study comprised encounters selected randomly from a previously compiled, de-identified database of 417 patients with chest pain presenting from January 1, 2016 to June 9, 2016.[14–18]

Inclusion criteria for this database were adults, aged 18 years or older, who presented with a chief complaint of "chest pain," "chest tightness," or "chest pressure" as identified on a systematic, retrospective search of our electronic medical record. The database did not include individuals who presented with chief complaints consistent with non-painful presentations of acute coronary syndrome without a component of some type of chest pain. Exclusion criteria for the database comprised (1) subjects less than 18 years of age; (2) those who did not have a recorded troponin or electrocardiogram; (3) those meeting criteria for an ST segment elevation myocardial infarction; (4) those without a subsequent, documented health care encounter in the sixth week after their ED visit; and (5) those without the necessary documentation from which to determine a HEART score. Additional details of the methods for database construction have been previously published and are consistent with the highest standards of chart review studies as described by Gilbert et al and Kaji et al.[5,19,20]

For the purposes of power calculation, the value for the expected Cohen's kappa was 0.3 or greater with a null hypothesis of no agreement between raters. Given an alpha of 0.05 and a power of 80%, we determined a need for a minimum sample size of 52 subjects for the detection of a minimum value of kappa coefficient of 0.3 with an assumption of an even distribution of scores within the cohort.[21]

From this database, we selected 60 patient encounters randomly for review by the scorers. To prevent scorers from being influenced by other aspects of the HEART score, we blinded the scorers to the patient's cardiovascular risk factors, electrocardiogram findings, laboratory results, disposition from the ED, and incidence of major adverse cardiac events (MACE) at 6-week follow-up. The data available for each scorer to evaluate included sex, chief complaint, and initial vital signs. We also provided scorers the data for characteristics of the pain and associated symptoms at the time of presentation. We provided the following pain characteristics: severity, time since onset, context of onset, gradual or sudden onset of the pain, persistence

The Bottom Line

Although the HEART score is an often-used and validated tool for risk stratification of chest pain patients, the more subjective components of the score remain largely variable among different specialties, potentially influencing its application. In this retrospective analysis of 60 patients admitted with chest pain, the authors found only fair interrater agreement (kappa=0.33) between ratings of physicians in emergency medicine and internal medicine of the history component of the HEART score.

and timing of the pain, quality, radiation, exacerbating factors, and ameliorating factors. We also presented the presence or absence of nausea and vomiting, diaphoresis, palpitations, cough, weakness, and dizziness.[5]

2.3 | Data collection

A total of 4 individuals (1 EM attending physician with 4 years' experience, 1 EM second-year resident, 1 IM attending physician with 7 years' experience, and 1 IM second-year resident) scored each patient's history from 0 to 2 in accordance with the history component of the HEART score. We oriented all physicians to the database and the data dictionary defining the variables. We did not provide any specific instructions to these physicians as to the application of the HEART score to better assess for variability in usual practice patterns among the physicians. We blinded the scorers to the results of the other scorers.

2.4 | Analysis

We reported the demographic and outcome data of the cohort using descriptive statistics. We calculated a correlation coefficient (Cohen's kappa) to assess variability between pairs of physician scorers. The primary pairwise comparison of interest was between the attending physician in EM and the attending physician in IM. We then made pairwise comparisons between the attending and resident physicians of the respective specialties and between the EM and IM residents. The 6 pairs we assessed were EM attending versus IM attending, EM attending versus EM resident, EM attending versus IM resident, IM attending versus EM resident, IM attending versus IM resident, and EM resident versus IM resident. Next, we calculated an intraclass correlation coefficient (ICC) for all 4 physician categories.[3,4,10] In addition to Cohen's kappa, we also calculated a percent agreement and 95% confidence interval for each pair listed previously. We calculated the median and interquartile range for each physician's scores. We used a Shapiro-Wilk test to assess for the normality of the data and determined the data had a non-normal distribution. For this reason, we used a Kruskal-Wallis test to assess for differences in the mean rank between the scores assigned by the physicians. The Kruskal-Wallis test serves as a single test comparing multiple groups, analogous to an analysis of variannce, but for non-normally distributed data.[9,22] We subsequently used post hoc pairwise testing using Mann-Whitney *U* test.[9,22] Finally, we isolated the cases in which there was disagreement in the history scores assigned and conducted Mann-Whitney *U* tests to assess for differences in mean rank. For the Mann-Whitney *U* tests, we applied a Bonferroni correction to decrease the threshold *P* value required for significance to 0.0083 to account for 6 comparisons.

We performed all statistical analysis using Microsoft Excel (Version 15, Redmond, WA) and IBM SPSS Statistics (Version 22).

TABLE 1 Patient demographics of selected cases

	Median/percent	Interquartile range/95% CI
Age	55	41.8–67.3
Sex (male, N = 31))	51.7	38.4–64.8
Hyperlipidemia (N = 12)	20	10.8–32.3
Hypertension (N = 28)	46.7	33.7–60.0
Diabetes (N = 12)	20	10.8–32.3
Family history of early CAD (N = 17)	28.3	17.5–41.4
Smoking (N = 10)	16.7	8.3–28.5
Obesity (N = 29)	48.3	35.2–61.6
CAD (N = 10)	16.7	8.3–28.5
MACE (N = 2)	3.3	0.4–11.5

Abbreviations: CAD, coronary artery disease; CI, confidence interval; MACE, major adverse cardiac event.

3 | RESULTS

3.1 | Characteristics of study subjects

This study included 60 subjects. The median age was 55 years and 51.7% were male (Table 1). Regarding comorbidities, 20.0% had a history of hyperlipidemia, 46.7% had a history of hypertension, 20.0% had a history of diabetes, 28.3% had a family history of early coronary artery disease, 16.6% had smoking history, 48.3% were obese (body mass index > 35), and 16.7% had known coronary artery disease.[6] The overall rate of MACE within 6 weeks in this population was 3.3%.

3.2 | Main results

The scores assigned by each of the raters were similar with each rater having a median assigned score of 1.00. However, the distribution of scores assigned varied among the raters (Table 2).

TABLE 2 History component scores assigned [N%]

	0	1	2
EM attending	16 (26.7)	23 (38.3)	21 (35.0)
IM attending	27 (45.0)	22 (36.7)	11 (18.3)
EM resident	16 (26.7)	22 (36.7)	22 (36.7)
IM resident	18 (30.0)	29 (48.3)	13 (21.7)

Abbreviations: EM, emergency medicine; IM, internal medicine.

The Cohen's kappa value between the scores assigned by the attending physician in EM and the attending physician in IM was 0.33 demonstrating only fair agreement (Table 3).[9] Of the additional pairwise comparisons, only the attending physician in EM and the resident physician in EM showed substantial agreement with a kappa value of 0.70. The attending physician in IM and the resident physician in IM showed moderate agreement with a kappa value of 0.46. The remainder of the pairs all showed fair agreement. Percent agreement in the pairwise comparisons showed a similar trend (Table 3). There was the greatest percent agreement between the EM attending physician and the EM resident physician at 80% with the other pairwise comparisons showing a lesser degree of agreement ranging from 48.3% to 65%. The overall ICC for the history component of the HEART score was 0.632 demonstrating fair agreement.[23]

The Kruskal-Wallis test showed a significant difference among the scores given by the respective physicians ($P = 0.033$). However, post hoc, pairwise comparisons with Mann-Whitney *U* tests between the individual physicians failed to show a statistically significant difference when the Bonferroni correction was applied for multiple comparisons.

In the cases in which there was disagreement in the assigned scores of the history component, there was a significant difference in the mean rank assigned in the both the comparison between the attending physician in EM and the attending physician in IM and the attending physician in IM and the resident physician in EM. In both cases the attending physician in IM assigned significantly lower scores than the attending physician in EM or the resident physician in EM ($P < 0.001$ for each). The other pairwise comparisons in this cohort of cases failed to demonstrate a significant difference in the mean rank assigned.

4 | LIMITATIONS

This was a retrospective analysis of a previously collected data set, and, therefore, scorers were limited to data abstracted from documentation by the physicians in EM during the initial history and physical exam. As such, the raters were unable to obtain the details of the history from the patients, preventing evaluation of the impact of the clinician's approach to history taking. Thus, although we were able to compare different interpretations of the same history data, there may be a bias toward increased agreement as physicians in EM documented all the encounters. An additional limitation of this study is that only a single physician represented each type of physician and level of training.

Another limitation is that the original study was a chart review study. Such studies suffer from multiple limitations, such as recall bias, inconsistent data entry, and the fact that necessary data elements may not be consistently collected.[20]

A final study limitation is that all the raters were from a single institution. This may limit generalizability to facilities in different geographic locations or with different patient populations. Although the institution is a tertiary referral center, it primarily serves active duty military, dependents, and retirees.[6,17,18]

5 | DISCUSSION

We report the first analysis of interrater reliability in the HEART score history component among physicians in EM and IM of different training levels. We found fair agreement between the attending physicians in EM and IM and fair to substantial agreement between the remaining pairs. Our statistical analysis demonstrated a significant difference in scores assigned by the raters but was unable to identify the pair responsible for this difference. However, in the subgroup in which there was disagreement between raters, the attending physician in EM and resident physician in EM were both found to assign significantly higher scores than the attending physician in IM.

Clinical pathways for patient disposition from the ED using the HEART score have been shown to decrease both objective cardiac testing and ED length of stay without an increase in MACE.[5] The history component of the HEART score is more subjective than other aspects

TABLE 3 Measures of agreement between physician pairs

	Kappa value (95% CI)	*P*-value	Percent agreement (95% CI)
EM attending - IM attending	0.33 (0.15–0.52)	<0.001	55.0 (41.6–67.9)
EM attending - EM resident	0.70 (0.54–0.85)	<0.001	80.0 (67.7–89.2)
EM attending - IM resident	0.22 (0.02–0.41)	0.016	48.3 (35.2–61.6)
IM attending - EM resident	0.37 (0.09–0.55)	<0.001	58.3 (44.9–70.9)
IM attending - IM resident	0.46 (0.27–0.65)	<0.001	65.0 (51.6–76.9)
EM resident - IM resident	0.30 (0.10–0.49)	0.001	53.3 (40.0–66.3)

Abbreviations: CI, confidence interval; EM, emergency medicine; IM, internal medicine.

of the scoring system. In facilities where IM physicians primarily admit patients with chest pain, the level of agreement between physicians in EM and IM regarding the history component of the HEART score can affect whether EM and IM agree on a patient's risk for MACE and the patient's disposition. This study demonstrates a discordance between physicians in EM and IM in the scoring of the HEART score history component. We further found, in the patients in which there was disagreement in the history score between physicians in EM and attending physicians in IM, that the physicians in EM tended to provide higher history score. Clinically, this seems to reflect a greater predisposition to keep these patients in the hospital by the physicians in EM.

In the context of existing knowledge, the prior studies have shown a slight to substantial interrater reliability of the HEART score history component with kappa scores ranging 0.10–0.66.[4,6–8,10,11] Our study's finding of fair agreement between attending physicians in EM and attending physicians in IM with a kappa 0.33 is consistent with these prior data. Interestingly, our study showed stronger agreement between attending physicians in EM and IM than that shown by Wu et al. between physicians in EM and cardiologists (kappa = 0.13).[10] Further, our study showed stronger agreement between EM attending and EM resident than shown in previous studies.[4] Although it has not been previously studied, we were not surprised to find a low level of agreement between physicians in EM and IM as this is consistent with the investigators clinical experience. However, we noted that the level of agreement between the IM attending and IM resident was lower than that between the EM attending and EM resident. This has not been previously studied so we are unsure if this is typical. This would be an interesting area of future investigation.

There is only a fair level of agreement regarding the history component of the HEART score between physicians in EM and IM . Further, in cases of disagreement, physicians in EM were more likely to assign a higher score to the history component.

CONFLICTS OF INTEREST

None.

AUTHOR CONTRIBUTIONS

Alec J. Pawlukiewicz, MD: Primary Investigator, data collection and analysis, manuscript writing. Matthew R. Geringer, DO: Associate Investigator, data collection and analysis, manuscript writing. W. Tyler Davis, MD: Associate Investigator, data collection and analysis, manuscript writing. Daniel R. Nassery, DO: Associate Investigator, data collection and analysis, manuscript writing. Michael D. April, MD, DPhil, MSc: Associate Investigator, data collection and analysis, manuscript writing. Matthew J. Streitz, MD: Associate Investigator, data collection and analysis, manuscript writing. Jessica M. Hyams, MD: Associate Investigator, data collection and analysis, manuscript writing. Alex W. Martin, LPN: Associate Investigator, manuscript writing. Sadie A. Martin, R.T.(R): Associate Investigator, manuscript writing. Joshua J. Oliver, MD: Associate Investigator, data collection and analysis, manuscript writing, research mentor.

REFERENCES

1. 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. https://doi.org/10.1016/j.ijcard.2013.01.255
2. Green SM, Schriger DL. A Methodological Appraisal of the HEART Score and Its Variants. *Ann Emerg Med*. 2021;78(2):253–266. https://doi.org/10.1016/j.annemergmed.2021.02.007.
3. Niven WGP, Wilson D, Goodacre S, et al. Do all HEART Scores beat the same: evaluating the interoperator reliability of the HEART score. *Emerg Med J*. 2018;35(12):732–738. https://doi.org/10.1136/emermed-2018-207540
4. Gershon CA, Yagapen AN, Lin A, et al. Inter-rater reliability of the HEART score. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2019;26(5):552–555. https://doi.org/10.1111/acem.13665
5. Mahler SA, Riley RF, Hiestand BC, et al. The HEART pathway randomized trial: identifying emergency department patients with acute chest pain for early discharge. *Circ Cardiovasc Qual Outcomes*. 2015;8(2):195–203. https://doi.org/10.1161/CIRCOUTCOMES.114.001384
6. Oliver JJ, Streitz MJ, Hyams JM, et al. An external validation of the HEART pathway among emergency department patients with chest pain. *Intern Emerg Med*. 2018;13(8):1249–1255. https://doi.org/10.1007/s11739-018-1809-y
7. Parenti N, Lippi G, Bacchi Reggiani ML, 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
8. van Meerten KF, Haan RMA, Dekker IMC, et al. The interobserver agreement of the HEART-score, a multicentre prospective study. *Eur J Emerg Med*. 2021;28(2):111–118. https://doi.org/10.1097/MEJ.0000000000000758.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
10. Wu WK, Yiadom MYAB, Collins SP, et al. Documentation of HEART score discordance between emergency physician and cardiologist evaluations of ED patients with chest pain. *Am J Emerg Med*. 2017;35(1):132–135. https://doi.org/10.1016/j.ajem.2016.09.058
11. Soares WE 3rd, Knee A, Gemme SR, et al. A prospective evaluation of clinical HEART score agreement, accuracy, and adherence in emergency department chest pain patients. *Ann Emerg Med*. 2021;78(2):231–241. https://doi.org/10.1016/j.annemergmed.2021.03.024.
12. Long B, Oliver J, Streitz M, Koyfman A. An end-user's guide to the HEART score and pathway. *Am J Emerg Med*. 2017;35(9):1350–1355. https://doi.org/10.1016/j.ajem.2017.03.047
13. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495–1499. https://doi.org/10.1016/j.ijsu.2014.07.013
14. Oliver JJ, Streitz MJ, Hyams JM, et al. The HEART score as a prognostic tool for revascularization. *Intern Emerg Med*. 2020;15(4):607–612. https://doi.org/10.1007/s11739-019-02206-0
15. Janes JL, Streitz MJ, Hyams JM, et al. Are patients discharged on the HEART pathway following up? *Mil Med*. 2020;30(185):e2110–e2114. https://doi.org/10.1093/milmed/usaa228
16. Oliver JJ, Streitz MJ, Hyams JM, et al. A HEART pathway pitfall in an admitted patient. *Am J Emerg Med*. 2019;37(1):177.e5–177.e6. https://doi.org/10.1016/j.ajem.2018.10.017
17. Streitz MJ, Oliver JJ, Hyams JM, et al. A retrospective external validation study of the HEART score among patients presenting to the emergency department with chest pain. *Intern Emerg Med*. 2018;13(5):727–748. https://doi.org/10.1007/s11739-017-1743-4
18. 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(4):549–557. https://doi.org/10.1016/j.jemermed.2017.12.038
19. Gilbert EH, Lowenstein SR, Koziol-McLain J, et al. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med*. 1996;27(3):305–308. https://doi.org/10.1016/s0196-0644(96)70264-0
20. Kaji AH, Schriger D, Green S. Looking through the retrospectoscope: reducing bias in emergency medicine chart review studies. *Ann Emerg Med*. 2014;64(3):292–298. https://doi.org/10.1016/j.annemergmed.2014.03.025
21. Bujang MA, Baharum N. Guidelines of the minimum sample size requirements for Cohen's Kappa. *Epidemiol Biostat Public Heal*. 2017;14(2):e12267–1-e12267-10. https://doi.org/10.2427/12267
22. Scheirer CJ, Ray WS, Hare N. The analysis of ranked data derived from completely randomized factorial designs. *Biometrics*. 1976;32(2):429–434.
23. Cicchetti D V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284–290. https://doi.org/10.1037/1040-3590.6.4.284

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Pawlukiewicz AJ, Geringer MR, Davis WT, et al. Interrater agreement of the HEART score history component: A chart review study. *JACEP Open*. 2022;3:e12732. https://doi.org/10.1002/emp2.12732