

Interrater Reliability and Accuracy of Clinicians and Trained Research Assistants Performing Prospective Data Collection in Emergency Department Patients With Potential Acute Coronary Syndrome

Carlos O. Cruz, BS
Emily B. Meshberg, BS
Frances S. Shofer, PhD
Christine M. McCusker, RN
Anna Marie Chang, MD
Judd E. Hollander, MD

From the Department of Emergency Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA.

Study objective: Clinical research requires high-quality data collection. Data collected at the emergency department evaluation is generally considered more precise than data collected through chart abstraction but is cumbersome and time consuming. We test whether trained research assistants without a medical background can obtain clinical research data as accurately as physicians. We hypothesize that they would be at least as accurate because they would not be distracted by clinical requirements.

Methods: We conducted a prospective comparative study of 33 trained research assistants and 39 physicians (35 residents) to assess interrater reliability with respect to guideline-recommended clinical research data. Immediately after the research assistant and clinician evaluation, the data were compared by a tiebreaker third person who forced the patient to choose one of the 2 answers as the correct one when responses were discordant. Crude percentage agreement and interrater reliability were assessed (κ statistic).

Results: One hundred forty-three patients were recruited (mean age 50.7 years; 47% female patients). Overall, the median agreement was 81% (interquartile range [IQR] 73% to 92%) and interrater reliability was fair (κ value 0.36 [IQR 0.26 to 0.52]) but varied across categories of data: cardiac risk factors (median 86% [IQR 81% to 93%]; median 0.69 [IQR 0.62 to 0.83]), other cardiac history (median 93% [IQR 79% to 95%]; median 0.56 [IQR 0.29 to 0.77]), pain location (median 92% [IQR 86% to 94%]; median 0.37 [IQR 0.25 to 0.29]), radiation (median 86% [IQR 85% to 87%]; median 0.37 [IQR 0.26 to 0.42]), quality (median 85% [IQR 75% to 94%]; median 0.29 [IQR 0.23 to 0.40]), and associated symptoms (median 74% [IQR 65% to 78%]; median 0.28 [IQR 0.20 to 0.40]). When discordant information was obtained, the research assistant was more often correct (median 64% [IQR 53% to 72%]).

Conclusion: The relatively fair interrater reliability observed in our study is consistent with previous studies evaluating interrater reliability for cardiovascular disease in the inpatient setting. With respect to research data, we found that prospective ascertainment of clinical data is more often correct when done by research assistants compared with clinicians simultaneously evaluating patients. [Ann Emerg Med. 2009;54:1-7.]

Provide feedback on this article at the journal's Web site, www.annemergmed.com.

0196-0644/\$-see front matter

Copyright © 2008 by the American College of Emergency Physicians.

doi:10.1016/j.annemergmed.2008.11.023

INTRODUCTION

Background

Although prospective data collection is generally considered more precise than data collected through chart abstraction,¹ collecting such data is complicated by substantial time and funding requirements.²⁻⁷ Busy emergency physicians often

perceive that they do not have the time to enroll patients, obtain consent, and complete data collection instruments. Moreover, difficulties in securing funds for research have contributed to a decline in the total number of clinical investigators.²⁻⁷ Although research can be facilitated by looking to industry for support, doing so may create conflicts of interest.⁸⁻¹¹ A different solution

Editor's Capsule Summary

What is already known on this topic

Valid clinical research requires high-quality data collection. Physicians are commonly considered the standard by which valid prospective data are obtained.

What question this study addressed

This study determined whether non-medically trained research assistants could reliably collect subjective historical data from emergency department patients with chest pain.

What this study adds to our knowledge

This prospective comparative study included 33 research assistants, 39 physicians, and 143 patients. Research assistants demonstrated fair to excellent reliability (as defined by crude agreement and κ) when obtaining cardiac histories and cardiac risk factors.

How this might change clinical practice

The results of this study will not change clinical practice. They do, however, provide evidence to support the use of trained research assistants for the collection of certain types of clinical data.

has been to expand available sources of labor by employing trained research assistants without a medical background to enroll patients, administer surveys, and obtain patient demographic information.¹²⁻¹⁴

Importance

The ability of research assistants to reliably collect prospective clinical data is unknown, regardless of presence or absence of a medical background. In addition to writing standard patient notes, we currently ask clinicians in our emergency department (ED) to document the guideline recommended,¹⁵ presenting symptoms and relevant medical history on standardized closed-question research forms. If this responsibility could be given to the research assistants, the clinicians caring for the patients would benefit.

Goals of This Investigation

Our goal was to test whether trained research assistants could obtain clinical research data as accurately as clinicians from patients presenting to the ED with chest pain. We hypothesized that they would be at least as accurate because they would not be distracted by clinical requirements.

MATERIALS AND METHODS

Study Design

This is a prospective comparative study of trained research assistants and clinicians who interviewed a convenience sample of patients presenting to the ED with chest pain. All data collectors were blinded to one another's data collection.

Setting

The study was based in the ED of an urban tertiary care center. The annual census is approximately 57,000 adult patients, of whom approximately 2,000 present with a chief complaint of chest pain.

Selection of Participants

We included patients who presented to the ED between September 3, 2007, and December 5, 2007, with chest pain symptoms and who received an ECG. Patients younger than 30 years who denied cocaine use were excluded. Broad inclusion criteria were intentionally chosen to ensure the generalizability of the data. The study was approved by the institutional review board of our hospital; verbal informed consent was obtained from all participants.

Research assistants were present in the ED 17 hours per day, 7 days per week. During these times, they consecutively identified and enrolled patients. The clinicians interviewing patients included emergency medicine residents, non-emergency medicine residents rotating in the ED, and emergency medicine attending physicians.

Training for the research assistants consisted of an initial 4-hour didactic session that included discussion of how the ED functions, methods of patient identification, informed consent, and patient enrollment. This was supplemented with Web-based training modules in patient-oriented research and patient confidentiality. All research assistants were supervised for a minimum of 8 hours while in the ED. During this time, they were directly observed to approach patients, administer informed consent, and conduct study surveys. Research assistants all attended 2 other sessions (1 per month), in which they could ask questions and discuss problems in a large group. The investigative team was available either in person or for 17 hours per day seven days per week during the study period.

Data Collection and Processing

One research assistant and 1 clinician separately interviewed each patient and recorded the history on a clinical research form. The history was recorded as answers to the form's closed-ended questions. In keeping with published chest pain research guidelines,¹⁵ the form's questions related to information considered to be important in the diagnosis of acute coronary syndromes: cardiac risk factors, other cardiac history, pain location, radiation, quality, and associated symptoms. All questions had to be answered either yes or no except for 2 questions that asked about the results of previous cardiac testing. If a patient had received a previous exercise stress test or

cardiac catheterization, the results had to be documented as normal, abnormal, indeterminate, or unknown. Both interviewers were free to use their own questioning styles during the patient interviews. To avoid delays in patient care, we did not control the order of the 2 interviews, but we did ask that one interview happen immediately after the other.

As soon as both interviews were completed, a second research assistant referred to as the tiebreaker examined the 2 forms for differences in the recorded answers. In cases of discrepancies, the tiebreaker reviewer would promptly ask the patient to verify his or her response, and the patient had to choose one of the 2 previously documented answers.

Outcome Measures

We examined the percentage of raw agreement and interrater reliability between the clinicians' and research assistants' recorded clinical histories. We also assessed the accuracy of each group's data in cases in which they differed from those of the other group. When both initial reviewers ascertained the same answer, it was considered correct. When they were discordant, the tiebreaker reviewer, who had directly asked the patient which of the previous 2 answers was correct, was considered the criterion standard of the data's accuracy. We did not use medical records to determine accuracy because we attempted to compare research assistants to clinicians with respect to obtaining information from the patient.

Primary Data Analysis

Crude agreement and interrater reliability between research assistants and clinicians were analyzed for each variable. For crude agreement, we present the percentage of agreement. We did not define an a priori measure of acceptable agreement because it varies according to the importance of the information. For interrater reliability, we used Cohen's κ statistic, a measure of the extent to which agreement is greater than expected by chance alone.¹⁶ κ Values range between -1.00 (no agreement) and $+1.00$ (complete agreement). A value of 0 indicates chance agreement. As described by Landis and Koch,¹⁷ κ less than 0.2 represents poor agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, good agreement; and 0.81 to 1.00, excellent agreement. We did not define an a priori measure of acceptable interrater reliability because it also may vary according to the importance of the information.

For purposes of data presentation, questions were grouped into 8 main categories, and percentage of agreement and κ statistics for each main category are summarized by using median and interquartile ranges (IQRs). Data were analyzed with SAS statistical software (version 9.1; SAS Institute, Inc., Cary, NC) and StatXact (version 6.1; Cytel Software Corporation, Cambridge, MA).

RESULTS

There were 143 patients recruited. Forty-seven percent were female patients, and they had a median age of 51 years (SD 13).

The interviews were performed by 33 research assistants (58% women) and 39 clinicians (59% women). The clinicians consisted of 19 emergency medicine residents, 16 non-emergency medicine residents, and 4 emergency medicine attending physicians.

Overall, across all individual items assessed (Table 1), the median agreement and interrater reliability were both fair (81% crude agreement [IQR 73% to 92%]; κ value 0.36 [IQR 0.26 to 0.52]). With regard to specific categories (Table 2), agreement was good for traditional cardiac risk factors (hypertension, diabetes, hypercholesterolemia, smoking, family history of myocardial infarction, and cocaine use) and was moderate for other cardiac history. There was good to excellent agreement, however, for certain components of the cardiac history (congestive heart failure, coronary artery disease, myocardial infarction, and coronary artery bypass grafting). There was moderate agreement for the presence and results of previous cardiac testing. The agreement was fair for pain location, radiation, quality, and associated symptoms. The general chest pain questions not falling into any of the other defined categories also had fair agreement. The discordance could not be accounted for by individual patients because the majority of patients had at least 1 variable that was discordant.

When there were discordances in data collection, the research assistant was more often correct than the clinician (64% [IQR 53% to 72%]) (Table 3). Furthermore, in half the categories, the research assistants obtained the accurate information more than 70% of the time.

LIMITATIONS

This study can be criticized for using a third "tiebreaker" who was not an experienced clinician. We chose the study design for several reasons. First, we were comparing the concordance of the treating clinician to the research assistant. We could have had the attending physician history serve as the tiebreaker, but we believed that a more directed approach, whereby the tiebreaker could directly present the patient with the 2 previous choices and ask him or her to choose between them, was more appropriate. We were concerned that the attending physician would have already evaluated the patient, possibly with the clinician in the room, and the patient would therefore feel compelled to answer in a manner similar to whatever they told the medical team. Direct confrontation by a third person who had not previously evaluated the patient was chosen to avoid that source of bias.

The use of a convenience sample may have created selection bias, but we attempted to limit this by screening all patients with chest pain when research assistants were available. We did not control for the clinicians' or research assistants' level of training. Previous studies suggest that there is a higher concordance among observers with similar training.^{18,19} Our research assistants were largely premedical post-baccalaureate students, who may have more medical interest or expertise than other research assistants. We did not control for whether the clinician or research assistant talked with the patient first.

Table 1. Reliability of clinical histories between clinicians and research assistants.

Characteristics	κ	95% CI	Crude Agreement, %	95% CI
General chest pain questions				
Chest pain chief complaint?	0.37	0.19 to 0.55	78	70 to 84
Chest pain at rest?	0.21	0.04 to 0.38	72	64 to 79
Multiple chest pain episodes?	0.33	0.17 to 0.49	67	59 to 75
Chest pain on ED arrival?	0.28	0.11 to 0.44	76	68 to 83
Worse than previous angina?	0.31	0.15 to 0.47	71	63 to 79
Chest pain pleuritic?	0.29	0.14 to 0.44	65	57 to 73
Chest pain quality				
None	0.39	-0.15 to 0.93	98	94 to 100
Crushing	0.07	-0.11 to 0.25	83	76 to 89
Pressure/tightness	0.30	0.15 to 0.46	66	58 to 74
Aching/dull	0.42	0.25 to 0.59	78	70 to 84
Burning	0.14	-0.07 to 0.35	87	81 to 92
Tearing	0.27	-0.18 to 0.71	97	92 to 99
Sharp/stabbing	0.26	0.10 to 0.43	68	60 to 75
Other/unknown	0.44	0.14 to 0.74	94	88 to 97
Chest pain location				
None	*	*	97	92 to 99
Substernal/midchest	0.47	0.32 to 0.61	73	65 to 80
Left anterior chest	0.52	0.38 to 0.66	76	68 to 83
Left arm	0.37	0.05 to 0.69	94	88 to 97
Right anterior chest	0.58	0.37 to 0.79	92	86 to 96
Epigastric	0.30	0.05 to 0.56	90	83 to 94
Other	0.21	-0.08 to 0.50	92	86 to 96
Unknown	-0.01	-0.02 to 0.00	99	95 to 100
Chest pain radiation				
None	0.35	0.21 to 0.49	67	59 to 75
Left arm	0.57	0.40 to 0.74	86	79 to 91
Jaw/neck	0.38	0.17 to 0.59	86	79 to 91
Back	0.23	-0.01 to 0.46	87	80 to 92
Right arm	0.44	0.09 to 0.78	95	90 to 98
Other	-0.05	-0.09 to (-0.01)	85	78 to 90
Associated symptoms				
Shortness of breath	0.44	0.30 to 0.59	73	65 to 80
Diaphoresis	0.31	0.14 to 0.48	77	69 to 84
Nausea	0.42	0.28 to 0.56	74	66 to 81
Vomiting	0.40	0.16 to 0.64	90	83 to 94
Light headed	0.25	0.14 to 0.35	57	49 to 66
Syncope	0.11	-0.03 to 0.24	81	74 to 87
Palpitations	0.23	0.09 to 0.37	68	59 to 75
Weakness	0.09	0.00 to 0.19	48	40 to 57
Cardiac risk factors				
Hypertension	0.75	0.63 to 0.86	88	82 to 93
Diabetes mellitus	0.86	0.76 to 0.95	94	89 to 98
Increased cholesterol level	0.63	0.49 to 0.76	84	77 to 90
Family history of myocardial infarction (<55 y)	0.35	0.20 to 0.50	73	64 to 80
Tobacco use (now or past)	0.61	0.49 to 0.74	80	73 to 87
Cocaine use (<1 week ago)	0.92	0.76 to 1.00	99	96 to 100
Other cardiac history				
Congestive heart failure	0.75	0.60 to 0.90	93	88 to 97
Coronary artery disease	0.77	0.64 to 0.91	93	88 to 97
Angina	0.35	0.17 to 0.54	79	71 to 85
Myocardial infarction	0.90	0.80 to 1.00	97	93 to 99
Undiagnosed chest pain	0.14	-0.02 to 0.30	73	65 to 80
GERD/peptic ulcer disease	0.26	0.10 to 0.43	73	65 to 80
Arrhythmia	0.29	0.11 to 0.46	80	73 to 87
Coronary artery bypass grafting	0.88	0.72 to 1.00	99	95 to 100
Known cancer	0.56	0.27 to 0.85	95	90 to 98
Cardiac testing				
Exercise stress test	0.52	0.39 to 0.65	75	66 to 81
Exercise stress test result	0.24	0.03 to 0.45	53	38 to 67
Cardiac catheterization	0.64	0.49 to 0.79	87	79 to 91

Table 1. Continued

Characteristics	κ	95% CI	Crude Agreement, %	95% CI
Cardiac catheterization result	0.46	0.16 to 0.76	67	45 to 84

CI, Confidence interval; GERD, gastroesophageal disease.

Unweighted κ scores for all items.

*A κ value was impossible to determine for "Chest pain location: None" because no clinicians answered this question in the affirmative.

Table 2. Reliability of clinical histories between clinicians and research assistants: Median κ values and crude agreement across question categories.

Category of Questions	κ	IQR	Crude Agreement, %	IQR, %
General questions	0.30	0.28-0.32	72	68-75
Chest pain quality	0.29	0.23-0.40	85	75-94
Chest pain location	0.37	0.25-0.49	92	86-94
Chest pain radiation	0.37	0.26-0.42	86	85-87
Associated symptoms	0.28	0.20-0.40	74	65-78
Cardiac risk factors	0.69	0.62-0.83	86	81-93
Other cardiac history	0.56	0.29-0.77	93	79-95
Cardiac testing	0.49	0.41-0.55	71	63-78

Unweighted κ scores for all items.

Table 3. Accuracy of research assistant–collected data in cases in which it differed from that of the clinicians.

Category of Questions	% Research Assistant Correct	IQR, %
General questions	53.0	46-59
Chest pain quality	71.0	68-82
Chest pain location	55.0	44-64
Chest pain radiation	66.0	57-68
Associated symptoms	71.0	64-76
Cardiac risk factors	77.0	60-85
Other cardiac history	60.0	50-67
Cardiac testing	76.0	66-87

Patients presenting with a disorganized list of complaints may sharpen their focus with subsequent interviews.²⁰ This "coaching" effect may have inflated the degree of concordance. It is also possible that the research assistants, who focused only on the study-related data, were likely to overcome this problem. It is also likely that the research assistants used the structured form while interviewing the patient and the resident did not complete the structured form until after the patient assessment; however, we did not directly assess this and cannot be sure whether this occurred.

It is difficult to quantify how much patient-related factors decreased the interrater reliability for each category of questions. As mentioned previously, patients often have poor or nonspecific recall that can vary with time. Moreover, the ability to interpret questions and provide coherent answers is generally more impaired in patients who are old, cognitively impaired, less medically knowledgeable, and not native English speakers.²⁰ These problems may be further compounded by a lack of honesty by patients as a result of mistrust or dissatisfaction with their care.²⁰ Although strong clinical and interpersonal skills should help the clinician mitigate these difficulties, it is

improbable that these difficulties are completely overcome routinely.

There are several well-described limitations to the use of the κ statistic that influence the interpretation of the results.^{21,22} When there is a very high prevalence of affirmative or negative answers, the level of expected agreement becomes so great that κ becomes difficult to interpret.²¹ For this reason, it is important to inspect the raw data for prevalence effects that artificially lower κ values. Statistical methods to adjust for prevalence have been proposed, and some experts believe that the intraclass correlation coefficient is a better measure, but nevertheless, κ remains the most widely used index of agreement.²³

Finally, our sample size was chosen on the basis of convenience rather than a priori statistical calculations. The IQR for categories of questions is up to 15 percentage points wide. Nonetheless, our percentage of accuracy favoring the trained research assistants supports our conclusion that they are at least as accurate as the clinicians, in general.

DISCUSSION

The relatively fair interrater reliability observed in our study is consistent with expected results based on previous studies evaluating interrater reliability for cardiovascular disease in the inpatient setting. To our knowledge, the only study that directly examined physician-physician observer variability in patients with chest pain was performed by Hickam et al.¹⁸ They assessed the interobserver agreement between 2 general internists in a study of 197 inpatients admitted with chest pain. The 2 internists had overall good agreement when asking similar questions to the ones found on our clinical data form. κ Values for pain radiating to the left arm, history of myocardial infarction, pain in substernal location, pain described as pressure, pain brought on by cough or deep breath, and pain described as sharp were 0.89, 0.78, 0.74, 0.57, 0.44, and 0.30,

respectively. Although their degree of agreement was higher than that found in our study, it was likely artificially inflated because the internists began the study by discussing how to interpret hypothetical responses and how to make it so that they approach the patients similarly. They also approached patients in a less hectic setting after the initial management decisions had already been made and the patients had already been questioned by the clinicians providing their medical care.

In contrast, our study demonstrated agreement for cardiac risk factors that was equal to or higher than that in a study by Hand et al.¹⁹ This study examined the interrater reliability among 3 physicians and a medical student who interviewed 98 patients presenting with potential stroke. Although a different patient population than our own, patients presenting with potential stroke also require inquiry about cardiac risk factors as an aid to diagnosis. The Hand study's κ values for a history of smoking, diabetes mellitus, ischemic heart disease, atrial fibrillation, and hypertension were 0.69, 0.65, 0.64, 0.54, and 0.47, respectively.

We observed several trends that paralleled those of previous chest pain studies. There was greater agreement for classic angina symptoms such as shortness of breath, nausea, vomiting, and diaphoresis than for atypical or nonspecific symptoms.¹ Indeed, nonspecific symptoms are often overlooked, possibly as a result of their being attributed to less severe illnesses.^{23,24} We also found that different observers were more likely to agree with regard to the location and radiation of chest pain than with regard to its quality,¹⁸ which makes intuitive sense because patients frequently have trouble describing chest pain whether they use their own words or those suggested by a physician.²⁰

In reference to our study's ultimate goal, we showed that the research assistants obtained the accurate information more often than the clinicians. We expected these results because, although less experienced, research assistants do not have to deal with the clinicians' competing clinical responsibilities. The research assistants do not have to conform to the same time requirements, and they can also focus entirely on asking each of the data collection form's questions explicitly.

In conclusion, we found that when clinical histories were recorded, the agreement between non-medically trained research assistants and clinicians is comparable to the agreement observed between physicians in previous chest pain studies. More important, we found that the research assistants were more likely than the clinicians to acquire the accurate data, as assessed by our tiebreaker. We thus conclude that it is appropriate for research assistants to prospectively record clinical histories in studies of patients presenting to the ED with chest pain. This utilization of research assistants should help expand limited labor resources and thus afford physicians the opportunity to manage additional clinical and research duties.

Supervising editors: Jason S. Haukoos, MD, MS; Steven M. Green, MD

Dr. Haukoos and Dr. Green were the supervising editors on this article. Dr. Hollander did not participate in the editorial review or decision to publish this article.

Author contributions: COC, EBM, CMM, AMC, and JEH contributed to data collection. COC, FSS, AMC, and JEH wrote the article. COC, EBM, FSS, CMM, and AMC critically reviewed the article. EBM, FSS, CMM, and JEH contributed to study design. EBM, FSS, CMM, AMC, and JEH analyzed and interpreted the data. FSS was overall statistician for the project. JEH was the principal investigator for the study, collated comments from other authors, and takes overall responsibility for the data. COC takes responsibility for the paper as a whole.

Funding and support: By *Annals* 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 that might create any potential conflict of interest. The authors have stated that no such relationships exist. See the Manuscript Submission Agreement in this issue for examples of specific conflicts covered by this statement.

Publication dates: Received for publication September 14, 2008. Revision received November 4, 2008. Accepted for publication November 26, 2008. Available online January 29, 2009.

Reprints not available from the authors.

Address for correspondence: Judd E. Hollander, MD, Department of Emergency Medicine, University of Pennsylvania, Ground Floor, Ravdin Building, 3400 Spruce Street, Philadelphia, PA 19104-4283; E-mail Judd.Hollander@uphs.upenn.edu.

REFERENCES

- DeVon HA, Ryan CJ, Zerwic JJ. Is the medical record an accurate reflection of patients' symptoms during acute myocardial infarction? *West J Nur Res*. 2004;26:547-560.
- Weissman JS, Saglam D, Campbell EG, et al. Market forces and unsponsored research in academic health centers. *JAMA*. 1999; 281:1093-1098.
- Williams GH, Wara DW, Carbone P. Funding for patient-oriented research. Clinical strain on a fundamental linchpin. *JAMA*. 1997; 278:227-231.
- Nathan DG, for the National Institutes of Health Directors Panel on Clinical Research. Clinical research: perceptions, reality, and propped solutions. *JAMA*. 1998;280:1427-1431.
- Schechter AN. The crisis in clinical research endangering the half century National Institutes of Health consensus. *JAMA*. 1998; 280:1440-1442.
- Wright SW, Wrenn K. Funding in emergency medicine literature: 1985 to 1992. *Ann Emerg Med*. 1994;23:1077-1081.
- Singer AJ, Homan CS, Stark MJ, et al. Comparison of types of research articles published in emergency medicine and non-emergency medicine journals. *Acad Emerg Med*. 1997;4:1153-1158.
- Panacek EA, Lewis RJ. Guidelines for clinical investigator involvement in industry-sponsored clinical trials. *Acad Emerg Med*. 1995;2:43-45.
- Drazen JM, Koski G. To protect those who serve [editorial]. *N Engl J Med*. 2000;343:1643-1644.

10. Lo B, Wolk LE, Berkeley A. Conflict of interest policies for investigators in clinical trials. *N Engl J Med.* 2000;343:1616-1620.
11. Van McCary S, Anderson CB, Jakovlevic J, et al. A national survey of policies on disclosure of conflicts of interest in biomedical research. *N Engl J Med.* 2000;343:1621-1626.
12. Hollander JE, Valentine SM, Brogan GX. Academic associate program: integrating clinical emergency medicine research with undergraduate education. *Acad Emerg Med.* 1997;4:225-230.
13. Hollander JE, Singer AJ. An innovative strategy for conducting clinical research: the academic associate program. *Acad Emerg Med.* 2002;9:134-137.
14. Bradley K, Osborn HH, Tang M. College research associates: a program to increase emergency medicine clinical research productivity. *Ann Emerg Med.* 1996;28:328-333.
15. Hollander JE, Blomkalns AL, Brogan GX, et al. Standardized reporting guidelines for studies evaluating risk stratification of emergency department patients with potential acute coronary syndromes. *Ann Emerg Med.* 2004;44:589-598.
16. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37-46.
17. Landis JR, Koch CG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
18. Hickam DH, Sox HC Jr, Sox CH. Systematic bias in recording the history in patients with chest pain. *J Chron Dis.* 1985;38:91-100.
19. Hand PJ, Haisma JA, Kwan JJ, et al. Interobserver agreement for the bedside clinical assessment of suspected stroke. *Stroke.* 2006;37:776-780.
20. Farmer SA, Roter DL, Higginson IJ. Chest pain: communication of symptoms and history in a London emergency department. *Patient Educ Couns.* 2006;63:138-144.
21. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ.* 1992;304:1491-1494.
22. Altman DG, Machin D, Bryant TN, et al. *Statistics With Confidence: Confidence Intervals and Statistical Guidelines.* London, England: BMJ Books; 2000.
23. Horne R, James D, Petrie K, et al. Patients' interpretation of symptoms as a cause of delay in reaching hospital during acute myocardial infarction. *Heart.* 2000;83:388-393.
24. Canto JG, Shlipak MG, Rogers WJ, et al. Prevalence, clinical characteristics, and mortality of patients with myocardial infarction presenting without chest pain. *JAMA.* 2000;283:3223-3229.