Journal of the American College of Cardiology
© 2001 by the American College of Cardiology
Published by Elsevier Science Inc.

Vol. 38, No. 7, 2001
ISSN 0735-1097/01/$20.00
PII S0735-1097(01)01639-4

## EDITORIAL COMMENT

# Risk Stratification for Arrhythmic Events: Are the Bangs Worth the Bucks?*

Mark J. Eisenberg, MD, MPH, FACC
*Montreal, Canada*

More than 800,000 North Americans are admitted to the hospital each year with an acute myocardial infarction (MI) (1,2). Most of these patients survive until hospital discharge, but a substantial number of the survivors will die of out-of-hospital arrhythmic events (3). Approximately half the deaths that occur following acute MI are attributable to recurrent ischemic events or to congestive heart failure related to left ventricular impairment (3,4). A plethora of clinical trials and guideline statements have addressed the issue of post-MI risk stratification with respect to these complications. However, lethal arrhythmias cause approximately half the deaths that occur during the first year after acute MI (5,6). Unfortunately, little information is available to guide us in the appropriate stratification of patients at risk for this complication.

**See page 1902**

Various tests have been proposed as being useful for post-MI stratification of patients at risk of major arrhythmic events. Investigators have found that measures of left ventricular function are particularly useful (7). Other tests that have been closely examined include signal-averaged electrocardiogram (SAECG) (8) and Holter monitoring. Several aspects of Holter monitoring have been correlated with the risk of arrhythmic events, including the incidence of premature ventricular contractions, the occurrence of major arrhythmias and a reduction in heart rate variability (HRV) (9). Electrophysiologic studies are thought to be especially predictive (10). More recently, various other techniques have been examined as possible predictors, including baroreflex sensitivity (11), QT dispersion (12), T-wave alternans (13), time-domain analysis (14), spectral turbulence analysis (14) and fractal dimension (15).

Because no single diagnostic test has been found to have adequate predictive ability, investigators have examined the ability of various combinations of tests to predict arrhythmic events in the post-MI population. Based on the results of these studies, a number of testing algorithms have been proposed that combine the use of left ventricular ejection fraction (LVEF), SAECG, HRV and electrophysiologic studies (16–21). Because of the cost and invasive nature of electrophysiologic studies, this test is usually reserved for the end of the algorithm when it can be used in small numbers of high-risk patients. Unfortunately, previous studies that examined combinations of tests were done in relatively small numbers of patients at a limited number of clinical centers. These studies had the advantage of performing several types of tests in the same patient population, but the studies were limited because of their size and generalizability. In this context, the report by Bailey et al. (22) in this issue of the *Journal* is a welcome addition to the rather limited data that are currently available.

The objective of the study by Bailey et al. (22) was to identify a staged approach to post-MI risk stratification with respect to arrhythmic events. To that end, the investigators performed a literature review to identify studies assessing the predictive abilities of five different tests for the identification of major arrhythmic events (SAECG, severe ventricular arrhythmias on Holter monitoring, HRV, LVEF and electrophysiologic study). Sensitivities and specificities for the five tests were pooled from 44 reports. Point estimates for the sensitivities and specificities were determined as well as 95% confidence intervals (CIs). The researchers then assessed various combinations of these tests to determine which staged application would afford the most efficient use of resources and be the most effective in identifying those patients who would benefit from the insertion of an implantable cardioverter-defibrillator (ICD). Finally, Bailey et al. (22) performed an analysis to determine the cost of a staged approach to risk stratification.

The pooled results of their study demonstrated that the sensitivities for the five diagnostic tests ranged from 42.8% to 62.4%, and the specificities ranged from 77.4% to 85.8%. From these results, the investigators identified a three-stage risk-stratification approach. This approach involves screening all post-MI patients with a SAECG as well as a measure of LVEF. Patients who are judged to be high risk by both studies receive an ICD; low-risk patients undergo no further risk stratification; and intermediate-risk patients undergo Holter monitoring both for examination of HRV and to detect the occurrence of significant ventricular arrhythmias. If the Holter is positive, the patient receives an ICD; if it is negative, the patient undergoes no further testing; and if it is indeterminate, the patient undergoes an electrophysiologic study. The investigators calculate that the risk-stratification approach that they advocate will lead to 11.8% of post-MI patients being identified as being likely candidates for an ICD (41.4% risk of major arrhythmic events at two years); 80.0% being identified as low risk and not requiring an ICD (2.9% risk at two years); and 8.2% of patients being identified as intermediate risk (8.9% risk at two years). The cost of this staged approach is estimated to be $415 per patient.

*Editorials published in the *Journal of the American College of Cardiology* reflect the views of the authors and do not necessarily represent the views of *JACC* or the American College of Cardiology.

From the Jewish General Hospital, McGill University, Montreal, Quebec, Canada.

**Table 1.** Characteristics of the Five Diagnostic Tests (Assuming a Prior Probability of 7.9%)

| | Sensitivity (%) [95% CI] | Specificity (%) [95% CI] | Positive Predictive Value (%) | Negative Predictive Value (%) | Accuracy (%) | Positive Likelihood Ratio | Negative Likelihood Ratio |
|---|---|---|---|---|---|---|---|
| SAECG | 62.4 [56.4–67.9] | 77.4 [73.6–80.8] | 19.1 | 96.0 | 76.2 | 2.8 | 0.5 |
| SVA | 42.8 [32.7–53.7] | 80.9 [75.0–85.7] | 16.1 | 94.3 | 77.9 | 2.2 | 0.7 |
| HRV | 49.8 [37.5–62.1] | 85.8 [82.1–88.9] | 23.1 | 95.2 | 83.0 | 3.5 | 0.6 |
| LVEF | 59.1 [53.1–64.6] | 77.8 [75.5–79.9] | 18.6 | 95.7 | 76.3 | 2.7 | 0.5 |
| EPS | 61.6 [48.2–73.4] | 84.1 [65.0–93.8] | 24.9 | 96.2 | 82.3 | 3.9 | 0.5 |

EPS = electrophysiologic study; HRV = heart rate variability; LVEF = left ventricular ejection fraction; SAECG = signal-averaged electrocardiogram; SVA = significant ventricular arrhythmia on Holter monitoring.

Although previous studies examined the utility of post-MI risk stratification, the study by Bailey et al. (22) highlights several important issues that deserve attention. First, their study involves a meta-analysis of previous studies that examined diagnostic tests thought to identify patients at high risk of arrhythmic events. In comparison with previous studies, Bailey et al. (22) performed pooled analyses to determine estimates of sensitivity and specificity for the diagnostic tests they examined. These pooled analyses involved large numbers of subjects, ranging from 4,022 patients for electrophysiologic studies to 9,883 patients for studies involving SAECG. As a consequence, the test characteristics that the investigators report may be more precise and more generalizable than the test characteristics reported by previous investigators.

To obtain reliable pooled estimates of sensitivity and specificity, meta-analyses of diagnostic test results must be performed in a rigorous manner (23,24). Similar to meta-analyses of clinical trial results, meta-analyses of diagnostic test results require meticulous attention to the identification of studies to be included, data abstraction and data pooling. When identifying potential studies for inclusion, for example, every effort should be made to identify all studies in the area—whether positive or negative. Thus, in addition to searching the English and non-English medical literature, authors of prior studies should be contacted to see if they are aware of unpublished studies examining the area of interest. Publication bias, in which negative studies are less likely to be published, can lead to an overestimation of the utility of a diagnostic test (25,26). Once thorough efforts have led to the identification of all potential studies, blinded reviewers should independently examine the candidate studies using systematic inclusion and exclusion criteria. Only when the reviewers agree does a candidate study come to be included in the meta-analysis. Once a study is included, the reviewers should independently abstract the data. For meta-analyses of diagnostic tests, particular attention must be paid to the definitions used by the investigators to define positive and negative test results. Finally, once the data have been abstracted, specific methods should be used to pool the characteristics of the diagnostic tests (27). Even after closely following the above steps, investigators have noted differences in the results of meta-analyses on the same subject, and meta-analyses that pooled large numbers of patients have been found to differ with subsequent large clinical trials (24,28).

Although the techniques used by Bailey et al. (22) largely conform to previous recommendations regarding the conduct of meta-analyses, several exceptions are present. For example, the investigators limited their literature search to English-language studies, and they do not appear to have contacted previous researchers about the existence of unpublished studies. The studies of diagnostic tests that they pooled varied in terms of the cut points used to determine a positive test, and importantly, these tests were administered to patients at varying times after their MIs. No blinding of the data abstracters appears to have been employed, and the methods that were used for the data pooling are somewhat unclear. In most meta-analyses of this type, a random effects model is used (27). If there is homogeneity among the studies, then the model will reduce to one that assumes homogeneity; if not, the between-study variation will automatically be incorporated into the model. Because several methodologic limitations are present in their meta-analysis, the investigators may have overestimated the sensitivities and specificities of the tests they examined.

A second issue is that even with the large numbers of patients included in their pooled analyses, because there were major variations among the individual studies, the 95% CIs for the pooled sensitivities and specificities were wide (Table 1). The staged approach to risk stratification advocated by Bailey et al. (22) is based on their point estimates for sensitivity and specificity. However, their conclusions may not be robust given the wide and overlapping ranges within which the true sensitivities and specificities for these tests lie. A sensitivity analysis might have been helpful in this regard. By varying the sensitivities and specificities used in their calculations between the extremes of the 95% CIs, the investigators would have been able to examine whether their conclusions were indeed robust.

A third issue to be considered is that the investigators reported for each diagnostic test the two-year probability of a major arrhythmic event if the test was positive (also known as the *positive predictive value*) or if the test was negative (equivalent to 100 − *negative predictive value*). However, because these test characteristics vary with the prior probability of disease, so as to ensure comparability between tests,

JACC Vol. 38, No. 7, 2001
December 2001:1912–5

these characteristics should be calculated using the same prior probability of disease. After recalculating these test characteristics using a prior probability of 7.9% for each of the tests (Table 1), we can see that the ability of these individual tests to distinguish high-risk patients is similar and remarkably poor. These results are confirmed when the accuracy and the positive and negative likelihood ratios are calculated. Accuracy is an overall measure of how well a diagnostic test performs. It is a composite measure of how frequently a diagnostic test correctly identifies patients with and without a disease or condition. The positive likelihood ratio is the odds that a patient with a disease will have a positive test divided by the odds that a patient without the disease will have a positive result. The negative likelihood ratio is the odds that a patient with a disease will have a negative test divided by the odds that a patient without the disease will have a negative result. Similar to sensitivity and specificity, likelihood ratios do not vary with different pretest probabilities of disease. When the above test characteristics are calculated, we again see that none of the diagnostic tests examined are exceptional predictors of which patients will go on to have arrhythmic events. In addition, although HRV and electrophysiologic studies may be modestly better tests, overall these tests are quite similar in their predictive abilities. It is for this reason that various investigators have attempted to devise a sequence of tests that will provide a better prediction than any one of the tests alone.

Thus, a fourth issue is the methodology by which the predictive value of combinations of tests is determined. Generally speaking, when we employ a combination or a sequence of diagnostic tests, we try to optimize the overall sensitivity and specificity of the combination. When screening for HIV, for example, we initially use a test with high sensitivity and low specificity so as not to miss any patients who might have the disease. For those patients who have a positive test, a second diagnostic test with high specificity is then required to confirm the diagnosis. In this way, we can efficiently identify patients with HIV even though the two tests individually may not have adequate discriminatory ability.

In the case of tests used to predict major arrhythmic events, however, we can see that each of the tests has low sensitivities and only modest specificities. Thus, it is hard to imagine how any combination of these tests will lead to a truly effective identification of patients at risk. Even with the staged approach advocated by the investigators (22), 11.8% of all post-MI patients will require an ICD, and 58.6% of these patients will not benefit from it. In addition, the optimal discriminative ability of a combination of tests occurs when there is no interdependence between the tests (29). For example, exercise treadmill testing and echocardiographic measurement of LVEF are likely to provide independent prognostic information after acute MI. The use, however, of HRV and identification of serious ventricular arrhythmias on the same Holter monitor may not provide completely independent prognostic information. Because Bailey et al. (22) assumed independence between the diagnostic tests used in their study, the prognostic ability of their staged approach may well be overestimated.

Finally, three practical issues deserve mention: 1) the relationship between ischemia and arrhythmic events; 2) the timing of the diagnostic tests; and 3) the cost of applying the risk-stratification approach advocated by the investigators. Perhaps the most important cause of major arrhythmic events in the post-MI population is myocardial ischemia. Most of the studies that were pooled by Bailey et al. (22) examined groups of patients in whom residual ischemia had already been excluded. Because ischemia is such a common cause of arrhythmic events, any staged approach to risk stratification must take this etiology into account. It is for this reason that most electrophysiologists are reluctant to employ the diagnostic tests examined by the investigators until the absence of residual ischemia has been confirmed.

The timing of diagnostic testing after MI is also important. A practical, population-based risk-stratification scheme would need to be performed in hospital while patients are recovering from an MI. However, diagnostic testing in the acute and subacute MI setting may not be predictive of out-of-hospital arrhythmic events. Testing patients after discharge, however, is impractical and likely to be associated with low rates of compliance.

Finally, cost is a major factor limiting the application of the investigators' risk-stratification approach. Several studies suggest that the cost associated with implantation of an ICD is >$30,000 (30,31). This figure only includes the cost of the device, the implantation procedure, and the in-hospital stay. Additional costs are incurred during the follow-up period with regular maintenance, battery replacements, and so forth. Using the approach advocated by Bailey et al. (22), 11.8% of post-MI patients would be identified as benefiting from a prophylactic ICD insertion. If more than 800,000 North Americans are admitted to hospital each year with acute MI, this approach would identify >80,000 patients who would require ICD insertion. At a cost of >$30,000/patient, this would amount to well over $2 billion annually. To this figure, one would have to add the overall costs for risk stratification (>$300 million/year) as well as the ongoing costs associated with ICDs. Thus, it can be seen that even a risk-stratification scheme that is reasonably effective at identifying high-risk patients will lead to high costs and to many patients receiving an ICD who do not require one.

In conclusion, the present study by Bailey et al. (22) is the first to conduct a meta-analysis of diagnostic tests thought to be predictive of post-MI arrhythmic events. Using pooled diagnostic test characteristics, the investigators have arrived at a risk-stratification algorithm that may be more reliable than those previously reported. Nevertheless, much work needs to be done in this important area. Even the staged approach advocated by the researchers is inadequate for population-based risk stratification. The approach has not

been tested prospectively in randomized clinical trials, would be immensely costly to implement, and would lead to large numbers of patients receiving ICDs who would never need them. Nevertheless, Bailey et al. (22) have made an important contribution. Their findings will doubtless serve as a springboard for future studies examining the issue of risk stratification for arrhythmic events following acute MI.

**Acknowledgments**

The author would like to thank Drs. Lawrence Joseph and Tom Hadjis for helpful comments as well as Mr. Jonathan Marsan for his assistance in the preparation of the manuscript.

**Reprint requests and correspondence:** Dr. Mark J. Eisenberg, Divisions of Cardiology and Clinical Epidemiology, Jewish General Hospital/McGill University, 3755 Cote Ste. Catherine Road/ Suite A-118, Montreal, Quebec H3T 1E2, Canada. E-mail: marke@epid.jgh.mcgill.ca.

## REFERENCES

1. Every NR, Hlatky MA, McDonald KM, Weaver WD, Hallstrom AP. Estimating the proportion of post-myocardial infarction patients who may benefit from prophylactic implantable defibrillator placement from analysis of the CAST registry. Am J Cardiol 1998;82:683–5.
2. Friedman PL, Stevenson WG. Unsustained ventricular tachycardia—to treat or not to treat? (editorial). N Engl J Med 1996;335: 1984–5.
3. American Heart Association. 2001 Heart and Stroke Statistical Update. Dallas, TX: American Heart Association.
4. Moss AJ, DeCamilla J, Davis H. Cardiac death in the first six months after myocardial infarction: potential for mortality reduction in the early posthospital period. Am J Cardiol 1977;39:816–20.
5. Greene HL, Richardson DW, Barker AH, et al, and the CAPS investigators. Classification of deaths after myocardial infarction as arrhythmic or nonarrhythmic (the Cardiac Arrhythmia Pilot Study). Am J Cardiol 1989;63:1–6.
6. The Multicenter Post-Infarction Research Group. Risk stratification and survival after myocardial infarction. N Engl J Med 1983;309: 331–6.
7. Bigger JT Jr, Fleiss JL, Kleiger R, et al., and the Multicenter Post-Infarction Research Group. The relationships among ventricular arrhythmias, left ventricular dysfunction, and mortality in the two years after myocardial infarction. Circulation 1984;69:250–8.
8. Steinberg JS, Regan A, Sciacca RR, et al. Predicting arrhythmic events after acute myocardial infarction using the signal-averaged electrocardiogram. Am J Cardiol 1992;69:13–21.
9. Verschueren LJ. Risk stratification for arrhythmic events in postinfarction patients based on heart rate variability. J Am Coll Cardiol 1992;19:1369.
10. Bourke JP, Richards DAB, Ross DL, Wallace EM, McGuire MA, Uther JB. Routine programmed electrical simulation in survivors of acute myocardial infarction for prediction of spontaneous ventricular tachyarrhythmias during follow-up: results, optimal stimulation protocol and cost-effective screening. J Am Coll Cardiol 1991;18:780–8.
11. La Rovere MT, Pinna GD, Hohnloser SH, et al. Baroreflex sensitivity and heart rate variability in the identification of patients at risk for life-threatening arrhythmias: implications for clinical trials. Circulation 2001;103:2072–7.
12. Barr CS, Naas A, Freeman M, Lang CC, Struthers AD. QT dispersion and sudden unexpected death in chronic heart failure. Lancet 1994;343:327–9.
13. Ikeda T, Sakata T, Takami M, et al. Combined assessment of T-wave alternans and late potentials used to predict arrhythmic events after myocardial infarction. A prospective study. J Am Coll Cardiol 2000; 35:722–30.
14. Vazquez R, Caref EB, Torres F, Reina M, Espina A, El-Sherif N. Improved diagnostic value of combined time and frequence domain analysis of the signal-averaged electrocardiogram after myocardial infarction. J Am Coll Cardiol 1999;33:385–94.
15. Anderson JL, Karagounis LA, Stein KM, Moreno FL, Ledingham R, Hallstrom A. Predictive value for future arrhythmic events of fractal dimension, a measure of time clustering of ventricular premature complexes, after myocardial infarction. J Am Coll Cardiol 1997;30: 226–32.
16. Farrell TG, Bashir Y, Cripps T, et al. Risk stratification for arrhythmic events in postinfarction patients based on heart rate variability, ambulatory electrocardiographic variables, and signal-averaged electrocardiogram. J Am Coll Cardiol 1991;18:687–97.
17. McClements BM, Adgey AA. Value of signal-averaged electrocardiography, radionuclide ventriculography, Holter monitoring and clinical variables for prediction of arrhythmic events in survivors of acute myocardial infarction in the thrombolytic era. J Am Coll Cardiol 1993;21:1419–27.
18. Hartikainen JEK, Malik M, Staunton A, Poloniecki J, Camm AJ. Distinction between arrhythmic and nonarrhythmic death after acute myocardial infarction based on heart rate variability, signal-averaged electrocardiogram, ventricular arrhythmias, and left ventricular ejection fraction. J Am Coll Cardiol 1996;28:296–304.
19. Pedretti R, Etro MD, Laporta A, Braga SS, Caru B. Prediction of late arrhythmic events after acute myocardial infarction from combined use of noninvasive prognostic variables and inducibility of sustained monomorphic ventricular tachycardia. Am J Cardiol 1993;71:1131–41.
20. El-Sherif N, Denes P, Katz R, et al. Definition of the best prediction criteria of the time domain signal-averaged electrocardiogram for serious arrhythmic events in the postinfarction period. J Am Coll Cardiol 1995;25:908–14.
21. Gomes JA, Winters SL, Stewart D, Horowitz S, Milner M, Barreca P. A new noninvasive index to predict sustained ventricular tachycardia and sudden death in the first year after myocardial infarction: based on signal averaged electrocardiogram, radionuclide ejection fraction and Holter monitoring. J Am Coll Cardiol 1987;10:349–57.
22. Bailey JJ, Berson AS, Handelsman H, Hodges M. Utility of current risk stratification tests for predicting major arrhythmic events after myocardial infarction. J Am Coll Cardiol 2001;38:1902–11.
23. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994;120:667–76.
24. Bailar JC 3rd. The promise and problems of meta-analysis (editorial). N Engl J Med 1997;337:559–61.
25. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000;356:1228–31.
26. Sterne JA, Gravaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000;53:1119–29.
27. Joseph L. A brief introduction to meta-analysis. In: Rosenberg L, Joseph L, Barkun A, editors. Surgical Arithmetic: Epidemiological, Statistical and Outcome-Based Approach to Surgical Practice. Georgetown, TX: Landes Bioscience, 2000:143–58.
28. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997;337:536–42.
29. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 2001; 57:158–67.
30. Olshansky B, Loots D, McGrory-Usset M. Implantable cardioverter-defibrillator practices and costs at an academic medical center. J Cardiovasc Electrophysiol 2001;12:162–6.
31. Venditti FJ Jr, O'Connell M, Martin DT, Shahian DM. Transvenous cardioverter-defibrillators: cost implications of a less invasive approach. Pacing Clin Electrophysiol 1995;18:711–5.