LETTERS TO THE EDITOR

Survival Analysis for Cardiac Risk Stratification

To the Editor:—In the January issue, Shaw and colleagues use the techniques of survival analysis to investigate the ability of exercise myocardial perfusion tomographic imaging to incrementally risk-stratify patients following risk assessment by history and exercise treadmill testing.¹ We are concerned that choices made in the analysis of their data bring their results and conclusions into question.

Shaw and colleagues used proportional hazards regression to assess the associations between clinical, exercise testing, and imaging variables and time to death. As noted in the editorial accompanying their article, they appropriately chose to use cardiac death, not coronary revascularization, as their outcome because the use of revascularization as an outcome would have introduced bias given that physicians use test results to choose therapy.² Their decision to censor patients at the time of coronary revascularization, however, violates one of the basic assumptions of survival analysis and most likely invalidates the analysis.

Survival analysis assumes that censoring is noninformative: that is, censoring occurs independent of the risk of outcome.³ If this assumption does not hold, a form of selection bias is introduced. In the study by Shaw and colleagues, the analysis assumes that had the censored subjects not been revascularized, they would have had the same risk of cardiac death as the other subjects. We believe that one cannot reasonably assume censoring in this case to be independent of the risk of cardiac death.

For patients with intermediate Duke treadmill scores, Shaw and colleagues present Kaplan-Meier survival curves by the number of infarcted or ischemic territories on nuclear scanning. The decision to censor those subjects who underwent revascularization may have resulted in an exaggerated separation between the survival curves for the strata defined by imaging results. For those patients found to have no infarcted or ischemic territories on imaging, censoring due to revascularization implies either a false-negative scan or the presence of very aggressive disease. In either case, these subjects would most likely have a higher cardiac death rate than that of other patients with negative scans who remained under observation. Censoring these subjects would make the prognosis for the group with negative scans look better than it is.

Conversely, consider subjects with three or more infarcted or ischemic territories on the scan. Under most circumstances, these subjects would be candidates for revascularization, and we

would expect them to be censored. Patients remaining under observation would be those who had a false-positive scan or were felt to be poor surgical candidates. If the majority from this high-risk group who remained under observation were, in fact, subjects with false-positive scans, the identified high-risk cohort would be artificially "healthier" and bias results toward the null. On the other hand, if the majority were poor surgical candidates, censoring those revascularized would leave a "sicker" group in the study and thus incorrectly magnify the differences in mortality based on scan result. We cannot know the net effect of the bias, but given that there is at least a reasonable possibility that it is causing a bias away from the null, we have to question the significance of the finding that the number of ischemic or infarcted territories predicts survival.

If extended follow-up data are available for those who underwent revascularization in this study, the authors might reanalyze their data without censoring at the time of revascularization. Instead, they could include revascularization as a time-dependent covariate in an extended Cox regression. In this way, they would preserve cardiac mortality as their main outcome while adjusting for the effects of revascularization within each group. We hope that this approach proves possible and look forward to learning of their revised results.—**CHRISTINA NICOLAIDIS, MD, MPH, and BRIAN JOHNSTON, MD, MPH, University of Washington, Seattle, Wash.**

REFERENCES

1. Shaw L, Hachamovitch R, Peterson E, et al. Using an outcomes-based approach to identify candidates for risk stratification after exercise treadmill testing. *J Gen Intern Med.* 1999;14:1-9.
2. Hlatky M. Exercise testing to predict outcome in patients with angina. *J Gen Intern Med.* 1999;14:63-5.
3. Katz M, Hauck W. Proportional hazards (Cox) regression. *J Gen Intern Med.* 1993;8:702-11.

In reply:—We thank Drs. Nicolaidis and Johnston for their insightful comments. Some background information might be helpful. Our group has been using Cox proportional hazards models to study the natural history of heart disease for over two decades, with too many citations to include in this letter. Our goal has been to advance the field of noninvasive diagnostic testing and develop adequate clinical and economic outcome methods that reflect the true nature of patient care. Drs. Nicolaidis and Johnston rightfully point out that it is difficult to study the natural history of heart disease because current therapies change that history. This problem makes the choice of analytic method important. We have explored a number of alternatives, including using revascularization as a time-dependent covariate and using a stratified Cox model.¹ In fact, we performed an analysis on the data we reported in our article using a stratified Cox model, and the results did not diverge from those we reported.

Although we agree that other models may reflect current medical practice more appropriately, it is unclear which models are better. Furthermore, we take issue with the assertion that censoring violates the proportional hazards assumption. Patients whose underlying risk is changed should be censored.² Censoring should be unrelated to the primary event, and in the case of

---

JGIM welcomes your letters and comments. Letters may contain brief commentaries on articles published in the Journal, illustrative case reports, general suggestions for improving the Journal, or other information of interest to readers. Letters to the Editor should not exceed 450 words in length and should be sent in triplicate (typed, double-spaced). We cannot guarantee publication and we may edit to improve readability.

You may send your letters to us by mail or fax: JGIM, JHU Division of General Internal Medicine, 1830 East Monument Street, Room 8068, Baltimore, MD 21205; fax (410) 955-0825.

coronary revascularization, many would argue that improved survival alters the model prediction for that patient.

Nikolaidis and Johnston also argue that censoring creates greater separation between strata. This is not the case when adequate samples are used. The biggest changes would be in small samples of fewer than 1,000 patients. Nikolaidis and Johnston further argue that cardiac survival appears better for patients with normal test results. Perhaps this would be the case if more than a handful of patients with normal test results proceeded to revascularization. During the follow-up in our study, only five patients had percutaneous transluminal coronary angioplasty. In our experience, 5% to 7% of patients with normal test results proceed to cardiac catheterization, with few undergoing coronary interventional procedures. Furthermore, fewer than half of patients with markedly positive scans undergo cardiac catheterization.^{3,4} This low rate of intervention limits the effect described by Nikolaidis and Johnston.

A key problem identified by Nikolaidis and Johnston is that we need to examine differences in survival for those patients who proceed to revascularization as compared to those who do not. We are enlarging both the length of follow-up and the number of observations to make more meaningful statements about these important issues.—**LESLEE J. SHAW, PhD**, *Emory Center for Outcomes Research, Atlanta, Ga*, and **RORY HACHAMOVITCH, MD**, *Cornell Medical Center, New York, NY*.

REFERENCES

1. Mark DB, Nelson CL, Califf RM, et al. Continuing evolution of therapy for coronary artery: initial results from the era of coronary angioplasty. *Circulation*. 1994;89(5):2015-25.
2. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-87.
3. Hachamovitch R, Berman DS, Shaw LJ, et al. Incremental prognostic value of myocardial perfusion single photon emission computed tomography for the prediction of cardiac death: differential stratification for risk of cardiac death and myocardial infarction. *Circulation*. 1998;97(6):535-43.
4. Travin MI, Duca MD, Kline GM, Herman SD, Demus DD, Heller GV. Relation of gender to physician use of test results and to prognostic value of stress technetium 99m sestamibi myocardial single-photon emission computed tomography scintigraphy. *Am Heart J*. 1997;134(1):73-82.

Response from the statistical editor:—An earlier issue of *JGIM* included an article that described how to use exercise electrocardiographic testing and exercise myocardial perfusion tomographic imaging to predict the risk of dying from ischemic heart disease.¹ We received two letters about the article, one written by readers concerned about a methodologic issue and the other a response from some of the article's authors. I will add to their comments by briefly explaining the methodologic issue and then synthesizing the contrasting comments.

In survival analysis, it is sometimes necessary to remove a subject who is still alive from the analysis, which is known as "censoring" the subject. All of the standard methods for survival analysis assume that the time a subject is censored is unrelated to the survival time that would have been observed if the subject had not been censored. This assumption may not be true. For example, consider a study comparing the survival rates of patients receiving two different treatments. During the study, several subjects receiving one treatment withdraw because they are near

death. They enter a hospice that is far away, cannot be observed, and thus are censored. Their withdrawal from the study signals that they do not have long to live, and thus the time of their censoring contains information about how much longer they will live. In a standard analysis, their survival rates would be overestimated because they would be based on patients who stayed in the study and had longer survival times. This overestimate would affect the difference in patient survival measured for the two treatments. Informative censoring is a special case of censoring that occurs in examples like this one in which the time of censoring contains information about the length of survival.

Conventional approaches for measuring the difference between two survival curves typically involve either the log-rank statistic or the Cox proportional hazards model. In the presence of informative censoring, conventional approaches do not compare the desired survival functions that underlie the curves. Rather, they compare biased survival functions.^{2,3} The result is an incorrect measure of the difference in survival with an incorrect confidence interval and *p* value. If censored subjects have shorter survival, as in the example above, the difference can either be too small (biased toward the null) or too large (biased away from the null), with the direction of the bias depending on the correlation and on the positioning of the underlying survival curves. The *p* value will be too large in the former case and too small in the latter case. In both cases, the confidence interval for the difference will be incorrect. These effects worsen as sample sizes get larger and with stronger correlations between the time of censoring and the time until death.⁴ These same issues are present when conventional methods, such as the Kaplan-Meier estimate, are used to analyze single curves. Therefore, informative censoring can be a serious issue when performing survival analysis.

The authors of the article used Kaplan-Meier estimation and the Cox proportional hazards model to analyze their data. In these analyses, they had to decide what to do about patients who underwent coronary revascularization, and they decided to censor these patients at the time of their procedures.

Nikolaidis and Johnston argue that this decision violated an assumption in the analyses that there was no correlation between the time of censoring and the survival time. They say it is likely that in some analyses censored patients had shorter survival and in other analyses they had longer survival and, further, that the effects of these differences could explain why some results were significant and others were not. Therefore, censoring was informative and the analyses could be invalid.

In response, Shaw and Hachamovitch point out that the approach they used is standard for this type of work. They also say that they analyzed their data using another method and found no important differences in their results, although this analysis was not reported in the article. In addition, they point out that relatively few patients had to be censored, which would lessen the effects of informative censoring even if it were present. Finally, they argue that the large sample of 3,620 patients also would have lessened any effects from informative censoring, but we know from the work of others that large samples make the problems worse, not better.⁴

In summary, we cannot determine whether informative censoring affected the results without more information about the duration of survival for censored patients, and this information is not available. New methods, however, have been developed for accommodating informative censoring in survival analyses. For example, Lin, Robins, and Wei describe a model-based approach that requires several assumptions.⁴ Robins describes an approach that

requires less stringent assumptions.^{5,6} Perhaps these approaches would be useful for additional analyses of the original data.—

WARREN BILKER, PhD, *Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Penn.*

REFERENCES

1. Shaw L, Hachamovitch R, Peterson E, et al. Using an outcomes-based approach to identify candidates for risk stratification after exercise treadmill testing. *J Gen Intern Med*. 1999;14:1-9.
2. Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA*. 1975;72:20-2.
3. Kalbfleish JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley; 1980.
4. Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*. 1996;83:381-93.
5. Robins JM. An analytic method for randomized trials with informative censoring, part I. *Lifetime Data Analysis*. 1995;1:241-54.
6. Robins JM. An analytic method for randomized trials with informative censoring, part II. *Lifetime Data Analysis*. 1995;1:417-34.

CALL FOR CLINICAL REVIEWS

The Editors of the *Journal of General Internal Medicine* remain interested in expanding the number of Clinical Reviews published in *JGIM*. We encourage authors and readers to submit Clinical Reviews on timely and relevant topics to the Journal. For more information on the kind of reviews we want, see our editorial "Up for Review," *J Gen Intern Med* 1995;10:293-4.

Please note that our approach to Clinical Reviews has broadened. We remain interested in systematic reviews, particularly those that address a specific clinical question and are evidence-based. However, we also are interested in more synthetic and summative reviews that address broader clinical issues and concepts. We recognize that some subjects, for example, updates, are better handled using formats that depart from the systematic review and follow a more traditional outline. If you are interested in submitting a review to the Journal or have questions about our requirements for authors, please contact our editorial office at:

Philadelphia VAMC, JGIM-111
University and Woodland Avenues
Philadelphia, PA 19104
Phone: (215) 823-4471 Fax: (215) 823-4450
E-mail: jgim@mail.med.upenn.edu

We look forward to receiving your submissions.