JACC Vol. 26, No. 6
November 15, 1995:1560–3

LETTERS TO THE EDITOR

Incremental Prognostic Value of Single-Photon Emission Computed Tomography

The report by Palmas et al. (1) in a recent issue of the Journal is incomplete. Interpretation of the study requires additional information about model adequacy and the methods used to assemble, follow and measure the study sample.

The authors used the terms Cox regression and logistic regression interchangeably. According to the methods section, they used Cox regression; however, they reported adjusted odds ratios and implied in the study limitations section that they used logistic regression. Although the Cox and logistic techniques produce similar results when duration of follow-up, event rate and censoring are appropriately constrained, the procedures are not synonymous. The former models the hazard rate, accommodates censoring and estimates relative risks; the latter models cumulative hazard to a fixed point in time, ignores actual survival time and estimates odds ratios, which may or may not approximate relative risks in a given context.*

If Palmas et al. used Cox regression, then they modeled the cause-specific hazard. If some deaths were noncardiac, they assumed independent competing risks. This is a strong assumption, considering that the mechanism of loss to follow-up was probably not end-of-study censoring and that the prognosis of the patients may have varied systematically over the long enrollment and follow-up period. If the assumption was invalid, the cause-specific hazard does not have the usual survival probability interpretation. Dependent competing risks can, in theory, distort the estimated prognostic effect to the point of reversing its sign; that is, the direction of the estimated clinical effect. Valid interpretation of the adjusted analyses requires a description of the censoring mechanism as well as a list of the noncardiac causes of death and their observed incidence.

Cause-specific mortality data are unreliable (2–4); yet there is nothing to indicate that the authors recognized or attempted to minimize this problem. To this end, they should describe the protocol that they used to ascertain cause of death and provide details about its implementation: for example, What rules did they use to classify cause of death? Were the rules applied by independent observers who were blinded to the single-photon emission computed tomographic (SPECT) results? Did they attempt to corroborate information derived from death certificates and medical record face sheets? What source took precedence when there were multiple sources of data concerning cause of death? How were deaths handled when the cause was uncertain?

The validity of the Cox model also depends on the legitimacy of the proportionality assumption; that is, the assumption that the hazard associated with different covariate patterns is constant over time. Proportional hazards, like independent competing risks, is a strong statistical assumption. It can and should be tested. There is no evidence in the report by Palmas et al. that it was. Failure to demonstrate that the model was a reasonable representation of the data further obfuscates the meaning of the SPECT coefficients.

Patient enrollment and testing in the study spans different therapeutic eras. Fifteen years elapsed between the time that the first patient was enrolled in the data base and the last patient underwent SPECT testing. The authors have tacitly assumed that there were no significant secular trends in the relation between known and unknown as well as measured and unmeasured postbypass prognostic indicators. This strong assumption is unrealistic in light of the profound changes in cardiovascular medicine since 1975 when the first study patient underwent coronary artery bypass graft surgery. The impact of secular trends on this analysis warrants investigation.

The results of the regression depend on the investigators' choice of candidate variables. They chose to evaluate the incremental prognostic value (beyond historical and treadmill data) of thallium-201 SPECT in patients ≥5 years after bypass surgery; however, the substantive question is whether SPECT testing provides incremental predictive value after accounting for routinely available prognostic indicators. In the preamble to the methods section, the authors list prognostic factors that are routinely available in postbypass patients, including extent and severity of disease in the native vessels as well as left ventricular function; however, they ignored these in the analysis. It is impossible to gage the impact of this oversight without additional data; however, depending on the correlations between covariates, inclusion of these factors in the regression could abolish the SPECT effects.

Palmas et al. disregarded several other modeling issues despite relying on an esoteric statistical paradigm that is probably incomprehensible to most readers of the Journal. 1) They made the tacit assumption that the adjusted relative risks are multiplicative. 2) They evaluated the plausibility of the assumption, that there is an exponential relation between continuous covariates and event rate, for only one of the reported continuous variables. 3) There is no indication that they searched for influential observations. 4) They did not describe variable scoring; thus, it is possible that they treated ordinal, and possibly nominal covariates, as interval data without offering a justification. 5) They did not investigate the stability of the model despite the possibility of a) multicollinearity, related to variables like the SPECT scores, and b) latent overfitting (i.e., the ratio of the number of hard events to the number of candidate variables appears to be considerably <10 to 1).

Questions regarding the validity of the statistical procedure notwithstanding, the study sample may not be an inception cohort. The use of an inception cohort minimizes selection of patients who are at particularly high or low risk for the event(s) of interest. Failure to assemble a proper inception cohort, and to obtain complete follow-up on it, "usually constitutes a fatal flaw" in prognostic studies (5). Consequently, a vague description of the assembly process is a major deficiency in prognostic studies. This is particularly true in retrospective analyses, where it is more difficult to know whether all patients who should have been counted were actually counted.

In this case, the description of the criteria and methods used to assemble the study sample is inadequate. The authors evaluated patients who underwent routine SPECT testing at least 5 years after bypass surgery. However, they did not stipulate that this included *all* patients who underwent routine SPECT, at least 5 years after bypass, at the institution in question. Moreover, they did not define "routine" SPECT testing. Referral for SPECT was not routine in the ordinary sense because only 8% of the patients in the data base had it. In addition, there is, in the methods section, a 741 patient difference between references to the total number of bypass patients; does the

*A direct estimate of relative risk can also be derived with logistic regression in cohort studies.

0735-1097/95/$9.50
0735-1097(95)00405-S

smaller number reflect the subset of data base patients without missing data? The apparent completeness of retrospective data can be a reporting artifact, that is, the data manager may have selected a subset of patients who had complete data on all study variables or the authors may have reported the number of patients who had SPECT testing, as opposed to the number of patients who contributed to estimation of the regression coefficients; that is, only patients with complete data on all candidate variables contribute to estimation of regression coefficients in the proportional hazards model. The latter appears to be the case (e.g., in Table 3, the electrocardiographic response to exercise was available in only 281 of the 294 patients).

To thoroughly characterize the sample, the authors should report the inclusion/exclusion criteria for the data base, as well as the analysis. They should demonstrate that covariates were measured at a uniform time, relative to bypass surgery, in all patients, with a summary of the distributions of time from bypass to measurement of historical, stress test and SPECT variables. In addition, they should recapitulate the indications for SPECT testing in the sample, and they should report the prescripts for terminating patient follow-up, as well as indicate the proportion of patients whose final data base entry was the last posting before loss of contact with data base personnel. They should also disclose the proportion of patients and events that were excluded by the regression algorithm.

The admonition at the end of the conclusions section would be adequate if the study design and analysis were beyond reproach. As it stands we do not know 1) how many patients there were or how they got into the analysis; 2) why they underwent bypass and SPECT testing; 3) how many were lost to follow-up and why; 4) accuracy of the cause of death assignment; 5) reliability of the regression coefficients; and 6) appropriateness of the statistical model. In short, in its present form the report adds little to the current state of knowledge regarding late prognostication after bypass surgery. It remains to be determined whether the observed increment in information is genuine and not, as suggested by the authors, whether it justifies the cost of thallium scans in these patients.

JOHN G. FERGUSON, MD
Director of Clinical Epidemiology
Memphis Vascular Research Foundation
910 Madison Avenue, Suite 710
Memphis, Tennessee 38103

References

1. Palmas W, Bingham S, Diamond GA, et al. Incremental prognostic value of exercise thallium-201 myocardial single-photon emission computed tomography late after coronary artery bypass surgery. J Am Coll Cardiol 1995;25:403–9.
2. Kirchner J, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. N Engl J Med 1985;313:1263–9.
3. Hanzlick R. Improving accuracy of death certificates [letter]. JAMA 1993;269:2850.
4. Kelsey JL, Thompson WD, Evans AS. Retrospective cohort studies. In: Lilienfeld AM, editor. Methods in Observational Epidemiology. New York, Oxford University Press, 1986:134.
5. Sackett DL, Haynes RB, Tugwell P. Clinical Epidemiology. A Basic Science for Clinical Medicine. Boston, Little Brown, 1985:162.

Reply

We appreciate Ferguson's comments with respect to our recent report (1). Ferguson outlines in great detail the various necessary steps for performing a prognostic study and raises questions regarding the epidemiologic and statistical validity of our analyses.

In our study, we used the Cox proportional hazards analysis throughout. We inadvertently labeled Table 4 as showing the odds ratio when, as Dr. Ferguson pointed out, risk ratio or relative risk would have been the correct label. We similarly referred to the logistic model in the limitations section of the discussion; Again, this is a misstatement, and the Cox proportional hazards analysis should have been mentioned instead. We apologize for these inaccuracies.

Statistical considerations. Although not specified in the report, the assumption of proportionality of hazards was tested, and variables were assessed for influential observations. The stability of the model was also investigated in a manner similar to that previously suggested by Ferguson et al. (2) and Diamond (3). Specifically, overfitting was scrupulously avoided by performing the regression analysis on only four variables, given the 41 observed events of interest. Furthermore, all variables were inspected for collinearity before entry into the model. In the case of highly related variables, only one of the two were introduced into the model. The variables entered into the Cox proportional hazards model were selected on statistical grounds; that is, no variables were entered or forced into a model based on previous performance or assumption. With respect to cardiac catheterization variables, only a minority of patients had a catheterization in close temporal proximity to their test date; thus, this information was not used in our analyses. All patients had catheterization data from the time of their bypass surgery, but these data would no longer be accurate. A small portion of the patient population had missing information; those variables with missing values were not included in the multivariate analysis but patients with missing data were not excluded from the multivariate analysis. No detectable trends were found with respect to these missing values. Ordinal data were evaluated as such in the multivariate model, and scores were treated as interval data.

Epidemiologic considerations. Dr. Ferguson raised a number of questions regarding patient selection, follow-up and determination of events. The patients were selected in the following manner: We identified all patients who had undergone exercise thallium single-photon emission computed tomography (SPECT) at least 5 years after their coronary artery bypass surgery. Ferguson correctly states that identification of all potentially eligible patients is crucial in this step. We have found that <5% of patients are missed by the searching techniques used for this study. The discrepancy between the patient numbers cited for the total population who had undergone bypass during the time interval of our study is explained by our reference to two different intervals; 3,741 patients underwent bypass between November 12, 1975 and December 27, 1984. A larger group of 4,186 patients were assembled in response to a reviewer's comments asking for the comparisons made in Table 2. The time interval used to define the 4,186 patients was inadvertently extended to patients who underwent bypass up to August 1, 1986. This error is limited to the analyses shown in Table 2, and probably did not materially affect the results shown.

Secular considerations. With respect to Ferguson's comments regarding the time span of this study, the purpose of the study was "to assess the incremental prognostic value . . . of thallium-201 myocardial perfusion single photon emission tomography in patients ≥5 years after bypass surgery" (1) rather than to assess the efficacy of bypass surgery itself. The study period examined was that between the date of the nuclear test (between June 1987 and September 1990) and the last follow-up date (rather than the time between initial bypass date and last date of follow-up). Thus, a maximal interval of <6 years elapsed during which the population was followed up (rather than the 15 years stated in the letter). We believe that this time period is too short for secular trends to have impacted on our group. Contrary to Ferguson's comments, no major changes in patient treatment or management over