LETTERS TO THE EDITOR

In response to:
Bodnar E, Blackstone EH. Editorial: An 'actual' problem: Another issue of apples and oranges. J Heart Valve Dis 2005;14:706-708

The recent editorial of Bodnar and Blackstone (1) suggests that the so-called 'actual' analysis has been misused at times and makes suggestions for limiting its future use. I generally agree with the editorial.

However, the method is mathematically valid, and does have some important uses. I fear that in overreacting to some misuses we run the risk of throwing out the baby with the bath water. In this note I discuss some of the mathematical background, and some applications where use of competing risks analysis is critical for proper understanding of a clinical situation. I agree with the editorial that the term 'actual' is potentially misleading, and other terminology should be used. The general area of analysis is often referred to as 'competing risks' analysis, and the term 'cumulative incidence' seems to be well accepted in this area. I agree with the suggestion that the term 'cumulative incidence' be used.

One point that must be made is that competing risks analysis rests on a completely sound mathematical footing, and cumulative incidence is a precisely defined mathematical concept. The general setting is that there are two (or more) competing risks. Each risk will have its own probability distribution, and the concept of the first event to be observed is precisely defined. Theoretical treatments are given in Kalbfleisch and Prentice (2), and Andersen et al. (3); formula 4.4.19 of the latter reference includes a derivation of the standard error. A rather more readable treatment, which clearly illustrates the difference between Kaplan-Meier (actuarial) analysis and competing risks analysis, is given by Gooley et al. (4).

The methodology of competing risks is standard in many medical areas. The paper of Gaynor et al. (5) discusses an oncology example with three competing risks; a standard error formula is given, but it is more difficult to use than the one in (3). The paper of Klein (6) is also in the oncology area; it discusses handling of covariates.

In all the above references the probability of a particular event being the first event observed is computed as a function of time; the term 'cumulative incidence' is used to denote the graph showing these probabilities. (Note that the term 'cumulative hazard' is often used for a different concept; the two should not be confused.)

In the valve examples studied by Blackstone and Kirklin (7) and by Grunkemeier at al. (8), there will be some true mathematical distribution for valve failure, and there will be another true mathematical distribution for death. In the actuarial analysis of valve failure death is merely a circumstance that prevents one from observing failure events, the end of follow-up in a clinical trial also prevents observation of failure. Both circumstances have the same meaning to the statistician, albeit not to the patient; in either case data are censored at the last time that the valve was known to be good. The Kaplan-Meier algorithm is well suited for estimating both the valve failure and death distributions, but there are circumstances where a parametric model is preferable.

There will also be some true mathematical distribution for the random variable *valve failure observed before death*. This distribution could be computed if the other two distributions were known; if one has a series of patient data the distribution must be estimated. In this situation a patient death directly affects the cumulative incidence estimate, because this represents a case where the first event will not be valve failure. The end of follow-up still causes censoring, since in such a case one does not know which event would ultimately occur first. The distribution to be estimated here is mathematically different from the distribution of valve failure, and the algorithms are accordingly different. Competing risks analysis is not implemented as a standard part of SAS®; however, SAS code, including the standard error estimates of (3), has been published by Anderson (9).

As an example of correct use of competing risks analysis, the valve studies (7) and (8) considered advising a patient as to his own future risks. Other situations would include an insurance company that might want to estimate the potential cost of replacement surgery, or a valve manufacturer that might want to estimate future sales of replacement valves. The Kaplan-Meier analysis would not be appropriate in any of these cases; meaningful results cannot be obtained without use of competing risks analysis.

Another example comes from left ventricular assist devices. There are current clinical studies involving a combined endpoint of death, device failure, and stroke. Whichever event comes first defines the endpoint for the particular patient; censoring is almost exclusively due to a patient being alive and event-free at the end of follow-up. In planning such a trial it is vital for a manufacturer to use competing risks analysis to study the cumulative incidence of the various components of the combined endpoint.

Finally, there are the numerous oncology studies, including (5) and (6) above; the paper of Freidlin and Korn (10) also includes a simulation comparing different analysis methodologies.

If one is going to present competing risks analysis, in

J Heart Valve Dis
Vol. 15. No. 2
March 2006

the above or other circumstances where the analysis is appropriate, the question remains as to how the results might be presented. I have two suggestions.

First, I suggest that it may be useful to present on one set of axes a graph with three (or more) curves. There will be one curve showing the cumulative incidence of each risk. There will be also one curve showing the freedom from all events. In these graphs the various cumulative incidence curves increase, and the freedom from all events decreases; at all times the values represented by the curves add to 1. There is no place in this graph for the Kaplan-Meier actuarial curves showing freedom from individual events; using the vocabulary of Blackstone and Bodnar, the competing risks curves present the apples, and the actuarial curves are the oranges.

An example of such a graph is shown in figure 1; the analysis is of explant due to SVD for a hypothetical bioprosthetic valve. Literature examples of such graphs include figure 1 in Banbury et al. (11), figure 3 in Blackstone and Lytle (12), and figure 3 in Kojori et al. (13)

Second, I suggest that one should never present the cumulative incidence curve without also presenting the corresponding actuarial curve somewhere in the manuscript. Even if the real interest is in the cumulative incidence, presenting both will go a way towards ensuring that readers do not misinterpret the results. The Kaplan-Meier graph for freedom from SVD, based on the same hypothetical data set, is shown in Figure 2.

The graphs are not complete without some measure of the errors involved. Both graphs presented here show the patients at risk, and these numbers are mathematically the same for the two graphs. The competing risks graph also shows 95% confidence limits, based on the matrix version of Greenwood's standard error (3), (9). One could also show error bars at selected time points. In figure 2 the error bars show 95% confidence limits, based on Greenwood's standard error and the asymmetrical logit transformation. At least one of the three representations of error should always be shown. The influence of age on valve failure presents an important issue, and it is natural to include age as a covariate in actuarial analysis. In this context actuarial analysis is generally performed using the Cox proportional hazards algorithm. The results of such an analysis are well known: valves last somewhat longer in older patients. If different valve models are being compared, the valve model effect can be separated from the age effect using proportional hazards.

Alternatively, patients can be stratified into age groups, and then the Kaplan-Meier algorithm can be used within each age group; slight differences in age distribution will not have much impact on the final results. Countless valve series have been presented using various age stratifications, and valuable comparisons can be made from published data.

The situation is considerably different with competing risks analysis. Age has a dramatic effect on patient survival, and this translates directly into a dramatic effect on cumulative incidence of valve failure. Since this effect goes in the same direction as the known effect of age on valve failure itself, the combination of the effects makes it virtually impossible to validly compare cumulative incidence curves from different valve series. I agree with the recommendation of Bodnar and Blackstone that this comparison should never be done. I would go farther, and suggest that cumulative incidence curves from different series should never be presented together in the same graph; the temptation to do an invalid comparison is simply too strong.

If one had complete outcome and covariate information for all patients, a proportional hazards version of the competing risks analysis could in principle be performed (2), (6), (14). No single analyst would generally have such data for valves from different manufacturers. Even if such data were available, I sug-

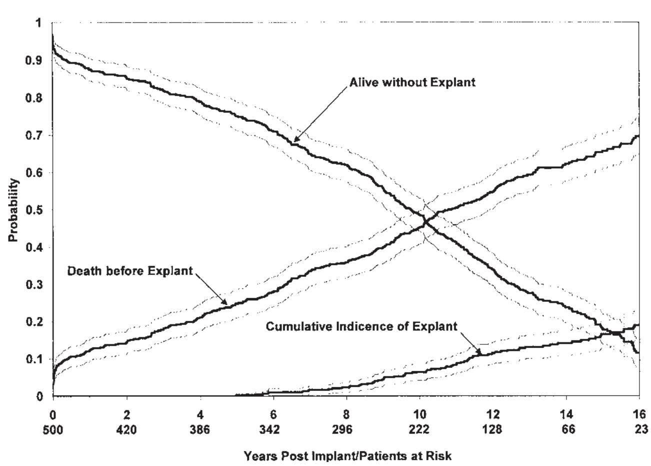

Figure 1: Competing risks for explant due to SVD
Hypothetical Bioprosthetic Valve
Dotted lines represent 95% confidence limits

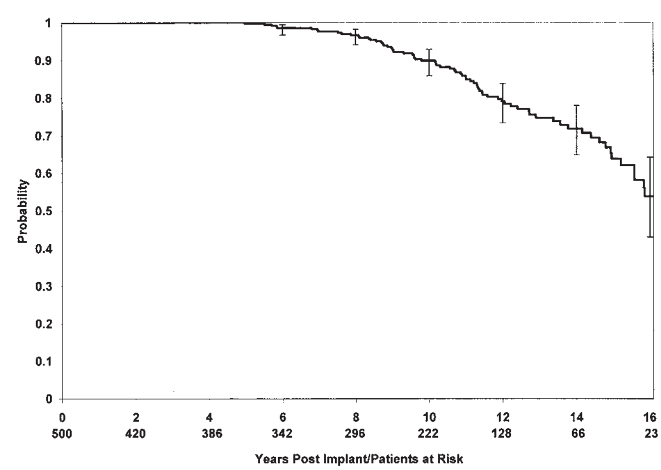

Figure2: Actuarial freedom from explant due to SVD
Hypothetical Bioprosthetic Valve
Error bars represent 95% confidence limits

gest that it would be much safer to compare valves using actuarial analysis; the latter removes one source of noise, and it does not seem that the competing risks analysis adds anything valuable to the comparison.

References:

1. Bodnar E., Blackstone EH. Editorial: An 'Actual' problem Another issue of apples and oranges. J Heart Valve Dis 2005; 14:706-708.
2. Kalbfleisch JD and Prentice RL. *The Statistical Analysis of Failure Time Data*, 2nd ed. Wiley-Interscience, Hoboken NJ, 2002.
3. Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*, Springer-Verlag, New York, 1993.
4. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Statistics in Medicine 1999;18:695-706.
5. Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data, J Am Stat Assoc 1993;88:400-409.
6. Klein JP. Modeling competing risks in cancer studies. Statistics in Medicine 2006 in press.
7. Blackstone EH, Kirklin JW. Recommendations for prophylactic removal of heart valve prostheses. J Heart Valve Dis 1992;1:3-14.
8. Grunkemeier GL, Jamieson WR, Miller DC, Starr A. Actuarial versus actual risk of porcine structural valve deterioration. J Thorac Cardiovasc Surg 1994;108:709-719.
9. Anderson WN. Algorithms for actuarial and actual analysis. In *Proceedings of Western Users of SAS Software conference*, FJ Sloan, ed. Scottsdale, AZ 2000.
10. Friedlin B, Korn EL, Testing treatment effects in the presence of competing risks, Statistics in Medicine 2005;24;1703-1712.
11. Banbury MK, Cosgrove DM, White JA, Blackstone EH, Frater RW, Okies JE. Age and valve size effect on the long-term durability of the Carpentier-Edwards aortic pericardial bioprosthesis. Ann Thorac Surg 2001;72:753-757.
12. Blackstone EH, Lytle BW, Competing risks after coronary bypass surgery: the influence of death on reintervention. J Thorac Cardiovasc Surg 2000;119:1221-1232.
13. Kojori F, Chen R, Caldarone CA, Merklinger S, Azakie A, Williams WG, Van Arsdell GS, Coles J, McCrindle BW. Outcomes of mitral valve replacement in childres: a competing risks analysis. J Thorac Cardiovasc Surg 2004;128;703-709.
14. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000.

William N. Anderson, PhD
21672 Montbury Drive, Lake Forest, CA 92630

In response to:
Bodnar E, Blackstone EH. Editorial: An 'actual' problem: Another issue of apples and oranges. J Heart Valve Dis 2005;14:706-708

The life of a bioprosthetic heart valve usually ends as a result of structural valve deterioration (SVD) or the death of the patient. In this competing-risk situation, the probability of SVD is estimated by the cumulative incidence function, sometimes referred to as 'actual' analysis. The recent editorial by Bodnar and Blackstone (BB) criticized this method and recommended its prohibition.

Specifically, BB object to:
1. Terminology: using the name 'actual' instead of 'cumulative incidence'.
2. Presentation: plotting its complement, and calling it 'actual freedom'.
3. Comparison: using this statistic instead of the Kaplan-Meier (KM) method to compare valve performance.

We disagree with BB on all three points:

Terminology

BB: *We should stop using the term 'actual freedom'... There is no reason to abandon the expression cumulative incidence...*

Statisticians use several names for this method, including 'cumulative incidence', 'crude probability', 'crude incidence', 'cause-specific failure probability', 'absolute cause-specific risk', and 'subdistribution function'. 'Actual' is just a shortcut name, like saying 'Ross' - which is not a medical term - instead of 'pulmonary autograft'. Moreover, it is appropriate because it estimates the percentage of patients who will 'actually' experience an event. 'Linearized rate' is not a statistical term, but was coined by cardiac surgeons at Stanford to mean '…the constant hazard rate of an exponential distribution'.

BB: *Synonyms of 'actual' are current, eventual, and real, and the implication of its use is that it is more real than the actuarial estimate.*

That is exactly the point; the term 'actual' is appropriate because it *is* 'more real' than the KM estimate (see below). However, since this term is no longer favored in this Journal, we will for now use the statistical terms 'subdistribution function' instead of 'actual failure' and 'subsurvival function' instead of 'actual freedom' - the terms favored by the prominent textbook on the subject (1).

Presentation

BB: *Other authors … displayed graphically the complement of cumulative incidence, which normally rises from zero to a certain positive value, so that the new curve declined from 100% to a certain value.*

Although the probability of having an event is the more direct concept, its complement - the probability of being event-free - contains the same information, and is usually preferred in survival analysis.

Comparison

BB: *Cumulative incidence... must not be used to define or compare valve performance. This should be done using actuarial [KM] methods.*

The first two issues are really a matter of taste, but this one is a matter of substance and statistical correctness. The subdistribution and subsurvival functions estimate the probability of having, or not having, respectively, SVD in the presence of the competing risk of death. As such, it would not be fair to use them to compare, say, valve A from a series with a high death rate to valve B from a series with a low death rate. However, if the death rates in two series were similar, then the comparison of SVD subdistribution or subsurvival functions would be appropriate (2).

The same can be said of KM estimates. Comparing high SVD rates with valve X in a younger population to low SVD rates with valve Y in an older population would not be correct, since the difference could be due to patient age alone. So to make a fair comparison of KM curves, the patient groups should also be similar. But there is a deeper, technical issue involved here. The subdistribution function gives a proper probability of experiencing SVD; the KM estimate does not.

BB: *We assume that patients who die before a non-fatal event occurs were just as likely, while they were alive, as anybody else to have experienced that event, even though they didn't." Now, that sounds more sensible, doesn't it?*

That is true, if indeed the events of death and SVD are independent. It may sound sensible, but it cannot be demonstrated using competing risks data (3,4), and for this reason the KM curve resulting from this assumption has dubious value. It has been universally criticized as: "…an incorrect use of the Kaplan-Meier method…" (2); "…a meaningless quantity" (5); "…irrelevant…" (6); and "…inappropriate for estimation purposes in the presence of competing risks, while the cumulative incidence is appropriate" (7).

That is the statistical argument. Then there is the clinical issue: the KM curve gives the probability SVD in the counterfactual situation where death has been eliminated. This of course will never happen.

BB: *In conclusion, the Journal of Heart Valve Disease will no longer publish 'actual freedom' results in articles reporting long-term performance of replacement valves.*

In light of the above, we feel that this editorial prohibition should be repealed.

References

1. Crowder M. Classical competing risks. Chapman & Hall/CRC, Boca Raton, 2001:200
2. Lin DY. Non-parametric inference for cumulative incidence functions in competing risks studies. Stat Med 1997;16:901-910
3. Cox DR. The analysis of exponentially distributed life-times with two types of failure. J Royal Stat Soc 1959;Series B, 21:411-421
4. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. Proc Natl Acad Sci USA 1975;72:20-22
5. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. Biometrics 2005;61:223-229
6. Pepe MS, Mori M, Longton G, Pettinger M, Fisher LD, Storb R. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? Stat Med 1993;12:737-751
7. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. Stat Med 1999;18:695-706

Gary L. Grunkemeier PhD
Medical Data Research Center
Providence Health System
Portland, Oregon, USA

Johanna J. M. Takkenberg MD, PhD
Dept. of Cardio-Thoracic Surgery
Erasmus University Medical Center
Rotterdam, The Netherlands

W. R. Eric Jamieson MD
St. Paul's Hospital - Heart Centre
University of British Columbia
Vancouver, British Columbia, Canada

D. Craig Miller MD
Department of Cardiothoracic Surgery
Stanford University Medical School
Stanford, California, USA

Reply

We are gratified by Dr. Anderson's solid endorsement of our editorial, helping to clarify this rather confused and confusing situation surrounding the use of the word 'actual' in the assessment of intrinsic valve performances. We fully agree with him regarding the importance of the concept of competing risk environment. His proposals on reporting long-term experiences with replacement heart valves are excellent, and we recommend that our future authors follow his advice.

The letter by Drs.Grunkemeier, Takkenberg, Jamieson and Miller incorrectly summarizes in its first paragraph the content of our editorial and therefore adds unnecessary confusion. Dr. Bodnar was probably the first to introduce competing risks to cardiac surgery; Dr. Blackstone and colleagues have used competing risks analyses extensively in multiple settings of adult and, particularly, congenital heart disease, and both will continue to do so to answer questions that methodology was designed to answer. The method is not the problem! The problem is its inappropriate use in answering questions related to intrinsic properties of heart valve substitutes, such as comparative durability. From its beginning, our editorial was clear about the specific, focused context of our remarks. Rather than restate the entire editorial to make one point clear, perhaps an analogy would be helpful.

Pretend we wish to compare durability (tread life) of three brands of tire. (We understand that, just like structural valve deterioration [SVD], this is a continuous process and not an event; for simplicity, however, we will estimate durability by the time to change tires because of excessive tread wear, just as we may estimate SVD by time to valve replacement.) Likely there are specific risk factors for tire durability, but let us imagine that miles traveled is the strongest (just as young age is the strongest correlate of SVD). Now, let us say that our study of tire durability for brand A is dominated by a fleet of cars operated by traveling salespersons, study of brand B by commuters living within 5 miles of work, and study of brand C by mothers with children involved in many after-school activities. After a stated period, most salespersons have had to change tires, a number of mothers have, but few short-distance commuters have. Should we conclude that A and C tires wear out too fast, and we should switch to brand B? If we instead compare the tires in a distance-specific fashion using time-to-event (actuarially based) analysis methods, we would be getting closer to a fair comparison of brands A, B, and C by isolating tire properties from the driving specifics of their owners. In the case of SVD, the only universally found risk factor is age of patient at implant, so age-specific durability provides a reasonable comparison of bioprosthetic device durability. As with the tires, estimates of durability are properly made by time-to-event-type analyses.

You may then ask, what tire should we recommend for the driving habits of a given person? The answer may well depend on how long different people intend to keep their cars. A person who leases a new car every 3 years no matter how little he or she drives may never need to change tires. Trading in a car, like death of a patient, is a competing risk. For someone who drives little, there is no reason to pay a premium for a tire of superior durability if at trade-in there is unlikely to be much tread wear! However, it is important to understand that the car-trading habits of owners do not themselves affect intrinsic properties of the tires. So, interesting as competing risks analysis is for answering some questions such as just posed, for simply comparing intrinsic durability of tires, these trading habits are extraneous and should not be allowed to confound the comparison.

Choice of a prosthesis for a given patient, like a tire, may well depend on a number of factors, such as the patient's expected longevity, which itself has many correlations that are unrelated to intrinsic properties of the device. These are appropriately evaluated by competing risks analyses. However, longevity of the person, like car trading habits, is a far more complex matter than the intrinsic properties of the prostheses. Estimates of whether or not a patient will live to experience SVD of their prosthesis will be different for every patient, despite the prosthesis retaining its unchanged intrinsic durability.

It is important to recognize that in *both* time-to-event and competing risks methods, time to SVD and time to death are assumed to be completely independent of one another. Thus, the fundamental "actuarial" calculation for each component of a competing risks analysis can be performed independently of one another, after which a mathematical combination is made. It follows that actuarial estimates are fundamental and unchanging, whereas competing risk estimates will vary, depending on what set of events is considered competing or ignored. (You may protest that it is irrational to think there is no linkage between SVD and death, and you are probably right. However, that is the assumption that both "actuarial" and competing risks methods assume. Methods for dealing with linked events, called methods for "informative censoring," are still not well developed.)

In conclusion, we suggest that Drs. Grunkemeier, Takkenberg, Jamieson and Miller think again, very carefully, to realize the magnitude of the mistake they make. Actuarial analysis and cumulative incidence assessment are *not competing but complementary methods*. One is apple the other orange, and oranges should not be blamed for not being apples. The question is not which is better, the question is which should be used for what purpose. Cumulative incidence cannot be used to assess let alone compare intrinsic valve properties. Therefore, we repeat, *The Journal of Heart Valve Disease* will no longer publish 'actual freedom' results in articles reporting long-term performances of replacement valves. The Journal will, however, readily publish cumulative incidence if relevant, but not as the measure of intrinsic valve characteristics.

Endre Bodnar MD
Eugene H. Blackstone MD