# From Signal/Noise to Information Content/Noise

## Reconsidering the Statistical Analysis of Continuous ST-Segment Data Streams With Gaps: Potential Optimization of Application-specific Information Content Using Left, Right, and Interval Censoring

Mitchell W. Krucoff, MD,* Cindy L. Green, MS,*
Frank E. Harrell, PhD,* Maarten L. Simoons, MD,†
Nancy M. Wilderman,* Kathleen M. Trollinger, RN,*
Sharon T. Sawchak, RN,* James E. Pope, MD,*
Akbar Shah, MD,* Christopher B. Granger, MD,*
Robert M. Califf, MD,* and Kerry L. Lee, PhD*

The philosophy, execution, and interpretation of how to optimize signal/noise content in computerized electrocardiographic (ECG) devices has traditionally focused on the fidelity of the record-playback loop, the accuracy of measurements, and the characteristics of the editing and display human interface environments. In an era in which derivation of computationally demanding continuous ECG parameters and transforms is facile, and with the development of integrated information systems in which ECG information overall is only a component of the total patient data analyzed, conceptual expansion from signal/noise optimization to information content/noise optimization could be useful. For continuous ST-segment data streams, this conceptual expansion could help track the influence of compromises in technical specifications or compromises in data quality on parameters affecting interpretative statements. The information content/noise focus could also be a useful way to consider whether limitations in raw signal recording or processing could be compensated for by statistical strategies applied to the parameters derived from the processed signal before they are actually interpreted for clinical or research purposes.

Two fundamental areas of ST-segment recording and analysis have received little explicit attention in the more than 500 papers published using this methodology: the definition of a baseline or reference ST level and the analysis of continuous data in which gaps are created by artifact.

*From the *Ischemia Monitoring Laboratory and Cardiovascular Databank, Duke University Medical Center, Durham, North Carolina, and †Thorax Center, Rotterdam, The Netherlands.*

Reprint requests: Mitchell Krucoff, MD, Duke University/VA Medical Center, 508 Fulton Street (111A) Durham, NC 27705.

We have previously detailed a self-referenced, continuously updated approach to baseline/reference level definition for the noninvasive detection of failed reperfusion using continuously updated ST-segment recovery analysis in patients presenting with acute myocardial infarction.[1,2] In the current discussion, considerations will be turned to the management of data gaps in the context of continuous ST-segment recovery analysis, also in patients suffering acute infarction treated with thrombolytic therapy. The methods proposed are considered of potential use for application in randomized trial designs comparing treatment regimens over populations but probably not for real-time use in individualized patient care.

### Overview of the Method

The application exampled will be a randomized clinical trial comparing drug A versus drug B for the induction of reperfusion during acute infarction. The measure of drug activity is taken as the speed and stability of 50% or greater ST-segment recovery, as a surrogate for angiographic reperfusion. The data from which the 50% ST-segment recovery variable is derived are taken from continuous ST-segment monitoring devices that archive ECG waveform measurements with the time and date of their acquisition. These digital archives are then uploaded to an independent computer analysis environment.

Using current standards for ST-segment analysis for such data streams, some arbitrary limits on absent or artifact-laden periods creating gaps in the data stream would be

imposed. Patient studies falling within the set limits would be included in the data analyzed for 50% or greater ST-segment recovery. Studies outside these arbitrary limits would be excluded from analysis as "technical failures." Such technical failures are generally of two types: either the data gap occurs at a crucial moment (eg, at the moment when some other correlated endpoint is simultaneously gathered) or the data gaps eliminate too great a portion of the monitoring period to support therapy-based conclusions (eg, artifact involving more than 50% of the total monitoring period). The technical failure codes from the Duke University Ischemia Monitoring Laboratory, used in the TAMI, DUCCS, GUSTO, and currently in other trials, are listed in Table 1.

Using such cutoffs, the whole study population denominator is thus divided into two groups: analyzable and technical failures. The analyzable data are used to produce parameters of interest, and an interpretative statistical analysis is performed to assess drug A versus drug B effects.

Using this classical approach, many individual studies included in the analysis contain substantial periods of artifact or gaps. Parameters derived from these studies are absorbed into the analytic conclusions while the noise content of each study itself becomes completely transparent

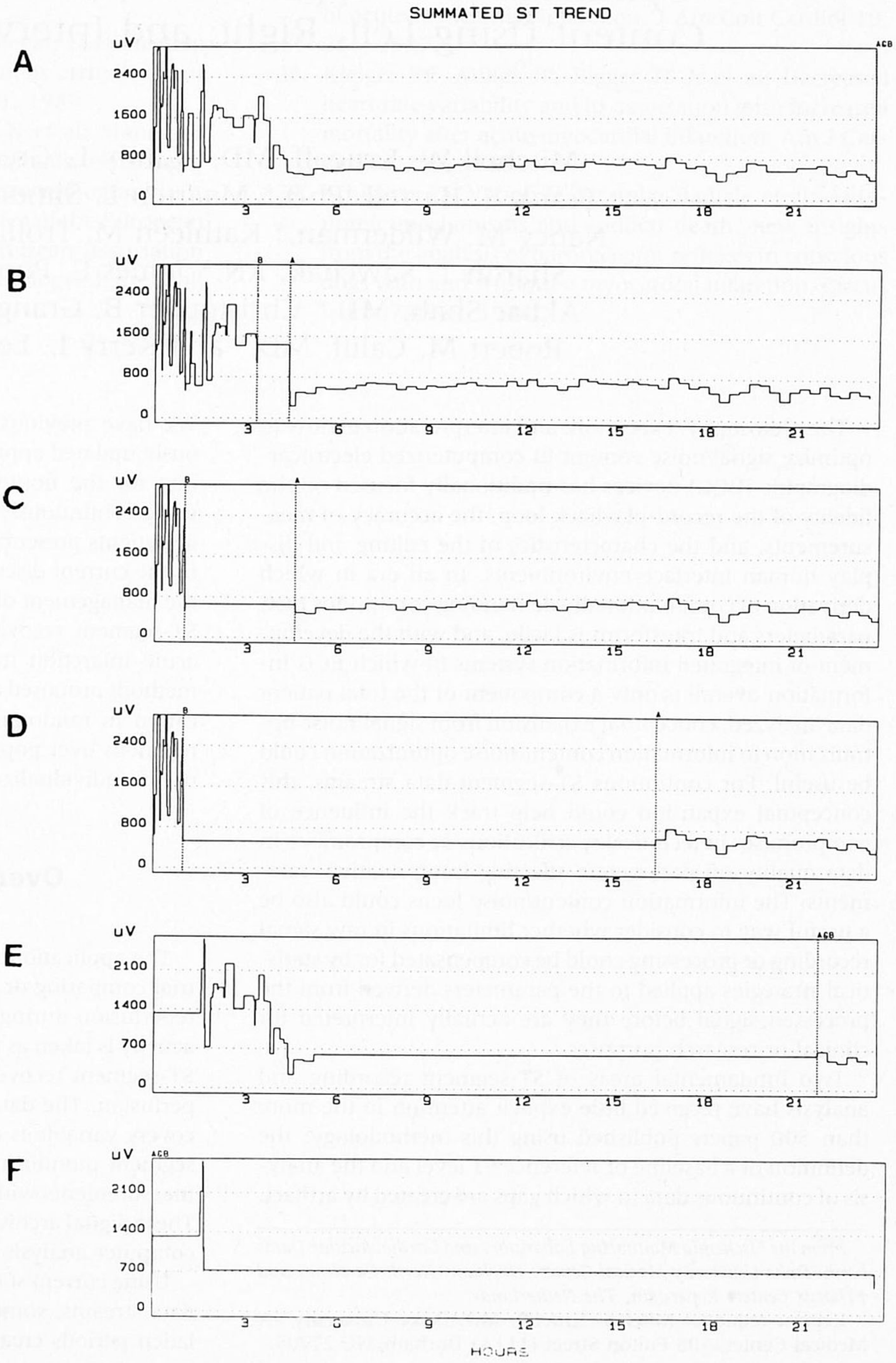


**Fig. 1.** Influence of gaps on continuous ST-segment trend of 12-lead ST levels versus time in a patient with inferior infarction and cyclic coronary flow following thrombolytics. (A) Complete study, no gaps. (B) 59-minute gap during ST recovery. (C) 3-hour gap during early ST recovery. (D) 15-hour gap prior to and following ST recovery. (E) 3.5 hours of ST recovery information with gaps before and after (F) total of nine ST levels from 24 hours of monitoring.

**Table 1.** Technical Failure Codes

| Code No. | Code Definition |
|---|---|
| 1 | Late hookup: >1 hour after lytics |
| 2 | Late hookup: >2 hours after lytics |
| 3 | Late hookup: >3 hours after lytics |
| 4 | Late hookup: >6 hours after lytics |
| 5 | Late hookup: >12 hours after lytics |
| 6 | No ST monitor hookup within 24 hours |
| 7 | Noise/artifact: >50% of monitoring time |
| 8 | Noise/artifact: >1 hour of first 3 hours |
| 9 | Noise/artifact: during catheterization or intervention |
| 10 | Noise/artifact: steady state indeterminate |
| 11 | Noise/artifact: other period of interest |
| 12 | Data gap: >50% of monitoring time |
| 13 | Data gap: >1 hour of first 3 hours |
| 14 | Data gap: during catheterization or intervention |
| 15 | Data gap: steady state indeterminate |
| 16 | Data gap: other period of interest |
| 17 | BBB/V-Pacer: >50% of monitoring time |
| 18 | BBB/V-Pacer: >1 hour of first 3 hours |
| 19 | BBB/V-Pacer: during catheterization or intervention |
| 20 | BBB/V-Pacer: steady state indeterminate |
| 21 | BBB/V-Pacer: other period of interest |

Codes are from the Ischemic Monitoring Laboratory, Duke University Medical Center.

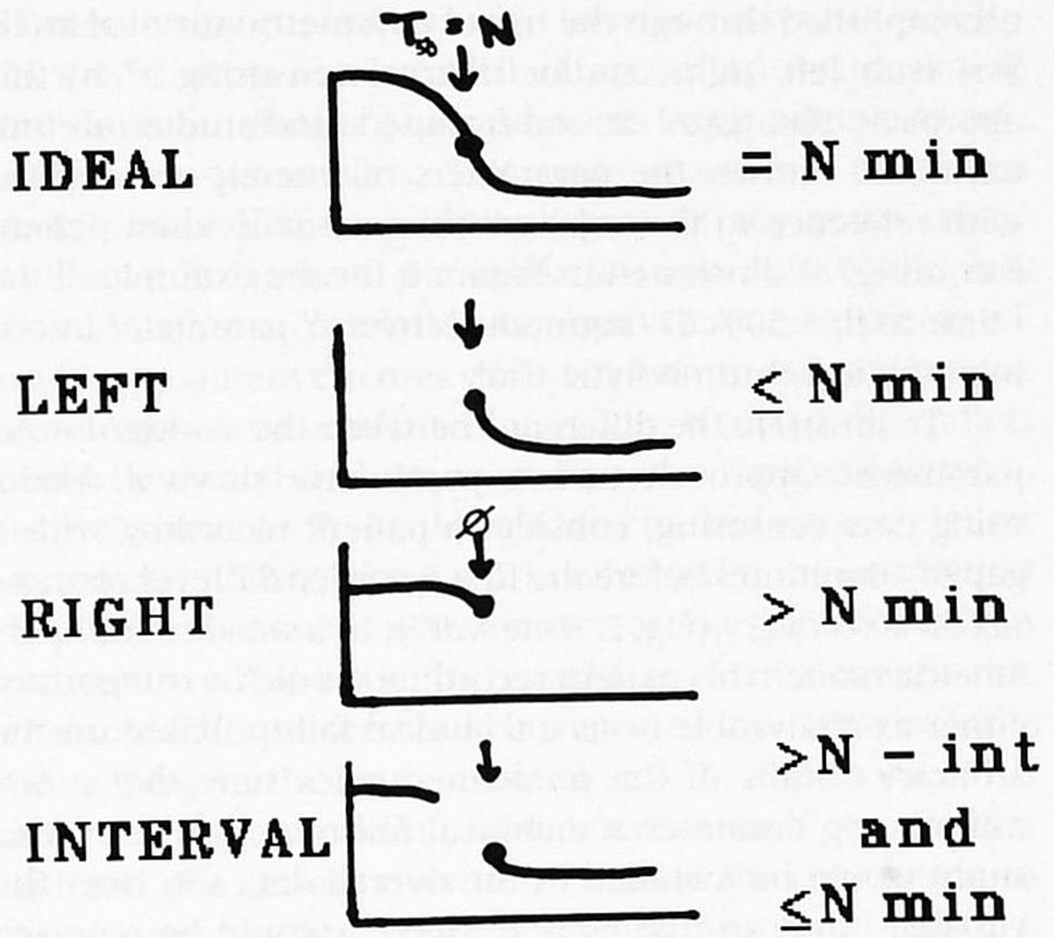


**Fig. 2.** Interval censoring of reported time to 50% recovery with data gaps in different locations. From top to bottom, trends showing ideal absence of gaps, early gap, late gap, and interval gap, which would be treated with no, left, right, and interval censoring, respectively. As shown at the right of each trend, this approach generates differentially weighted parameter statements for time to 50% recovery: precise number of N minutes (ideal), ≤ number of N minutes (left), > number of N minutes (right), and > and ≤ number of N minutes (interval).

within the analysis of the final parameters considered to reflect drug effect. Conversely, many studies excluded as technical failures may nonetheless have substantial information content. The loss of such studies from the analysis not only excludes the available information in the recordings but also reduces the overall denominator of patients analyzed and so reduces the statistical power of the study to detect more subtle differences or to characterize important subgroups within the drug treatment effect analysis.

Most trials published using ST-segment monitoring endpoints do not even report either the exclusion criteria defining technical failures or the number of patients excluded based on such criteria. Many trials omit both, either in the analytic process or in the reporting of the data. From available information in acute infarction trials using any ST-segment monitoring device, from 25 to 50% of patients entering the trials may be excluded from at least one primary analysis if criteria are carefully defined and applied.

In Figure 1, a series of ST recordings during acute infarction is shown. In Figure 2A, the data stream shows an uninterrupted depiction of the continuous surveillance mode of the ST-100 12-lead ST monitor (Mortara Instrument, Milwaukee, WI), as has been detailed previously.[3] As can be seen Figure 1B, in some recordings the overall profile of ST-segment recovery is so well preserved that there is almost an intuitive impulse to ignore the gap or "connect the dots" and impute the missing ST values. In other studies, such as in Figure 1C and D, the gaps get larger, and even the subjectively perceived ability to impute accurate information wanes. In other studies, the gap duration predominates over ST-segment values altogether. In studies such as Figure 1E, despite the absence of data overall, the period in which data are recorded is highly active and potentially informative. In other studies, such as Figure 1F, so little data are actually recorded over the monitoring period that there is almost no impression as to the status of the infarct artery or the effect of the drug being measured.

An accurate method for imputing the missing data would be the most ideal solution. Such a method could even potentially be applied at the bedside in individual patient care. However, as can be seen in Figure 1, the range of potential ST-segment behavior even in short time periods is extremely variable, from no change at all to hundreds of microvolts in only seconds. Given this heterogeneity, a method of imputation that would suffice for these purposes seems remote, unless a large enough database could be gathered to develop a robust probability of activity in any definable time period around some index event. If, for instance, the likelihood of coronary reocclusion with ST-segment activity was far greater in the first hour after thrombolytic therapy than in the twenty-first hour after thrombolytic therapy as assessed in 1,000 patients with complete ST recordings, then a probability of ST activity within a data gap might be imputed differently in a data gap in the first hour compared with a gap of similar duration in the twenty-first hour. The number of assumptions incorporated into this approach and the database that would be necessary to generate a robust basis for imputation seem prohibitive.

The statistical treatment of data gaps within continuous ECG streams as a part of the derivation of the endpoint parameters themselves might be better considered as an alternative to the standard approach of excluding studies based on arbitrary thresholds. Such an approach might be

accomplished through the use of parametric survival analysis with left, right, and/or interval censoring.[4,5] In this approach, data gaps are not imputed, and studies are not excluded. Rather, the parameters of interest are defined with reference to the gaps within each individual patient recording, as illustrated in Figure 2 for the example of the "time to first 50% ST-segment recovery" parameter in our hypothetical thrombolytic trial.

To illustrate the difference between the standard nonparametric approach and a parametric survival model using data censoring, consider a patient recording with a gap of 40 minutes before the first recorded ST level connoting 50% recovery (Fig. 2 "Interval"). In a standard nonparametric model, this patient recording would be categorized either as analyzable or as a technical failure based on the arbitrary cutoffs. If the predefined rules state that a 50-minute gap connotes a technical failure, analysis of this study would be included in the overall data set. How the variable "time to first 50% recovery" would be reported in minutes from therapy, however, would remain arbitrary, as either the first real value at the end of the gap, as the last real value at the beginning of the gap, or as some intermediate value between those two. In every case, such a value would include some imputed estimation of the real timing of 50% ST recovery. Once assigned a value in minutes, however, the value would enter the database and be included in the final analysis of drug effect without distinction from more precise values generated from truly continuous studies. Thus, in the final database, each study deemed technically analyzable would produce a time assigned (in minutes) to the moment of 50% recovery. The amount of imputed information incorporated in the derivation of any of these values would be completely invisible when they were matched to drug treatment assignments to address the primary investigational hypothesis. In a well-designed trial, this problem might be solved by randomization per se, producing an equivalent amount of imputed information within each drug arm's ST recovery parameter; however, there is no systematic way to ensure that such was the case for any given trial. In smaller trials or nonrandomized trials, the problem could be even more pronounced.

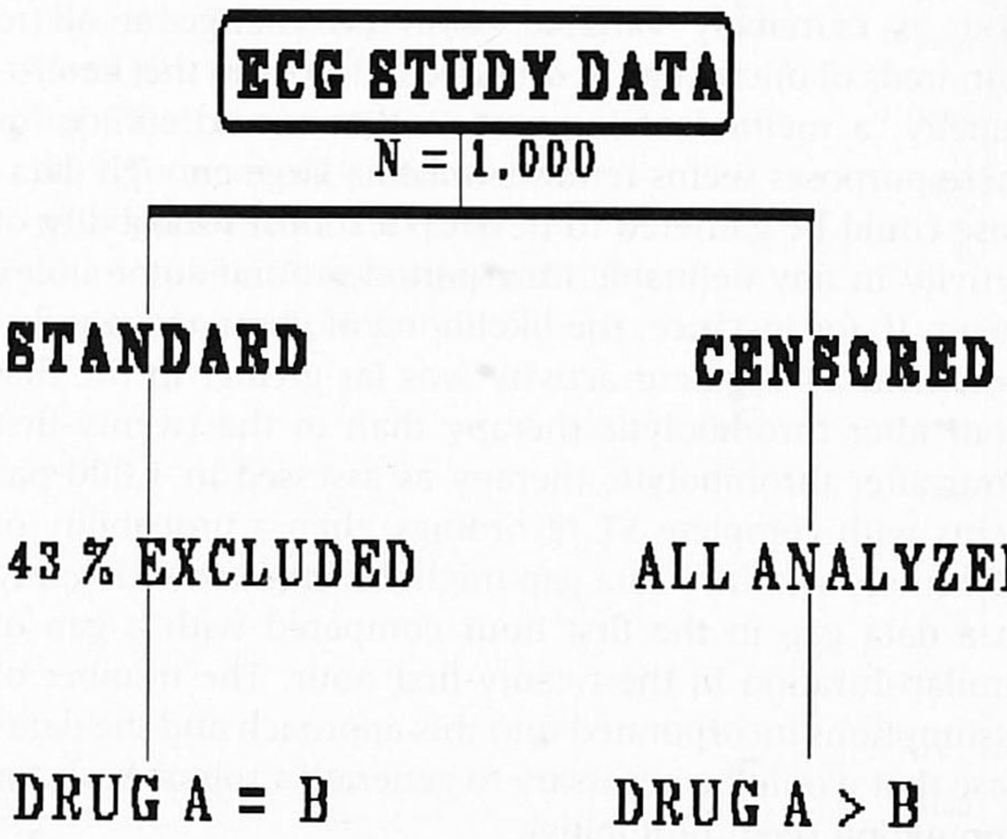


**Fig. 3.** Dilemma of the greater than standard method and censored model, yielding different analytic statements with no referee as to whether a smaller denominator (standard) or highly fit censored model gives more insight into relative drug effects.

Interval censoring and a parametric modeling approach could be adopted as a statistical attempt to circumvent both the necessity of excluding studies from the denominator—all studies would be included—and the necessity to impute or assign values that were any more specific than the actual recorded data allow. Thus, for the study in Figure 2 "Interval", the value (in minutes)assigned to 50% ST recovery would correspond to the duration of the interval between the last recorded ST level (showing less than 50% recovery) and the onset of recording after the gap (showing more than 50% recovery). If the time of onset of the gap is X minutes from the time of therapy and the duration of the gap is 40 minutes, the value assigned to the time to 50% recovery parameter would be "$>$X minutes and $\leq$X + 40 minutes." In an ideal parametric model, when values for 50% recovery were correlated to treatment assignments to address the primary investigational hypothesis, this censored statement would be incorporated with less weight than the more precise values (in minutes) derived from truly continuous patient recordings but with more weight than statements from other studies with even longer gaps. Thus, at least conceptually, this approach to analysis of ST recordings would be an attempt to combine the preservation of the overall denominator with a censoring of noise/gaps information instead of an invisible incorporation of noise/gap information.

Although conceptually intriguing, the actual performance of this approach in an acute myocardial infarction ST-segment data set would itself have to be tested. Theoretical assumptions within parametric survival models using data censoring might be inappropriate for such nonnormally distributed, nonlinear data. "Forcing" the application if such a mismatch is substantial could lead to total nonsense rather than to the intended improvement in analytic capabilities. For example, as illustrated in Figure 3, say a data set of 1,000 patients is collected for our randomized trial of thrombolytic therapy. In the standard approach, the technical failure rate based on predefined rules is 43%. Of the remaining 570 patients, randomized across the two therapies, time to 50% recovery suggests that there is no difference between drug A and drug B. Using the parametric model, however, including precise and censored statements to described 50% recovery in all 1,000 patients, time to 50% recovery suggests that drug A is more effective than drug B. On a theoretical basis, it is impossible to predict which conclusion is actually correct. In the final section of this discussion, we consider at least one approach to the assessment of the strengths or weaknesses of parametric modeling in ST-segment data stream information.

The focus of an approach to this methodologic validation is well served by returning to the concept of signal/noise content of the ST data streams, and therefore of any parameters derived from ST data streams. In that sense, the standard approach of predefining technical failures is similar

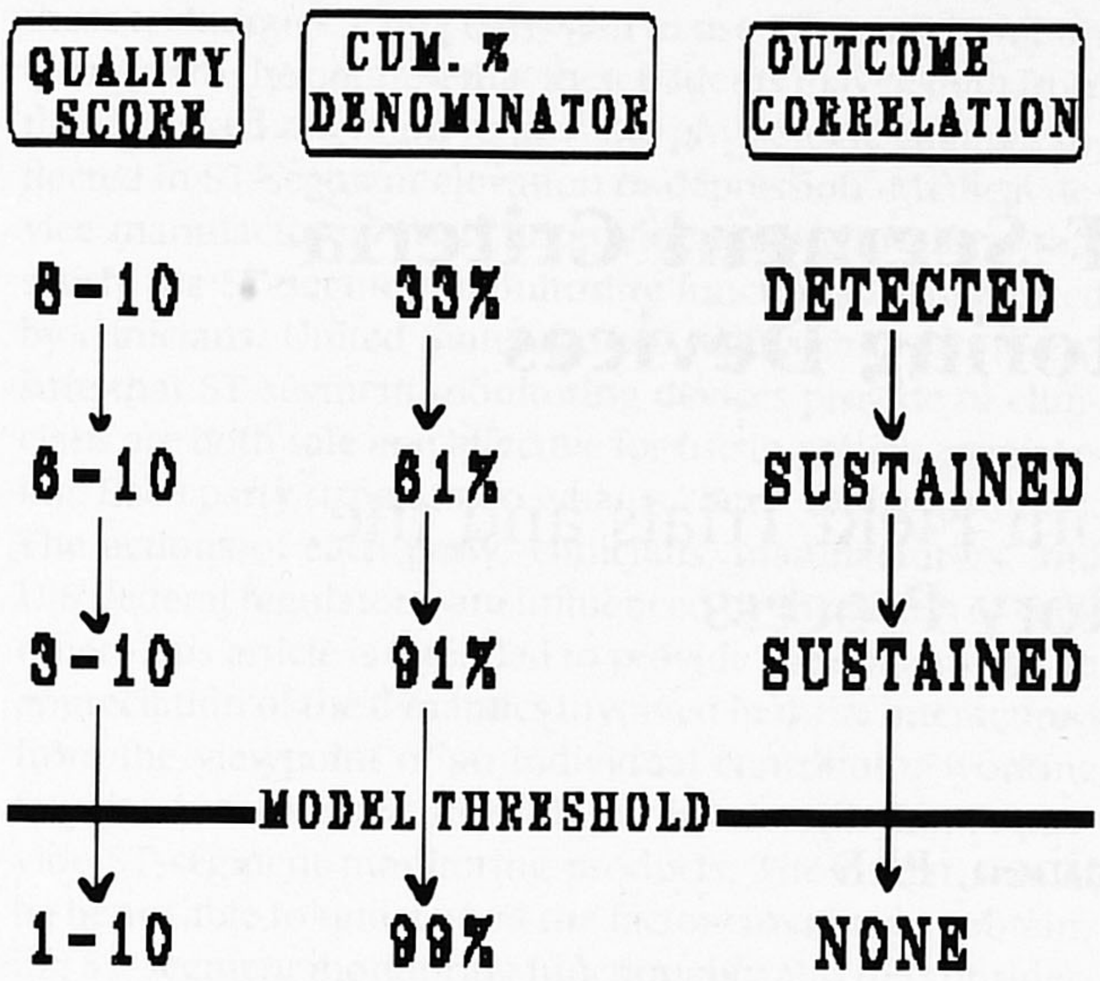


**Fig. 4.** Incremental testing scheme to determine the performance of the censoring model over progressively poorer information content studies.

to a bandpass filter threshold. All information beyond the threshold is simply eliminated on the presumption that that information contains more obscuring noise than useful signal. From this perspective, the goal of data censoring within a parametric model could be regarded as a shift from a fixed bandpass filter to a statistical filter executing an attempt to enhance the quality of data around the threshold level and even to recover information and analytic power from studies that are well beyond that threshold (ie, studies that are heavily polluted with noise or gaps). The test of such a statistical filter's performance would thus need to be a test of how much power versus how much pollution was incorporated by its use.

There are many possible approaches to such validation testing. One might be to look at the conclusions reached compared with other markers within the same clinical trial, such as angiographic information, and see if they are consistent. If the angiographic data implied that drug A and drug B, for instance, were no different, one might suspect that the modeling process created a nonsense statement if it suggests that drug B was better. However, it would be impossible to discriminate this conclusion from the possibility that the modeled ST data was, in fact, elucidating some component of the disease and its response to therapy that was not captured by the angiogram.

A more potentially robust approach to validation would be to establish a known relationship between the marker and some outcome of interest and then observe the performance of the modeling process as more and more "polluted" studies were added in a stepwise fashion. This validation approach is illustrated in Figure 4. An essentially qualitative grading scale for the early and late content of ST monitoring studies recorded from acute myocardial infarction trials is first established according to the timing and duration of data gaps and noise levels relative to primary study parameters. This scale provides a crude but useful index of the "pollution" level of any given study. As shown in Figure 4, the validation approach would first seek to define a correlation between the most optimal ST studies (qualitative scores of 9–10) and an outcome of interest, such as survival after infarction. If such a correlation could be established, then studies of a "more polluted" nature (scores 6–7) would be added in, and the correlation reanalyzed. If the "filter" works effectively, the correlation should be sustained, with an increase in the information content by virtue of the increased size of the denominator analyzed. This process could be repeated in a stepwise fashion, adding in the studies scored at 5, 4, and so on, with the parametric modeling reanalyzed at each step. When addition of a certain noise level caused the correlation to deteriorate, the limit of the modeling process to improve information content would essentially have been defined. If that limit occurred very early, one might interpret it to mean that the assumptions of the model about the nature of the data are being violated in too primary a manner to use the strengths of the censoring process. If that limit occurred late or not at all, however, then one might be more comfortable concluding that the standard nonparametric approach showed no difference between drugs A and B because the analysis became underpowered after eliminating more than 30% of the denominator and that the suggestion that drug B was better using the parametric approach to left, right, and interval censoring was, in fact, a statistically sound conclusion.

Thus, parametric modeling using left, right, and interval censoring suggests a unique and innovative approach to the analysis of continuous ECG data streams for investigational applications. This method does not impute data into data gaps or otherwise reactivity noise and so is not proposed as a solution to individual patient study problems for real-time clinical management per se. Rather, this method appears to have potential to optimize the information content of analyses performed on parameters derived from ST-segment data streams with gaps by including the entire population but creating differential weighting of analyzed variables based on the actual continuity of each individual data stream from which the variable is derived. Before applying such a model to comparisons of unknowns, such as therapeutic effects, the modeling process will need validation of its ability to actually distill information out of noisy ST studies. This validation could be established by showing incremental information gain from the enhanced denominator more than information loss from data characteristics that do not precisely fit the underlying assumptions of the model itself. In this proposal, the salient theme remains, in an era of tools that allow us to pursue truly unique analytic approaches to all levels of data management, to broaden the theme of signal/noise content to one of information/noise content.