Draft Recommendations
Clinical Trials

Criteria for markers as surrogate endpoints

A true "surrogate" endpoint defines a clinically meaningful endpoint. In doing so, it must be shown that the surrogate fully captures the effect of an intervention on the endpoint of interest (DeMets and Fleming). To do this requires knowing the effect of the intervention on the clinical endpoint of interest. Thus, an issue becomes one of validation, which requires phase III clinical trials. For validating surrogate endpoints in phase III trials the criteria of Prentice are recommended (Prentice). That is the proposed surrogate endpoint must be a correlate of the clinical endpoint, and the proposed surrogate must fully capture the net effect of treatment on the clinical outcome.

Use of markers as surrogate endpoints and issues of trial design

To use biomarkers as true surrogate endpoints, the criteria outlined above must first be met. How biomarkers are employed as intermediate endpoints depends on the phase of study and a study's sample size requirements (or restrictions) and goals for regulatory approval and labeling. Often biomarkers are used as intermediate endpoints to guage for "a signal" of the effect of treatment in early phase clinical testing (e.g decrease in BNP or hsCRP with new treatment) while minimizing sample size, but with the intent to pursue further definitive testing on clinically meaningful endpoints if there is a promising signal. Biomarkers alone should not serve as the only criterion for measuring treatment effect unless the above criteria for a surrogate have been met, however. Example: Hormone replacement therapy has favorable effects on lipids, but has no effect or in come cases increases clinical cardiovascular events.

Biomarkers are also commonly used as a component of the definition for a clinical endpoint, such as myocardial infarction. In addition, assessing biomarkers may provide insight into pathophysiology or mechanisms of action of a new agent, particularly when employed in early phase testing. Combining biomarkers with other variables, for example area under the CK-MB release curve with continuous ST-segment monitoring may be useful in reducing sample size and detecting "a signal" for effect of a new agent used to treat myocardial infarction. This would not substitute for ultimately testing the effect of the treatment in reducing mortality from myocardial infarction, however.

Implications of various "cutoff" values for a biomarker

The cutoff used to define an event based on biomarker assessment has implications for both identifying patients for enrollment in clinical trials and in defining the occurrence of a clinical event. In terms of enrollment, the cutoff selected can affect selection of the correct population, one that is likely to experience benefit from treatment greater than risks of treatment. It must be remembered that there is both assay and biological variability at cutoff that leads to imprecision in diagnosis. In addition, elevations of a biomarker in a disease state other than the disease of interest/under study could lead to erroneous diagnosis and inclusion in clinical trials or registries. For this reason, in defining myocardial infarction, for example, both clinical and electrocardiographic characteristics must accompany elevation of CKMB or troponin.

In defining the endpoint of a clinical trial (again using myocardial infarction as an example) many factors must be considered, including signal to noise ratio at various "cutoffs" and the need to optimize the ability to detect a treatment effect. Although standardization across CKMB mass assays has largely been achieved, particularly in international trials, activity assays are still used. Standardization of the troponin I assays remains to be completed and none of the currently available troponin assays meets the ESC/ACC requirements for precision of 10% CV at the reference limit.

One way to circumvent the limitations of standardization of existing assays and use of different assays at different sites is to employ a core laboratory in which a single assay is used. Although this will not solve the problem of assay precision, the degree of imprecision is at least uniform. Downsides to this approach, particularly in large, multicenter clinical trials are logistics and cost, both at the site and to the sponsor. However, for smaller studies this may be a reasonable option.

Short of a core laboratory, it has been proposed that sites be allowed to use their own local assays, but to centrally define the assay's "cutoff" value for the study to be the level at which 10% CV can be achieved with that assay (Apple, Wu and Jaffe). This approach is easier for the site and less costly than a core laboratory, but particularly for large, multicenter clinical trials cataloging and tracking this information and changes in it over the course of a trial would be impossible.

Finally, particularly for multicenter clinical trials, allowing sites to use their own assays, but centrally defining a multiple of the local reference limit to define myocardial infarction is widely used. Although this approach cannot address standardization or precision, it places the definition in context of clinical practice at that site. Using a cutoff of 2x the ULN at the site accomplishes a similar objective as the approach of using the $\leq$10% imprecision level, but is simpler to implement. While awaiting assay standardization and improvements in precision, existing databases could be employed to define what the best cutoffs (for troponin or CKMB) are to optimize signal to noise ratio in defining the clinical endpoint of myocardial infarction. It is recommended that regardless of which approach is used, the myocardial infarction endpoint be adjudicated by a committee blinded to treatment assignment and using both biomarker data and clinical data to determine the occurrence of the endpoint.