The introduction of newer classes of chemotherapeutic agents, with varying mechanisms of action by which they affect tumor growth and viability, has challenged the traditional norms of clinical trial design and drug approval in oncology. Most notably, the emergence of cytostatic biologic agents with antitumor efficacy has necessitated reassessment of appropriate primary endpoints for phase II and III trials in advanced disease from both a clinical and regulatory standpoint. Recent data in the field establishes an endpoint hierarchy, which places progression-free survival (PFS) between overall survival (OS) and response rate (RR) as appropriate primary endpoints for assessing the clinical efficacy of cytostatic and cytotoxic agents.

The success of clinical trials and the relevance of their results to clinical practice depend on the choice of appropriate endpoints for the hypothesis being examined. In this context, endpoints must meet certain requirements. First, an endpoint must be measurable; second, it must be sensitive to the effect of treatment. Biologic activity can be used early in the development of a new agent to assess the potential for patient benefit, but in the longer term, proof of clinical efficacy is required for successful phase II and III clinical trials. Third, endpoints must be clinically relevant. Clinical relevance may be related to the outcome for an individual patient, or more commonly, it is related to a shift in the distribution of outcomes for an entire study population.

Statistical Endpoint Hierarchy

In a phase III trial setting (and in general), an endpoint hierarchy must be considered. The most preferable endpoint for any trial design is a true clinical efficacy measure. Overall survival (OS), historically the "gold standard" of phase III trials in oncology, is a true clinical efficacy measure and remains the most unambiguous and universally agreed upon primary endpoint. However, in most tumors, the use of OS as an endpoint requires considerable time for patient follow-up, a high associated cost for clinical trials, and thus delays the introduction of possibly beneficial therapy to patients. An endpoint that could be assessed more rapidly would more quickly answer/address the trial hypotheses, limit the number of patients required per study, and potentially expedite the drug approval process. Additionally, OS may be insensitive to a true beneficial drug effect when patients receive subsequent lines of therapy with an impact on OS. Thus, identification of alternate but appropriate primary endpoints for assessing the effect of new treatment regimens on patient outcome has been central to the evolution of clinical research in oncology.

These considerations naturally lead to the search for validated surrogate endpoints that accurately predict clinical efficacy. Validated surrogate endpoints are obtained sooner, at less cost, or less invasively than the true endpoint of interest. Validated surrogates are rare, as the formal process of validation is burdensome, time-consuming, and sometimes despite all efforts, simply fails. Current efforts are ongoing to validate progression-free survival (PFS) as an endpoint for palliative phase II and III trials in various tumor entities as it relates to OS.[1] The advantages of a validated surrogate, were it available, would make it an endpoint of choice in the drug development and approval process.

The next lower endpoint in the hierarchy (below a validated surrogate) is a surrogate endpoint that is reasonably likely to predict clinical benefit (but not formally validated). This has been the status of tumor response rate (RR) from the perspective of the US Food and Drug Administration (FDA). Because tumors rarely spontaneously regress, RR is considered a measure of clinical activity. There are sufficient data and literature to suggest that there is some patient benefit to receiving a response, and for this reason, RR has been used as the basis for an accelerated approval process for multiple chemotherapeutic agents in oncology.[2,3] However, in the era of biologic therapies, RR would not be an appropriate endpoint for a trial investigating a purported cytostatic or other novel chemotherapeutic agent that impacts tumor growth by nonconventional mechanisms. Additionally, surrogate endpoints that are reasonably likely to predict clinical benefit are rarely accepted as an
endpoint for full drug approval.

At the bottom of the endpoint hierarchy is the great majority of other possible trial endpoints that are solely measures of biologic activity, with unclear clinical relevance. The relationship of these endpoints to patient outcome may not fully be understood, and therefore their use as a measure of clinical efficacy is questionable.

Surrogate Endpoints

The overall goal associated with the use of a surrogate endpoint is to allow the same inference to be made as if the desired clinical endpoint had been used more quickly, but to allow this inference and at less cost. Various endpoints have historically been used as surrogates: tumor burden outcomes; time to progression (TTP, or progression-free survival [PFS]) and objective response rate (ORR); and biomarkers, such as carcinoembryonic antigen (CEA) in colon cancer, prostate-specific antigen (PSA) in prostate cancer, or CA-125 in ovarian cancer. While the use of any of these endpoints can establish biologic activity, their use as a primary endpoint in a study when it has yet to be understood whether they truly reflect clinical efficacy may result in an outcome of uncertain clinical relevance.

Surrogate endpoints fail for a variety of reasons. As outlined by Fleming and DeMets,[4] a particular surrogate may not be intrinsic or active in the causal pathway of the disease process. The disease process may indeed affect the surrogate endpoint, but if the surrogate is not linked to the true clinical outcome, then the treatment may impact the surrogate endpoint but this impact would not translate into efficacy as measured by the true clinical endpoint (Figure 1A). For example, the tumor marker CEA has been used as a surrogate marker in advanced colon cancer. However, it remains unclear that a treatment that impacts CEA levels will in turn impact clinically relevant outcome measures such as OS.

Surrogate endpoints may also fail if the disease impacts the true clinical outcome via multiple processes, but the surrogate endpoint is contained in only one of those (Figure 1B). The intervention may affect the disease pathway containing the surrogate endpoint, but not other relevant pathways. In this instance, a large effect of the treatment on the surrogate can translate to a small effect on the true clinical outcome, thus the surrogate may suggest a false-positive result.

Alternatively, the surrogate endpoint may either not be in the pathway of the intervention's effect, or may be insensitive to its effect (Figure 1C). The intervention could have no effect on the surrogate endpoint, but still impact the true clinical outcome. This would, in turn, lead to a false-negative result. This scenario is particularly germane to trials investigating cytostatic agents, and it demonstrates why the use of RR as a surrogate endpoint could miss a beneficial effect on clinically relevant patient outcomes such as OS.

Finally, the use of a surrogate endpoint may fail when the intervention has multiple mechanisms of action related to the true clinical outcome, and these mechanisms of action may (or may not) include the surrogate endpoint (Figure 1D). For example, the surrogate endpoint of RR will not detect the impact of a grade 5 adverse event on OS, as in this case the therapy negatively impacts the true
clinical outcome independently of the surrogate.

Validated Surrogate Endpoints

Because of the many reasons that a surrogate endpoint may fail to accurately predict a true endpoint, appropriate validation concerning how the surrogate relates to patient outcome can enhance its validity as an endpoint in phase II and III trial settings. Validation is the formal establishment that the effect of the intervention on the surrogate endpoint reliably predicts the effect of the intervention on the clinical endpoint. Thus, we consider the topic of how a surrogate endpoint may be properly validated.

The marriage of both the clinical and statistical perspective is necessary to provide a compelling argument when validating a surrogate endpoint for a specific disease setting. From a clinical perspective, validation requires a comprehensive understanding of both the causal pathways of the disease process and the intervention's intended (and unintended) mechanism(s) of action. Without a sound clinical basis for an endpoint's relevance to a particular trial, any statistical validation is without merit. From the statistical perspective, surrogate-endpoint validation requires a pooled or meta-analysis of clinical trial data. Multiple clinical trials—and three trials with 500 patients each are preferable to one with 1,500—must be considered in such an analysis to demonstrate a consistency of effect. Reasonable heterogeneity in the trials is desirable, as heterogeneity minimizes the impact of any one particular treatment or trial, and allows for a more broadly based conclusion and robust validation.

Our group (the Adjuvant Colon Cancer Endpoints Group, ACCENT) recently validated 3-year disease-free survival (DFS) as a surrogate endpoint for 5-year OS for phase III trials in the adjuvant setting for colon cancer.[5] This pooled-analysis included 18 different clinical trials that tested various fluorouracil (5-FU)-based regimens, and included 20,898 patients, with a median follow-up of 8 years for all trials and with ≤ 5 years of follow-up in 93% of patients. Multiple techniques were used to formally examine the surrogacy. In one simple graphical approach, we plotted the hazard ratio between arms within study for 3-year DFS against the same for 5-year OS, and observed a tight linear relationship with a correlation coefficient of 0.94 (Figure 2). Notably, the hazard ratios correlated in both the positive and negative direction. For example, for 33 of 43 study arms analyzed, differences between 3-year DFS and 5-year OS were ≤ 3%, with the largest single difference for any single study arm at 8%. On the basis of this and multiple other analyses, we concluded that we can reliably predict 5-year OS using 3-year DFS in adjuvant trials for colon cancer.

In light of these data, the Oncologic Drug Advisory Committee (ODAC) of the FDA recently accepted 3-year DFS as an endpoint to allow full drug approval in stage II and III colon cancer. Three-year DFS subsequently served as basis for the FDA-approval of oxaliplatin (Eloxatin) and capecitabine (Xeloda) as adjuvant treatment for stage III colon cancer in 2004 and 2005.

Response Rate Caveats

Tumor response rate is widely used as an endpoint for phase II trials in oncology. Despite the fact that RR has repeatedly been shown not to correlate with OS, RR has been useful as a trial endpoint.
due to the rapidity, relative ease, and presumed accuracy of measurement. Indeed, it has been noted that improvements in RR do not necessarily translate to OS benefit from a clinical or regulatory perspective,[2-4,6-8] and, conversely, it has been shown that, for instance, patients with advanced colon cancer still receive an OS benefit from "better" treatment even if they do not respond.[9] In a subset analysis of their pivotal, randomized controlled, phase III trial (n = 813) in responding and non-responding colorectal cancer patients receiving irinotecan/5-FU/leucovorin (IFL) + bevacizumab (Avastin), Mass et al [9] concluded that the magnitude of clinical benefit associated with bevacizumab treatment (measured by hazard ratio [HR] for PFS and OS) was statistically similar regardless of objective tumor response. For PFS, respective HRs were .66 (all patients), .60 (responders), and .76 (nonresponders); for OS, HRs were .54 (all patients), .53 (responders), and .63 (nonresponders). Objective tumor RR measured by the Response Evaluation Criteria in Solid Tumors (RECIST) criteria were significantly different between study arms: IFL + bevacizumab (Avastin) (n = 402), 44.8% vs IFL + placebo (n = 411), 34.8% (P < .004).[10] However based on this analysis, the benefit associated with bevacizumab on OS was independent of objective tumor response. The authors indicated that "response-independent survival benefit" might be uniquely attributable to the targeted therapy's anti-angiogenic (cytostatic) mechanism of action, or may reflect a more general indication that RR does not measure the full potential impact of an agent on OS.

Alternative Endpoints for Clinical Trials

As exemplified above in a large pivotal trial of targeted therapy, RR is inadequate as a comprehensive measure of patient benefit as it lacks sensitivity for measuring a cytostatic mechanism of action, such as those putative mechanisms possessed by novel biologic agents. In view of such shortcomings, several alternative endpoints for phase II trials should be explored. The duration of stable disease or the proportion of patients who are progression-free at a particular time point (such as 6 or 12 months from the initiation of therapy) are proving to be powerful alternatives to RR in this setting. Another possible surrogate endpoint is the proportion of patients who achieve the target biologic response.

From a statistical standpoint, there are numerous myths concerning the use of PFS as a phase II endpoint. The first myth is that more patients are required for trials with PFS than with RR as the endpoint. In phase II trials, it is useful to evaluate the efficacy of a treatment regimen by evaluating the percentage of patients who are progression-free at a particular time point. Compared to use of an RR endpoint, correctly choosing a clinically relevant time point abrogates any need for an increased sample size using a PFS endpoint. In terms of the trial design, the time point for the analysis of the PFS endpoint may need to be extended, compared with the detection of response, but a larger sample size is not necessarily required. However, assessing RR as an endpoint also takes time, in particular, since per RECIST all observed responses must be confirmed. However, this point does underscore the fact that correctly picking the time point is critical for efficiently performing a PFS analysis.

A second myth involving PFS as a primary endpoint in a phase II trial is that its use necessitates concurrent controls. Whether (or not) this is true depends more critically on the trial setting rather than the type of endpoint one wishes to use, as well as the tumor type and treatment under investigation. In many cases, historical data for PFS are arguably as reliable as those for RR. If one has faith in the historical data for RR, then one should also have faith in the PFS historical data. However, it is noted that in multiple tumor types (eg, renal cell carcinoma [RCC]), historical controls may not accurately be used as a comparator, depending on patient risk factors, staging, performance status, etc. Clearly, in tumor types where the body of historical evidence does not allow a consensus, concurrently randomized, controlled PFS data are desirable.

Regulatory Issues and Trial Design

At the May 2004 ODAC meeting, in addition to accepting DFS as a validated surrogate for OS in the adjuvant colon cancer setting, ODAC also unanimously voted to prefer PFS (time from randomization until objective tumor progression or death) to TTP (time from randomization until objective tumor progression) as a potential phase III trial endpoint in advanced colon cancer. An important point that influenced this decision is that PFS accounts for fatal toxicity whereas TTP does not. Additionally, TTP analysis can be complicated because the exact date of progression may be difficult to assess. Patients can be taken off therapy due to clinical progressions not prespecified in the study protocol, such as elevation of tumor markers. Patients also go off study for reasons other than progression, such as toxicity and/or withdrawal of consent. These patients are generally not available for a rigorous follow-up schedule, and the date of progression therefore may be difficult to assess. ODAC also preferred PFS to TTP because it accounts for the fact that drugs have mechanisms of action.
unrelated to their tumor activity. When considering reliability, reproducibility, and sensitivity to biologics, PFS should be considered superior to TTP as a trial endpoint.

Conclusion
The field of surrogate endpoints in oncology is rapidly evolving. Large data sets are required to validate surrogate endpoints, but the possible time and cost savings provide the impetus to embark on such analyses. PFS and, in some situations, TTP, are measurable, clinically relevant constructs that in most cases require further validation regarding how they relate to OS as phase II and III trial endpoints that may be gained through the use of a validated surrogate. If so validated, TTP/PFS could replace RR as an endpoint for phase II trials of both cytotoxic and cytostatic agents. It is hypothesized that in tumor types other than advanced colon cancer (provided a large patient data set provides the opportunity for meaningful statistical analysis), PFS may eventually be able to be applied as a surrogate endpoint for OS (eg, RCC, lung cancer).

The importance in achieving a gold standard, OS, or a suitable validated surrogate, reinforces the need to recognize that RR may not necessarily predict clinical benefit in many tumor types or with novel agents. Additionally, the use of both OS and PFS as primary and surrogate endpoints, respectively, underscores the importance of the statistical endpoint hierarchy in the study design of phase II and III clinical trials. For example, while PFS may or may not be valid as a surrogate endpoint in tumor types such as RCC (pending statistical validation), it does not appear that RR is a surrogate for survival in this tumor type (based on the data to date). Further prospective and retrospective investigation is warranted using data from trials of both standard chemotherapy as well as targeted agents to attempt to validate surrogates for overall survival, similar in concept to the analytical and validation processes for 3-year DFS as an endpoint in adjuvant colon cancer.

Disclosures:
Dr. Sargent is a consultant for Sanofi-Aventis.

References:


**Source URL:**
http://www.cancernetwork.com/oncology-journal/general-and-statistical-hierarchy-appropriate-biologic-endpoints

**Links:**
[1] http://www.cancernetwork.com/review-article
[2] http://www.cancernetwork.com/oncology-journal
[5] http://www.cancernetwork.com/authors/daniel-sargent-phd