Surrogate Endpoints And FDA's Accelerated Approval Process

The challenges are greater than they seem.

by Thomas R. Fleming

ABSTRACT: There is interest in approaches allowing more rapid availability of new interventions, particularly for diseases providing risks of death or serious illness. The accelerated-approval regulatory process is intended to address this need by allowing marketing of interventions shown to have strong effects on measures of biological activity, if those measures are potential "surrogates" for true measures of tangible clinical benefit. To use surrogate endpoints and the accelerated-approval process, challenging issues must be addressed to avoid compromising what is truly in the best interest of public health: the reliable as well as timely evaluation of an intervention's safety and efficacy.

I
N CLINICAL TRIALS INTENDED TO PROVIDE the pivotal evidence for regulatory approval for marketing of drugs, biologics, or devices, the primary goal typically is to obtain definitive evidence regarding the benefit-to-risk profile of the experimental intervention relative to a placebo or an existing standard-of-care treatment. One of the most challenging and controversial issues in designing such trials relates to the choice of the primary-efficacy endpoint or outcome measure used to assess benefit. Given that such trials should provide reliable evidence about benefit as well as risk, the primary-efficacy endpoints preferably should be clinical efficacy measures—that is, measures that unequivocally reflect tangible benefit to patients. For example, for life-threatening diseases, one would like to determine the effect of the intervention on mortality or on a clinically significant measure of quality of life, such as relief of disease-related symptoms, improvement in ability to carry out normal activities, or reduced hospitalization time.

Establishing that an experimental drug can provide quality-of-life or survival benefit in a newly diagnosed patient with prostate or breast cancer, or that a vaccine can reduce the spread of HIV, or that a device can reduce risk of serious illness or death from cardiovascular disease could require trials that are large, long term, and financially costly.

In many instances, sponsors have proposed alternative endpoints (that is, "surrogates") for these clinical endpoints, to reduce the duration and size of the trials. A common approach has been to identify a biological marker that is "correlated"

. . . .

Thomas Fleming (tfleming@u.washington.edu) is a professor and chair of the Department of Biostatistics, University of Washington, in Seattle.

with the clinical efficacy endpoint (meaning that patients having better results for the biological marker tend to have better results for the clinical efficacy endpoint) and then to document the treatment's effect on this biomarker, where this effect is expected to be relatively large in magnitude and evident early in time. In oncology, for example, one might attempt to show that the experimental treatment regimen induces tumor shrinkage (likely transient), delays tumor growth in some patients, or improves levels of biomarkers such as carcinoembryonic antigen (CEA) in colorectal cancer or prostate-specific antigen (PSA) in prostate cancer. Although these effects do not prove that the patient will derive symptom relief or prolongation of survival, such effects on the biomarker are of interest because it is well known that patients with worsening levels of these biological markers have greater risk for disease-related symptoms or death.

Unfortunately, demonstrating treatment effects on these biological "surrogate" endpoints, while clearly establishing biological activity, may not provide reliable evidence about effects of the intervention on clinical efficacy measures. This paper considers issues related to validating surrogate endpoints—that is, identifying when effects on biological markers would accurately predict when treatment truly provides tangible benefit to patients. It proposes an endpoint hierarchy characterizing the relative reliability of outcome measures when used to evaluate clinical benefit. Finally, it considers the controversial issues in the implementation of the Food and Drug Administration's (FDA's) accelerated-approval process, where treatments only known to be biologically active can be marketed to the public while scientific trials are under way to determine whether these agents truly are more effective than toxic.

'A Correlate Does Not A Surrogate Make'

How could it be that a treatment could provide an effect on a biological measure that is correlated with a clinical-efficacy endpoint but not also provide a meaningful effect on that endpoint? Here I consider some illustrations to explain why a biological marker could be a "correlate" of a clinical efficacy measure yet might not be a valid "surrogate," or why "a correlate does not a surrogate make."[1]

The first of these illustrations is a situation where a disease causally influences a biomarker as well as the true clinical-efficacy endpoint (Exhibit 1). As a result, the biomarker is correlated with the clinical endpoint. However, if this biomarker does not lie in the pathway by which the disease process actually influences the occurrence of the clinical endpoint, then affecting the biomarker might not, in fact, affect the clinical endpoint.

Consider the examples presented in Exhibit 1. First, suppose one would like to identify an intervention that reduces the risk of transmitting HIV from an infected mother to her infant. It is well known that pregnant women who have more advanced stages of HIV infection have higher viral loads. There are at least two important consequences of these high viral loads: These women tend to have lower

EXHIBIT 1
A Reason For The Unreliability Of A Proposed Surrogate Endpoint: The Proposed Surrogate Is Not In The Causal Pathway Of The Disease Process
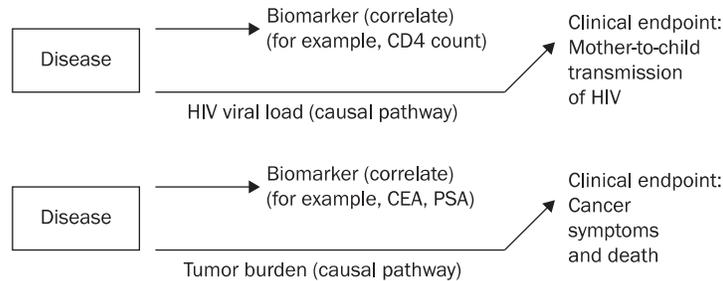
SOURCE: T.R. Fleming and D.L. DeMets, "Surrogate End Points in Clinical Trials: Are We Being Misled?" *Annals of Internal Medicine* 125, no. 7 (1996): 605–613.
NOTES: Correlates are useful for disease diagnosis or for assessing prognosis. Valid surrogates are useful for replacement endpoints. See discussion in text. CEA is carcinoembryonic antigen. PSA is prostate-specific antigen.

CD4 cell counts, and they are more likely to transmit HIV infection to their infants. It then is not surprising that clinical data have clearly shown that maternal CD4 count is strongly correlated with the risk of mother-to-child transmission of HIV.[2] However, if interleukin-2 were provided to the mother before labor and delivery to raise her CD4 count, this would not be expected to affect the risk of transmission. Rather, the causal approach here would be to do something to reduce her viral load.

In the oncology setting, the causal mechanism by which the disease process induces risk of mortality or serious morbidity, such as cancer-related symptoms, is likely predominantly through advancing tumor burdens. Elevated levels of tumor markers such as CEA and PSA in colon cancer and prostate cancer, respectively, also are the result of advancing tumor burdens. Hence, such markers are clearly correlated with the level of disease and with morbidity/mortality risks and thus may be effective tools for early detection of disease or for assessing prognosis. However, CEA and PSA are not the mechanism through which the disease process induces increased risk of the clinical-efficacy outcomes, so it is questionable whether treatment-induced changes in these markers could be relied upon to accurately predict treatment-induced effects on the clinical endpoints.

In essence, as noted in Exhibit 1, it is adequate for markers such as CEA and PSA to be "correlates" to be useful tools for disease detection, or to counsel patients regarding their prognosis. However, "validated surrogates" are required if one wants to use replacement endpoints in trials designed to reliably provide accurate estimates of the level of tangible benefit to patients.

Additional explanations for why a biological marker could be a "correlate" of a clinical-efficacy measure yet might not be a valid "surrogate" are provided in Exhibit 2. First, usually there are multiple pathways through which the disease process influences the risk of the clinical-efficacy endpoints. If the proposed surrogate

EXHIBIT 2
Additional Reasons For The Unreliability Of Proposed Surrogates: Disease Processes Having Multiple Causal Pathways And Interventions Having Mechanisms Of Action Independent Of The Disease Process
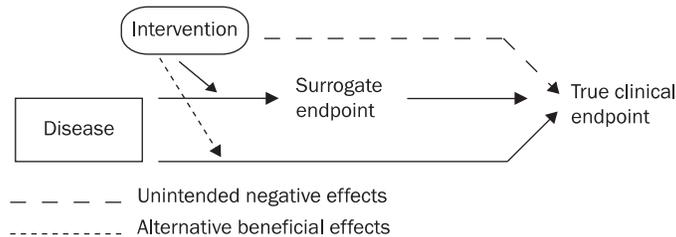
endpoint lies in only one of these pathways and if the intervention does not actually affect all pathways, then the effect of treatment on clinical efficacy endpoints could be over- or underestimated by the effect on the proposed surrogate.

Second, the intervention itself could have mechanisms of action that are independent of its intended effects on the disease process. Very often, because such effects are unintended, they are unanticipated, unrecognized, and unrecorded. A classic example of this was seen in the setting of ventricular arrhythmia after myocardial infarction. Arrhythmias are a known risk factor for sudden cardiac death, and the drugs encainide and flecainide are very effective in suppressing them. As a result, patients and cardiologists were persuaded that these drugs would have a positive effect on the clinical outcome of primary interest—sudden cardiac death—mediated through this effect on arrhythmias. In fact, they were so persuaded that between a quarter-million and a half-million patients each year in the United States alone were receiving these drugs for this purpose. Many were so confident that the drugs provided important therapeutic benefits that they thought it would not be ethical to withhold these drugs from patients in the control group of a randomized controlled trial (RCT) designed to reliably evaluate their effects on overall mortality. (Similar arguments are made today by advocates for continued widespread use of antibiotics in children with acute otitis media, even though we lack scientific evidence to establish that antibiotics meaningfully decrease complications or reduce the time to resolution of symptoms.)

Fortunately, a controlled trial of encainide and flecainide was conducted. The Cardiac Arrhythmia Suppression Trial provided results that astounded cardiologists.[3] These two anti-arrhythmia agents, while suppressing arrhythmias effectively, not only did not provide an improvement in survival, but actually tripled the death rate. Encainide and flecainide may have provided some benefit though suppression of arrhythmias, yet they also had unintended and previously unrecog-

nized mechanisms that ultimately led to an adverse effect on overall survival, mechanisms that would not have been detected if there had not been a trial to directly assess the effects on the clinical-efficacy endpoint of overall survival.

Many other examples can be provided of the situation shown in Exhibit 2.[4] Consider one additional illustration: antimicrobials and their frequent use to provide decolonization (the biomarker) in the hope of being able to achieve the tangible benefit of reducing the risk of serious symptomatic infections. While decolonization might be achieved in one body system, such as the nasal tract, such effects might not be durable, or decolonization might not have been achieved in other pathways (such as skin or gastrointestinal tract) that also could lead to infection. For example, Trish Perl and colleagues performed a trial in 4,000 surgical patients with *Staphylococcus aureus* present in their nasal passages.[5] Such patients are known to have risk for the postsurgical clinical event of *S. aureus* infection at surgical sites. While the experimental intervention, intranasal mupirocin, achieved a highly statistically significant reduction in the postoperative nasal carriage of *S. aureus* (4.6 percent on treatment versus 21.3 percent on placebo, $p < .001$), there was no effect on the clinical efficacy endpoint of *S. aureus* infection at the surgical sites (2.3 percent from treatment versus 2.4 percent from placebo). In a second example, Richard Chaisson and colleagues reported results from a randomized trial in AIDS patients with bacteremic *Mycobacterium avium* complex disease (MAI bacteremia), showing that increasing doses of clarithromycin produced an 80 percent reduction in bacterial load relative to the lowest dose, yet twelve-week mortality sharply increased with increasing doses (5.7 percent at a twice-daily dose of 500 mg, 25.5 percent at 1,000 mg, and 28.0 percent at 2,000 mg).[6] Hence, with increasing doses, as greater reductions in bacteremia were achieved, these were associated with increases in mortality.

Validation Of Surrogate Endpoints

How can one "validate" a proposed surrogate endpoint, specifically establishing that the effect of the intervention on the clinical-efficacy endpoint is reliably predicted by the effect of the intervention on the surrogate? Ross Prentice identified two conditions that, if simultaneously valid, would be sufficient: (1) The biological marker must be correlated with the clinical endpoint; and (2) the marker must fully capture the net effect of the intervention on the clinical-efficacy endpoint.[7] Although many have had the misunderstanding that the first condition would be adequate to validate a surrogate, the second required condition is less likely to be satisfied and is much more difficult to verify.

Validation of a surrogate should be based on both in-depth clinical insights and empirical evidence. Ideally, one should have a comprehensive understanding of the causal pathways of the disease process and of the intervention's unintended and intended mechanisms of action. Admittedly, achieving such understanding is an extremely complicated challenge. Hence, as recognized by several researchers,

validation of a potential surrogate endpoint typically also requires a meta-analysis of many RCTs.[8] As a result, it is easier to directly show the effect of an intervention on the clinical-efficacy endpoint than to actually validate the surrogate.

An illustration of an informative approach obtaining statistical evidence about the validity of a surrogate is provided by Daniel Sargent, in the setting of stage 3 colon cancer.[9] In this clinical setting, even though a surgeon has removed all clinically apparent disease, it is known that approximately 50 percent of patients will die within five years after surgery because of the recurrence of undetected microscopic residual disease. One is interested in using chemotherapeutic regimens after surgery to eliminate this residual disease and, ultimately, to improve patient survival. The most commonly applied regimens proven to have substantial effects on reducing the risk of disease recurrence and improving survival are those involving the anticancer drug 5-fluorouracil (5-FU) in combination with other agents that enhance its effect.

Sargent performed a meta-analysis of fifteen RCTs, a dozen that compared a 5-FU–based experimental treatment regimen with a control and three that compared two such regimens with a control. In each of these trials, it was possible to compute the effect of each experimental treatment on the surrogate endpoint, "disease-free survival," which was specifically defined to be the time to detection of recurrence of disease (or to death if that occurred first) during the first three years after surgery. It was also possible to compute the effect of the experimental treatment in reducing the risk of death, which specifically was defined to be the time to death during the first five years after surgery. In Exhibit 3, for each of the eighteen comparisons of a 5-FU–based experimental regimen versus control, Sargent plotted the relative risk reduction (called the hazard ratio) for the surrogate—disease-free survival—versus that for the true clinical-efficacy endpoint—overall survival. This clear relationship provides evidence that for regimens having similar mechanisms to those of the 5-FU–based interventions, effects on disease-free survival would provide reliable predictions for effects on overall survival.

Once a surrogate is "validated" for one pharmacologic class of treatment regimens, it is tempting to consider that it can be validly used as a replacement endpoint when evaluating other classes of agents as well. However, as seen in Exhibit 2, one must be able to conclude that the "alternative beneficial effects" and "unintended negative effects" on the clinical-efficacy outcome that are not directly captured by the surrogate endpoint will yield the same net effect for the other classes of agents as for the class of agents used in the validation analyses. For example, when considering short-course antiretroviral regimens that sharply reduce viral load in HIV-infected pregnant women, reduction in the rate of mother-to-child transmission of HIV likely can be shown to be a valid surrogate endpoint for the reduction in the risk of AIDS-defining events and death. However, if formula feeding is used as an alternative to breastfeeding to reduce mother-to-child transmission of HIV, the actual improvement in infant morbidity and mortality could be

EXHIBIT 3
Hazard Ratios For Disease-Free Survival (DFS) Versus Overall Survival (OS), Studies Of Adjuvant Colon Cancer, November 2003
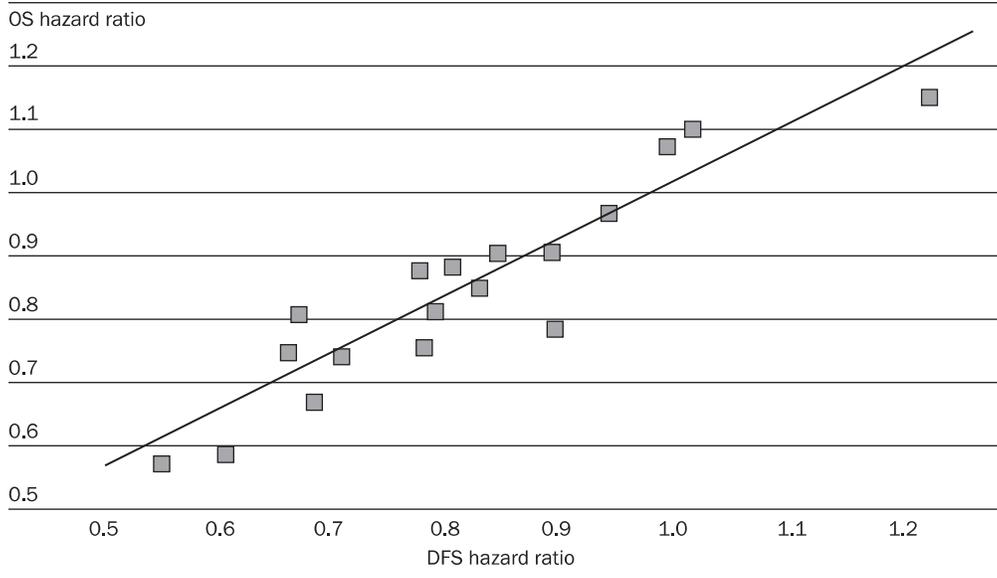
SOURCE: U.S. Food and Drug Administration, data from randomized trials, 20 November 2003.
NOTE: Three-year DFS versus five-year OS as an endpoint for adjuvant colon cancer studies.

greatly overestimated in developing countries, where impurities in the water supply could result in unintended risks of serious bacterial infection and death. For this reason, Ruth Nduati and colleagues performed an RCT of formula versus breastfeeding strategies to reliably assess relative intervention effects on morbidity and mortality as well as on HIV transmission.[10]

In summary, it is important to avoid overstating the conclusions that can be justified from the validation process. Suppose one has performed a meta-analysis of clinical trials evaluating a class of therapeutic regimens in a given disease setting. Suppose further that this meta-analysis reveals that treatment-induced effects on a proposed surrogate endpoint reliably predict treatment-induced effects on the true clinical endpoint. It would be an overstatement to conclude that this process has validated the surrogate endpoint; rather, this process has validated the surrogate endpoint in the given disease setting, for the class of agents studied in those clinical trials.

An Endpoint Hierarchy For Outcome Measures

Based on the insights from the previous discussion, one might form the following hierarchy for outcome measures. Level 1: a true clinical-efficacy measure; Level 2: a validated surrogate endpoint (for a specific disease setting and class of interventions); Level 3: a nonvalidated surrogate endpoint, yet one established to be "reasonably likely to predict clinical benefit" (for a specific disease setting and

class of interventions); and Level 4: a correlate that is a measure of biological activity but that has not been established to be at a higher level.

The highest level, Level 1, would include those outcomes that directly reflect tangible benefit to patients. For example, in cardiovascular disease, reducing the risk of stroke or myocardial infarction could be surrogates for reducing the risk of death (often the supreme Level 1 measure for life-threatening diseases), but at the same time they are also direct measures of clinical benefit through improved quality of life.

An outcome is at Level 2 if it is a surrogate that has been validated in the manner described previously. An example is prevention of mother-to-child transmission of HIV when using short-course antiretroviral regimens. This outcome, while not directly representing tangible clinical benefit, can be used to reliably predict the levels of such benefit. A second example, based on insights from hundreds of clinical trials in patients with cardiovascular disease, is blood pressure reduction as a surrogate for risk of stroke, for well-studied classes of antihypertensive agents such as beta-blockers and low-dose diuretics that are known to have favorable safety profiles. Unfortunately, validated surrogates are rare.

In Level 3 are those outcomes that satisfy the constraint that meaningful effects on these measures are "reasonably likely to predict clinical benefit." Evidence to conclude that an outcome is at Level 3 typically is based on an aggregation of statistical reasoning and clinical insight. To be established to be at least at Level 3, an outcome measure (and the intervention, or the pharmacologic class of interventions such as beta-blockers, to be evaluated using this measure) should satisfy the following criteria: (1) There is considerable clinical evidence that the intervention's effect on the outcome measure will accurately represent the intervention's effect on what is thought to be the predominant mechanism(s) through which the disease process induces risk of clinically tangible events; (2) there is considerable clinical evidence that the experimental intervention does not have important adverse effects on the clinical-efficacy endpoints that would not be captured by the outcome measure; (3) statistical analyses suggest that the net effect of the intervention (or of any member of the class of interventions) on the true clinical-efficacy measure is consistent with what would be predicted by the level of effect on the outcome measure; and (4) the targeted effect on the outcome measure is sufficiently strong and durable that, based on the relationships specified by criteria 1–3, this is reasonably likely to predict meaningful clinical benefit on clinical-efficacy measures.[11] Illustrations of such effects on outcome measures would be reduction of viral load to undetectable levels for six months duration in patients with advanced HIV infection, or a three-month delay in the endpoint of progression-free survival (that is, time to occurrence of death or progression of cancer disease) in advanced cancer patients who have an expected survival duration of six to eight months. In the third of these criteria, the "net effect" of an intervention on the true clinical-efficacy measure refers to its overall effect that is achieved by

the aggregation of its intended and unintended mechanisms of action.

Unfortunately, the majority of outcomes that are found to be correlated with clinical-efficacy endpoints likely lie in Level 4, where treatment effects on such measures establish that an intervention is biologically active, yet where these effects have not been established to provide reasonably likely evidence of clinical benefit. Biological markers, such as CEA and PSA as discussed earlier, that almost certainly do not represent the biological mechanism through which the disease process induces risk of clinically tangible events, likely would be at Level 4. Other measures may be at Level 4 because of the inadequacy of current evidence to establish their validity at a higher level.

Controversial Issues Regarding Accelerated Approvals

In 1992 the FDA formulated a new regulatory process, often referred to as "accelerated approval" (AA).[12] Under the AA process, marketing approval can be provided for interventions when they have been shown to have compelling effects on Level 3 biological markers, where these effects are "reasonably likely to predict clinical benefit." Once AA has been granted to an intervention, the sponsor then is responsible to complete, in a timely manner, one or more clinical trials that will validate that the intervention truly does provide meaningful benefit on tangible measures of clinical benefit. These validation trials should meet all of the usual criteria for quality of trial conduct and reliability of conclusions, including usual levels for statistical strength of evidence, which would be required for providing full regulatory approval in non-AA settings.[13]

The motivation behind AA is to provide patients earlier access to promising new interventions for diseases that are life-threatening or induce irreversible morbidity, when the inadequacy of existing therapies leaves an important unmet clinical need. Although this motivation is easily justified, the actual implementation of the AA process is controversial.

The FDA Oncology Drugs Advisory Committee (ODAC) held a meeting in March 2003 that illustrated many challenges resulting from the AA process.[14] At that meeting, the FDA presented to ODAC the status of validation trials for eight products that had received AA during 1995–2000—the first five years during which the AA process had been implemented in the oncology setting.

One of the disturbing facts revealed in that meeting was that the average time between the granting of marketing through AA and the completion of ongoing validation trials for these eight products was projected to be ten years. It was apparent that when an AA has been granted and a validation trial needs to be done, there often are major difficulties enrolling patients into clinical trials when the experimental therapy has been made available for use in a nonresearch setting. Furthermore, after receiving authorization to market the product, the sponsor often has a loss of the sense of urgency that in premarketing settings is a powerful driving force for the sponsor to obtain timely evaluation of the benefit-to-risk profile

of the intervention. This loss of sense of urgency was illustrated in the validation trial in T-cell lymphoma for Denileukin diftitox (Ontak); that trial enrolled only eight patients per year during the three years before the ODAC meeting.[15] Such a rate of enrollment is far below what typically would be acceptable to a sponsor in a premarketing trial.

Furthermore, in some other cases presented to ODAC, the initial validation trials that had been completed indicated minimal treatment benefit, yet marketing of the intervention continued. For example, ethyol injection received AA in March 1996 for use in non–small cell lung cancer (NSCLC) patients who receive repeated courses of platinum-based therapy.[16] Enrollment in the validation trial, WR-0053, was completed in June 1999, and results were submitted to the FDA at the trial's completion. The trial confirmed that ethyol provided protection from kidney toxicity, but it was unclear whether this effect would affect the clinical risks of dialysis or renal failure. Furthermore, the results regarding the clinical-efficacy measures in the trial were very unfavorable. Patients receiving ethyol had the same frequency of tumor shrinkage but had 20 percent shorter duration of tumor response than the control patients, who were not receiving ethyol. The ethyol group also had shorter time to worsening or progression of their tumors and had shorter survival duration (8.7 months on ethyol versus 9.9 months in the control group), where these unfavorable survival differences were nearly statistically significant. Given that one goal of achieving reduced nephrotoxicity is to enable improved efficacy through better tolerability of platinum-based therapy, these unfavorable trends in the ethyol injection group call into question the appropriateness of continued marketing and use of this agent in nonresearch settings. Indeed, if it is a high priority to reduce nephrotoxicity, a simple alternative would be to use reduced doses of the platinum-based regimens. Such a strategy does not require ethyol and may well result in less loss of efficacy than is apparent with use of ethyol.

Even after extensive involvement in the process, this author has not been able to identify a clear pathway that the FDA plans to use when the validation trial itself is not conclusively positive. This matters because the AA process as implemented allows products that are biologically active but fail to truly provide tangible benefit to patients, and that potentially induce serious safety risks that cannot be detected reliably in the relatively small and short-term trials that can serve as the basis for achieving AA, to be marketed for an indefinite period of time.

If sponsors are allowed lengthy periods to complete validation trials and if the FDA allows marketing of these products to continue after a validation trial fails to provide conclusive evidence for a favorable benefit-to-risk profile, then AA appears tantamount to receiving full regulatory approval. If so, then why should more lenient criteria be allowed for regulatory approval of agents in the AA setting?

Congress should require and the FDA should ensure that AA be granted only when diseases are life-threatening or induce irreversible morbidity; the new intervention is expected to provide important benefit over existing therapies; the pro-

posed surrogate endpoint upon which AA will be granted satisfies the criteria for a Level 3 outcome measure; clinical trials are in place at the time of the AA that can reasonably be expected to provide statistically compelling evidence, within a well-defined rapid time frame, about whether the intervention has a favorable benefit-to-risk profile by being safe and by providing clinically meaningful tangible benefit to patients; and the product will be withdrawn from the market promptly if the validation trial does not conclusively provide this required positive evidence.

Discussion

There is considerable interest in identifying approaches that will allow more rapid availability of drugs and biologics that are being evaluated in new clinical indications and that appear promising in early stages of clinical experimentation. This is particularly apparent for diseases providing risks of mortality or irreversible serious morbidity, when available interventions provide limited benefit. This has been the motivation for the implementation of the AA process, which allows products to be used in nonresearch clinical care settings before they have been reliably established to have a favorable benefit-to-risk profile.

Some factors motivating interest in a streamlined regulatory process can be in conflict with patients' best interests. Often, sponsors of drugs and biologics view the AA process as the easiest way to get their products onto the market. Not only does the AA process allow sponsors to get marketing approval much sooner and with much less research expenditure, but also, quite frankly, it allows them to market products that likely are biologically active but less likely to provide truly important effects on clinical-efficacy endpoints.

Although the goal of achieving the rapid availability of drugs having improved benefit-to-risk profile is laudable, one should not compromise what is truly in the best interest of public health: the reliable as well as timely evaluation of both the safety and the efficacy of new interventions. Why is it in patients' best interest to have more drugs from which to choose, if there are less-reliable insights to guide their caregivers and themselves in making those choices? And why is it in patients' best interest to have earlier access to biologically active interventions, if these therapies may be inconvenient to receive, costly, and potentially more toxic than effective? And might earlier access to ineffective treatments delay or chill the development and proper testing of other interventions that really do work?

Biological markers can give useful insights into the drug development process. When used as primary endpoints in Phase 2 screening trials or as supportive endpoints in definitive Phase 3 trials, they provide evidence about whether the intervention has effects on the biological pathway through which it is hoped that true clinical benefit will be achieved. Yet to obtain reliable evidence about the benefit-to-risk profile of experimental regimens, the primary approach should be to conduct Phase 3 trials having outcome measures that are Level 1 or Level 2 endpoints in the hierarchy described earlier. In turn, when AA has been granted, con-

tinued marketing of the product should be contingent on the sponsor's providing timely and conclusive evidence from validation trials that establishes that the experimental regimen is safe and provides tangible clinical benefit.

. . . .

NOTES

1. T.R. Fleming and D.L. DeMets, "Surrogate End Points in Clinical Trials: Are We Being Misled?" *Annals of Internal Medicine* 125, no. 7 (1996): 605–613.

2. L.A. Guay et al., "Intrapartum and Neonatal Single-Dose Nevirapine Compared with Zidovudine for Prevention of Mother-to-Child Transmission of HIV-1 in Kampala, Uganda: HIVNET 012 Randomised Trial," *Lancet* 354, no. 9181 (1999): 795–802; and J.B. Jackson et al., "Intrapartum and Neonatal Single-Dose Nevirapine Compared with Zidovudine for Prevention of Mother-to-Child Transmission of HIV-1 in Kampala, Uganda: Eighteen-Month Follow-up of the HIVNET 012 Randomised Trial," *Lancet* 362, no. 9867 (2003): 859–868.

3. D.S. Echt et al., "Mortality and Morbidity in Patients Receiving Encainide, Flecainide, or Placebo: The Cardiac Arrhythmia Suppression Trial," *New England Journal of Medicine* 324, no. 12 (1991): 781–788.

4. Fleming and DeMets, "Surrogate End Points in Clinical Trials"; and R. Temple, "Are Surrogate Markers Adequate to Assess Cardiovascular Disease Drugs?" *Journal of the American Medical Association* 282, no. 8 (1999): 790–795.

5. T.M. Perl et al., "Intranasal Mupirocin to Prevent Postoperative *Staphylococcus aureus* Infections," *New England Journal of Medicine* 346, no. 24 (2003): 1871–1877.

6. R. Chaisson et al., "Clarithromycin Therapy for Bacteremic Mycobacterium Avium Complex Disease: A Randomized, Double-Blind, Dose-Ranging Study in Patients with AIDS," *Annals of Internal Medicine* 121, no. 12 (1994): 905–911.

7. R.L. Prentice, "Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria," *Statistics in Medicine* 8, no. 4 (1989): 431–440.

8. Fleming and DeMets, "Surrogate End Points in Clinical Trials"; L.S. Freedman, B.L. Graubard, and A. Schatzkin, "Statistical Validation of Intermediate Endpoints for Chronic Diseases," *Statistics in Medicine* 11, no. 2 (1992): 167–178; V. De Gruttola et al., "Perspective: Validating Surrogate Markers—Are We Being Naïve?" *Journal of Infectious Diseases* 175, no. 2 (1997): 237–246; D.Y. Lin, T.R. Fleming and V. De Gruttola, "Estimating the Proportion of Treatment Effect Explained by a Surrogate Marker," *Statistics in Medicine* 16, no. 14 (1997): 1515–1527; and M. Buyse et al., "The Validation of Surrogate Endpoints in Meta-Analysis of Randomized Experiments," *Biostatistics* 1, no. 1 (2000): 49–67.

9. D. Sargent, "Disease Free Survival versus Overall Survival as an Endpoint for Adjuvant Colon Cancer Studies: Data from Randomized Trials," Meeting Transcript of the FDA Oncology Drugs Advisory Committee, Washington, D.C., 5 May 2004.

10. R. Nduati et al., "Effect of Breastfeeding and Formula Feeding on Transmission of HIV-1: A Randomized Clinical Trial," *Journal of the American Medical Association* 283, no. 9 (2000): 1167–1174.

11. Regarding point 3, see Freedman et al., "Statistical Validation"; De Gruttola et al., "Validating Surrogate Markers"; and Lin et al., "Estimating the Proportion of Treatment Effect."

12. 21 CFR, Secs. 314.500–314.560

13. T.R. Fleming and B.A. Richardson, "Some Design Issues in Microbicide HIV Prevention Trials," *Journal of Infectious Diseases* 190, no. 4 (2004): 666–674; and Meeting Transcript of the FDA Oncology Drugs Advisory Committee, Washington, D.C., 12–13 March 2003.

14. Meeting Transcript of the FDA Oncology Drugs Advisory Committee (12–13 March 2003); and R. Dagher et al., "Accelerated Approval of Oncology Products: A Decade of Experience," *Journal of the National Cancer Institute* 20, no. 96 (2004): 1500–1509.

15. Ibid.

16. Ibid.