

Surrogate Endpoints: Application in Pediatric Clinical Trials

by

Geert Molenberghs, PhD and Camille Orman, PhD

Key Words:

Adjusted association, Biomarker, Meta-analysis, Pediatrics, Proportion Explained, Relative Effect, Surrogate Endpoint, Validation

Introduction

As part of the Critical Path Initiative¹, the U.S. Food and Drug Administration (FDA) called for targeted research in six areas to stimulate the drug and device pipeline². Highlighted was the development of biomarkers to more rapidly and/or more efficiently determine the benefit risk profile of a new therapy. Potential uses of biomarkers that were listed included genomic tests to identify patients at high risk for serious toxicity, markers of drug metabolism to individualize drug dosage, and new imaging techniques to assess treatment efficacy. In addition, the qualification of new surrogate endpoints, a subset of biomarkers targeted at later-phase clinical trials, was identified as an important area to drive more rapid drug development. As recognized by the FDA, the use of surrogate endpoints holds great promise for improving the efficiency in clinical research. Given the pressing need for new pediatric therapies, incorporating surrogate endpoints could be a significant aid to accelerate development. However, the use of surrogate endpoints has been controversial; with the unique aspects of pediatric research, it is vital that surrogate endpoints be used appropriately, and potentially more frequently, in this population. In this chapter, we will examine the use of surrogate endpoints in clinical research in general and the role that they can play in pediatric research in particular.

Definition of Surrogate Endpoint

Although surrogate endpoints have been present in the scientific debate for over two decades, varying definitions have been used, the earliest one going back to

Prentice³. In 2001, a National Institutes of Health (NIH) working group recommended the following terms and definitions⁴.

Biological Marker (Biomarker): A characteristic that is objectively measured and evaluated as an indicator of normal biologic process, pathogenic process, or pharmacologic responses to a therapeutic intervention.

Clinical Endpoint: A characteristic or variable that reflects how a patient feels or functions, or how long a patient survives.

Surrogate Endpoints: A biomarker intended to substitute for a clinical endpoint. A clinical investigator uses epidemiologic, therapeutic, pathophysiologic, or other scientific evidence to select a surrogate endpoint that is expected to predict clinical benefit, harm, or lack of benefit or harm.

A surrogate endpoint, therefore, does not directly measure clinical impact, but rather reflects the desired therapeutic treatment effect. For example, tumor shrinkage is an obvious candidate surrogate endpoint for long-term survival from carcinoma.

History of Surrogate Endpoints

Assessment of the benefit and risk associated with a therapeutic intervention is the underpinning of a clinical development program. The traditional “gold-standard” in assessing efficacy is the evaluation of treatment effect on a well-defined clinical endpoint. The choice of endpoint is a critical factor in determining

the duration and complexity of the trial. However, often, the most sensitive and relevant clinical endpoint, the so-called “true” endpoint, can pose severe challenges for evaluation. For example, the use of clinical endpoints such as survival in newly diagnosed patients with breast cancer, or short-term mortality in patients following acute myocardial infarction (a relatively infrequent event) would result in large, long, and expensive trials. An effective strategy in these situations is to identify alternative endpoints, or surrogates, that are less costly to be measured, are more conveniently assessed, or occur earlier or more frequently than the true clinical endpoint.

In the 1980s, with the alarming rise in HIV infections and AIDS-related deaths, the approval process of new therapies relying on traditional clinical outcomes was questioned. The scientific and regulatory “gold-standard” for a clinical trial endpoint was one where distinct clinical impact could be shown. In the case of demonstrating a treatment effect on HIV infections, this required assessing the clinical outcomes of either prevention of progression to AIDS or increased overall survival. However, the time and cost associated with trials using these outcome measures were unacceptable given the epidemic of HIV infection. In response to the demand for more rapid approval, the accelerated approval provisions were added to the US new drug regulations⁵ with a guidance following in 1998⁶ (subsequently updated in 2004). This new regulation allowed for accelerated approval of drugs targeting serious or life-threatening diseases if the drug appeared to show a benefit over current therapy. In addition, the approval could

be provisionally based on surrogate endpoints that were reasonably likely to predict clinical benefit (Subpart H). Under this regulation, sponsors must demonstrate long-term clinical benefit on the ultimate outcome measure if the association between the true clinical endpoint and the surrogate endpoint has not been demonstrated. Both Europe and Japan have similar regulatory provisions for accelerated approval based on surrogate endpoints. In 1998, the International Conference on Harmonization (ICH) provided guidance on the use of surrogate endpoints that is also comparable⁷.

Regulatory authorization allowing approval based on surrogate endpoints coupled with the need to reduce the ever-rising costs of developing new therapies led to an increased use of surrogate endpoints as the regulatory basis for approval⁸. Examples include CD4⁺ T-lymphocyte counts (CD4) in HIV-infected subjects rather than progression to AIDS or death, cholesterol levels in lieu of occurrence of myocardial infarction, and tumor response rates instead of reoccurrence.

Controversy With Surrogate Endpoints

In spite of the potential advantages, the use of surrogate endpoints has been controversial due to several dramatic failures of a surrogate endpoint to adequately substitute for a clinical endpoint⁹. During the early 1990s, FDA approved the drugs encainide and flecainide after effective suppression of ventricular arrhythmias was demonstrated. It was believed that since arrhythmia

is associated with a four-fold increase in sudden death following myocardial infarction, these drugs would reduce the death rate after myocardial infarction. Following approval, however, results from the Cardiac Arrhythmia Suppression Trial (CAST)¹⁰, showed that the death rate among subjects treated with encainide and flecainide was more than twice that observed in the placebo subjects. In addition, relying on surrogate endpoints in smaller clinical trials has raised concerns about detecting safety issues that can only be detected in large randomized trials¹¹.

In contrast, reliance on a surrogate endpoint may fail to show the true effect of a new therapy. For example, the evaluation of interferon gamma on recurrent infections in patients with chronic granulomatous disease, a surrogate endpoint for phagocytic function, did not show a therapeutic effect; but clinical benefit was demonstrated by substantial reduction in serious infections^{4,12}.

These failures in the use of surrogate endpoints for evaluating therapeutic effect were fundamentally due to the misconception that an association between a true clinical endpoint and an observed biomarker is sufficient to declare a biomarker as a surrogate. The mere existence of an association between a candidate surrogate endpoint and the true endpoint is not sufficient for using the former as a surrogate: “a correlate does not a surrogate make”⁹. Although an association between a potential and true clinical endpoint is desirable, what is required is that the effect of the treatment on the surrogate endpoint reliably predict the effect on

the true endpoint. Unfortunately, partly owing to the lack of appropriate methodology, this condition was not met in the early attempts to use surrogate endpoints and, consequently, negative opinions about the use of surrogates in the evaluation of treatment efficacy have been voiced^{9,13,14}.

Why Continue with Surrogate Endpoints?

In spite of the failures, there are compelling reasons to continue the evaluation and use of surrogate endpoints. Technological advances have dramatically increased the number of new biomarkers fueling the pool of potential, new surrogate endpoints. Improved understanding of proposed therapies' mechanism of action at the molecular level facilitates the use of relevant biomarkers in the evaluation of benefit and risk¹⁵. Continued public pressure for fast approval of promising new drugs, particularly for serious illnesses where the effect on the true clinical endpoint is distant, encourages the use of surrogate endpoints that could reduce the time and cost of the required trials¹⁶.

Surrogate endpoints also can be used for early detection of safety issues that could point to toxicity problems of a new therapy. The duration and size of clinical trials designed to evaluate efficacy of a new drug are often insufficient to detect rare adverse events or events that occur after prolonged therapy^{17,18}. The use of surrogate endpoints in this context of toxicity-related clinical endpoints might allow one to obtain information about such effects even during the clinical testing phase.

Further, new discoveries in medicine and biology are creating an exciting range of possibilities for the development of many potentially effective treatments for a particular disease. This unquestionably is an achievement, but in turn it creates a challenge to rapidly evaluate a large number of new, promising treatments. Surrogate endpoints in the development program can offer an efficient route.

Finally, shortening the duration of a clinical trial using a surrogate endpoint not only can decrease the cost of the evaluation process, but also can limit potential problems with non-compliance and missing data, thereby increasing research effectiveness and reliability^{19,20}. Benefits to the subject are also obvious in terms of reduced time and the number of potential studies related to the clinical trial burden.

The potential of surrogate endpoints to accelerate and improve the quality of clinical trials is clear. However, early experiences also demonstrate that only thoroughly evaluated surrogate endpoints should be used. The following section discusses this issue further.

Statistical Evaluation of Surrogate Endpoints

Thus, while some of the past failures have led a number of researchers to the conclusion that surrogate endpoints should be avoided altogether, practice has clearly shown that sometimes surrogate endpoints are the only reasonable and

plausible alternative to evaluate a new drug. Nevertheless, past attempts to use surrogate endpoints have made it clear that, before deciding on the use of a candidate surrogate endpoint, it is of the utmost importance to evaluate its validity. Developing a definition of a valid surrogate endpoint and operational criteria to assess a proposed surrogate were needed. Statistical methods to evaluate proposed surrogate endpoints have become the subject of intensive research since the 1980s. Note that, as in most clinical decisions, statistical arguments will play a major role but must be considered in conjunction with clinical and biological evidence.

The first formal statistical framework for the evaluation of potential surrogate endpoints dates back to 1989 when Prentice proposed a formal definition of surrogate endpoints and outlined a set of evaluation criteria, all within a hypothesis testing paradigm³. A perfect surrogate endpoint, as described by Prentice, can be represented as:

$$X \iff S \iff T,$$

where X is the treatment, S is the surrogate endpoint and, and T is the true clinical endpoint. In this paradigm, the surrogate endpoint mediates all of the effect of the treatment on the true clinical endpoint. Prentice proposed four operational criteria to validate a proposed surrogate endpoint:

1. treatment (X) has a significant effect on the surrogate endpoint (S);
2. treatment (X) has a significant impact on the true endpoint (T);
3. the surrogate endpoint (S) has a significant impact on the true endpoint (T);
4. the full effect of treatment (X) upon the true endpoint (t) is captured by the surrogate (S).

Although intuitively appealing, much debate ensued, for the criteria set out by Prentice are not straightforward to verify^{21,22}. The fourth criterion is particularly challenging, as it requires that the surrogate must explain 100% of the treatment effect on the true clinical endpoint. In addition, Prentice's criteria could only be applied to binary endpoints (*e.g.*, success *versus* failure)^{19,23}.

Freedman supplemented Prentice's approach by introducing the term *proportion of treatment explained* (PE), aimed at measuring the proportion of the treatment effect mediated by the surrogate²². This proposal was important as it shifted the interest in the validation of surrogate endpoints from significance testing to estimation of the treatment effect explained by the surrogate. However, properties of the PE made it difficult to reliably estimate^{19,23}, *e.g.*, the denominator of the proportion explained (the effect of treatment on the true clinical endpoint) usually cannot be estimated with precision²⁴. Moreover, and fundamentally, the PE is flawed in the sense that it is not restricted to the unit interval. Attempting to further refine this approach, Buyse and Molenberghs²³ and Molenberghs *et al*²⁵

showed that the PE can be decomposed into three different quantities: the ratio of the surrogate and true endpoint variances, the relative effect, and the adjusted association. This approach reflects the two dimensions of the problem of validating a surrogate endpoint. The first dimension is the capability of the surrogate to predict the treatment effect on the true clinical endpoint, while the second one is the capability to predict the outcome of the true clinical endpoint^{23,26}.

The earlier proposals for a statistical framework to validate a surrogate endpoint have been based on utilizing data from a single trial. However, combining data from multiple studies, or meta-analyses, can lead to a more accurate assessment of a surrogate. Similar to the evaluation in single studies, meta-analyses examine the association between treatment effects on the surrogate endpoint and the true clinical endpoint. Based on the results of the association, the model assesses the reliability for predicting the treatment effect on the true clinical endpoint leading to an observed effect on the proposed surrogate. Daniels and Hughes, and Buyse and colleagues, focusing on continuous response endpoints, employed linear mixed-effects models^{20,27,28} to predict the treatment effects on the true clinical endpoint based on data from the surrogate. In this approach, the quality of a surrogate is quantified using two coefficients of determination: R^2_{trial} and R^2_{indiv} . Both measures are unit-less and can range from 0 to 1. Calculated from earlier trial results, R^2_{trial} measures how precisely the effect of treatment on the true clinical endpoint can be predicted based on the treatment effect on the

surrogate endpoint in a new trial. If $R^2_{\text{trial}} = 1$, then the treatment effect on the true clinical endpoint can be predicted without error based on the treatment effect on the surrogate, whereas, if $R^2_{\text{trial}} = 0$ then the treatment effect of the true clinical endpoint and the proposed surrogate are independent and therefore no meaningful prediction can be made. R^2_{indiv} has a very similar interpretation but it quantifies at the individual patient level how precisely the outcome on the true endpoint can be predicted using the outcome on the surrogate.

This meta-analytic method fully captures both dimensions of validation of a proposed surrogate. Nevertheless, a question that immediately arises in this setting is which of these two dimensions is the most important one in practice. There is no single answer to this question but will depend on the context. For a trialist who wants to use the surrogate to predict the treatment effect on the true endpoint, the trial dimension will clearly be the most interesting one. However, for a treating physician who has observed a tumor response in a specific patient and wants to know how this can predict the survival of the patient, the individual dimension will be most useful.

Many extensions of this meta-analytic approach to surrogate validation have been developed to encompass a range of endpoints including binary, time-to-event, or repeated measures¹⁹. Each of these extensions has led to different ways of quantifying a proposed surrogate which in turn could potentially lead to varying interpretations. Alonso and Molenberghs²⁹ have proposed a unifying

framework for the validation of surrogate endpoints using information theory. This method applies to a wide variety of endpoints and reduces to the quantities previously introduced in the literature providing a unified theoretical basis for the variety of statistical approaches to validate a surrogate endpoint. Recently, a new approach has been introduced based on causal inference concepts^{30,31}. Though this approach is based on the single-trial setting and some strong assumptions are required, it appears to be a promising line of research, especially to evaluate potential surrogates in the initial stages of development when little information about the surrogate is available.

With the advent of new biomarkers and the increase in understanding of disease mechanism, there will be a continuing need for developing new statistical models for testing the validity of new proposed surrogates. To best assist in the evaluation of a biomarker as a potential surrogate endpoint, a statistical framework must be established during all stages of therapeutic development including the exploratory phases of a new compound³².

Surrogate Endpoints in Pediatric Trials

The recognition of the importance of conducting clinical trials in the pediatric population began in 1997 with the Food and Drug Administration Modernization Act (FDAMA) that allowed for market exclusivity based on pediatric clinical trials³³. In 1999, the ICH issued a draft consensus guideline, "Clinical Investigation of Medicinal Products in the Pediatric Population" (ICH-E11), that

encouraged drug development in children while recognizing the unique challenges associated with the pediatric population³⁴. Replacing the long-held belief that using children in clinical trials was unethical was the recognition that only through empirical evaluation in a clinical trial could the risk benefit of a new therapy be accurately assessed³⁵. The Best Pharmaceuticals for Children Act (BPCA), signed in 2002, reauthorized the FDAMA exclusivity provision to further encourage pediatric trials³⁶. In 2003, the Pediatric Research Equity Act (PREA) required a review of all new active ingredients, dosage forms, routes of administration, indications and dosing regimens for assessment of safety and efficacy in a pediatric population³⁷. Through both requirements and incentives, regulatory authorities have attempted to encourage the study of new therapies in pediatrics resulting in an understanding of the benefit risk profiles that is equivalent to that required for adults. For further details regarding the current regulatory framework affecting trials in children and the incentives and requirements, please see chapters by Rose, Maldonado and Nakamura for discussion of EU, US and Japan considerations in other sections of this textbook.

Although ICH-E11 and FDA regulations accept extrapolation of efficacy data from adults to children or from older to younger children, but this is only suitable if the disease process and the outcome of therapy are comparable. However, as noted by FDA, "Children's bodies are not just small versions of adult bodies," so extrapolation is often inappropriate and clinical trials assessing both efficacy and safety are required³⁹. Selecting clinical endpoints appropriate for a chronological

or developmental age are critical in the design of a pediatric clinical trial. For example, specific assessment tools may be necessary to evaluate pain in infants and children³⁹.

Similar arguments for using surrogate endpoints in adult studies can be applied to pediatric clinical trials. The FDA guidance on pediatric oncology studies states that, “in the absence of available therapies to treat refractory stages of pediatric cancers, the FDA expects to use flexible regulatory approaches in approving drugs for pediatric research,” including the use of surrogate endpoints such as the effect on tumor size in place of survival⁴⁰. In addition to reducing the time to evaluation and the number of subjects, surrogate endpoints may be of value in pediatric research where patient-reported outcomes traditionally measured in adult studies are impossible to collect. Parents’ resistance to invasive techniques to measure clinical outcome can drive the need for a surrogate endpoint that is more easily tolerated. Dosing may be better determined by sensitive surrogate measures.

Several statins (e.g., lovastatin, atorvastatin, simvastatin, and pravastatin) have been approved for familial hypercholesterolemia in pediatric patients. Similar to the approval process for these medications in adults, efficacy was based on the surrogate endpoint of lowering LDL cholesterol. Although pediatric patients with this diagnosis are at greater risk for coronary heart disease (CHD), no studies

have examined the long-term safety of statin therapy or decrease in CHD morbidity and mortality with chronic exposure.

Vaccines are difficult to assess based on true clinical outcome due to the long duration of observation and, for some indications, the rarity of infection. Vaccines can be approved using a responder analysis demonstrating an immune response, e.g., seroconversion in those subjects initially seronegative or the maintenance of an increase above pre-vaccination concentrations in subjects who were initially seropositive. The FDA will generally require Phase 4 commitments to study adverse effects and long-term monitoring is required to ensure adequate protection. The debate over the association of autism and childhood immunizations highlights the difficulty in fully assessing the benefit risk profile in pediatrics using a short-term surrogate endpoint. Even if true developmental safety issues exist, they can take years to assess.

As directed in the 2002 guidance for accelerated approval of antiretroviral drugs⁴¹, the five pediatric exclusivity approvals and one PREA approval to treat HIV infections have been based on the surrogate endpoints of HIV-1 RNA levels < 400 copies/ml and increases in CD4 counts. Substantial scientific work has focused on HIV-1 RNA levels and CD4 counts as surrogate endpoints, including in-depth meta-analyses aimed at validating these measures as useful surrogates⁴². Such meta-analyses have focused on adult patients and it is not

known whether these surrogate endpoints will prove reliable predictors for long-term outcome in the pediatric population treated early in life.

As new medications are developed with different proposed mechanisms of action, new surrogate endpoints must be identified and validated in the pediatric population. Gastroesophageal reflux disease in infants can cause marked clinical problems including poor weight gain, failure to thrive, esophagitis, persistent irritability, pain, and feeding problems. Aspiration of refluxate in premature infants can lead to pneumonia which can be life-threatening and lead to chronic respiratory problems. Although a few available medications are approved to treat GERD in the pediatric population, approvals have been based on bridging studies from adult clinical trials using the surrogate endpoint of control of gastric pH levels. However, for other classes of medications that are believed to alter gastrointestinal motility, there is little scientific basis to use gastric pH as an endpoint. Although two recent methods to detect reflux (intraluminal impedance and the ¹³C-substrate breath testing) are promising as surrogate endpoints, neither have been validated⁴³.

Numerous biomarkers have been associated with asthma, including measures of lung function (e.g., peak expiratory flow rate and FEV₁), exhaled nitrous oxide levels, and sputum eosinophils with FDA approvals based on FEV₁⁴⁴. Although inhaled corticosteroids are the most effective current treatment for asthma, there are concerns about potential long-term effect on growth and final adult height⁴⁵.

However, conducting a study of sufficient duration to assess final adult height is logistically difficult. Various surrogate endpoints have been used to assess the impact corticosteroids may have on growth, including knemometry (to measure short-term changes in lower-leg length), growth velocity, and changes in height centile⁴⁵. Confounded with asthma-related height effects, measurement error, and cyclical changes in height have made these short-term surrogate endpoints problematic as predictors of final height.

Although statistical methodology has been applied to qualifying surrogate endpoints in adult indications, little work has been done in the pediatric population. This may reflect the general lack of pediatric clinical trials and availability of data to perform the required analyses, but is clearly an area in need of further investigation to adequately assess the value of proposed surrogate endpoints.

Future Areas for Surrogate Endpoints in Pediatric Research

There are several areas where surrogate endpoints might be particularly beneficial for developing new pediatric therapies. Surrogate measures of inflammation collected by non-invasive techniques such as biomarkers in induced sputum (e.g., total and differential cell counts, cytokines and neutrophil products) may prove useful to assess new cystic fibrosis therapies⁴⁶. Biomarkers of the underlying inflammatory process of Crohn's disease (e.g. C-reactive protein, nitric oxide levels) may prove useful both in identifying

subpopulations of patients as well as surrogate endpoints to evaluate treatment response^{47,48}. Brain-imaging technology, allowing insight into the possible biological basis for psychiatric disease, may provide objective surrogate measures of drug response, as has recently been demonstrated in children with attention-deficit/hyperactivity disorder⁴⁹. Developing surrogate measures of pain in babies in neonate intensive care units is needed to better evaluate the effectiveness and risks associated with analgesia particularly in ventilated preterm infants⁵⁰.

Concluding Remarks

In this chapter, we have described the history of surrogate endpoints, the early misconceptions in assessing a proposed surrogate, and the development of statistical methodologies to accurately evaluate, in a quantitative fashion, surrogate endpoints. Surrogate endpoints in pediatric trials offer potential benefits in reducing the time to evaluation of a new therapy, as well as the number of exposed patients, thereby allowing for less invasive measurements, and collecting developmentally appropriate measures.

With the hoped for increase in pediatric clinical trials, a need for surrogate endpoints will likely increase as well. We must proceed cautiously to ensure that the selection of surrogates is based on sound scientific rationale. Noted failures of surrogate endpoints to adequately assess treatment effect in adults must not be repeated in pediatric clinical trials. The concern over inadequately assessing

the benefit risk ratio with a surrogate endpoint is of even greater concern in the pediatric population. Close collaboration between basic scientists, clinicians, and statisticians can facilitate the appropriate use of surrogate endpoints. Beginning early in the development cycle of a new drug, not only should emphasis be placed on identifying potential biomarkers associated with the proposed mechanism of action, but biomarkers that reflect developmental and chronological differences within the target patient population. Biomedical research will inevitably lead to new biomarkers that must be evaluated, and statistical models must be developed to assess new disease mechanisms and the association of surrogate endpoints. Trial designs should facilitate the evaluation of proposed surrogates, including standardizing collection methods and frequency of biomarker measurements to allow for across-study analyses to be performed. Careful and sound scientific methods will ensure that surrogate endpoints can effectively be use to assess new therapies in children.

References

1. Food and Drug Administration. Challenge and opportunity on the critical path to new medical products 2004. US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
2. Food and Drug Administration. Critical path opportunities list 2006. US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
3. Prentice R. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* 1989; 8:431-440.
4. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* 2001 69:89-95.
5. Food and Drug Administration. Code of Federal Regulations 21 314.50 and 601.40. US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
6. Food and Drug Administration. Guidance for Industry: Fast Track Drug Development Programs —Designation, Development, and Application Review 2004. US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
7. International Conference on Harmonisation. General Considerations for Clinical Trials E8 1997.
8. Ellenberg SS, Hamilton JM. Surrogate endpoints in clinical trials: cancer. *Statistics in Medicine* 1989;8: 405-413.
9. Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are we being misled? *Annals of Internal Medicine* 1996;125: 605-613.
10. Echt BS, Liebson PR, Mitchell LB. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The cardiac arrhythmia suppression trial. *The New England Journal of Medicine* 1991; 324: 781-788.
11. Temple, R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA* 1999;282:790-795.
12. The International Chronic Granulomatous Diseases Cooperative Study Group. A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. 1991. *New England Journal of Medicine*: 324:509-516.
13. De Gruttola VG, Fleming TR, Lin DY and Coombs R. Perspectives: validating surrogate markers – are we being naïve? *Journal of Infectious Diseases* 1997;175:237-246.

14. Fleming TR. Surrogate markers in AIDS and cancer trials. *Statistics in Medicine* 1994;13:1423-1435.
15. Ferentz AE. Integrating pharmacogenomics into drug development. *Pharmacogenomics* 2002;3:453-467.
16. Lesko LJ and Atkinson AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making criteria, validation, strategies. *Annual Review of Pharmacology and Toxicology* 2000; 41:347-366.
17. Dunn N and Mann RD. Prescription-event and other forms of epidemiological monitoring of side-effects in the UK. *Clinical & Experimental Allergy* 1999;29 (suppl):217-239.
18. Jones TC. Call for a new approach to the process of clinical trials and drug registration. *BMJ* 2001;322:920-923.
19. Burzykowski T, Molenberghs G, Buyse M (Eds.). *The Evaluation of Surrogate Endpoints* 2005. New York: Springer.
20. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. 2000. New York: Springer-Verlag.
21. Fleming TR. Surrogate markers in AIDS and cancer trials. *Statistics in Medicine* 1994;13:1423-1435.
22. Freedman, LS Graubard, BI and Schatzkin, A. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* 1992;11:167-178.
23. Buyse M, and Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 1998;54:1014-1029.
24. Molenberghs G, Buyse M, Burzykowski T. The history of surrogate endpoint validation. In: Burzykowski T, Molenberghs G, Buyse M (Eds.), *The Evaluation of Surrogate Endpoints* 2000, New York: Springer: 67-82.
25. Molenberghs G, Buyse M, Geys, H, et al. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Controlled Clinical Trials* 2002;23:607-625.
26. Burzykowski T and Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics* 2006;5:173-186.
27. Buyse, M, Molenberghs G, Murzykowski, T et al. The validation of surrogate endpoints in meta-analysis of randomized experiments. *Biostatistics* 2000;1:49-67.
28. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 1997; 16:1965-1982.

29. Alonso A and Molenberghs G. Surrogate marker evaluation from an information theoretic perspective. *Biometrics* 2007;63:180-186.
30. Frangakis CD and Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58:21-29.
31. Chen H, Geng Z and Jia J. Criteria for surrogate end points. *Journal of the Royal Statistical Society* 2007;69:919-932.
32. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health workshop. *Controlled Clinical Trials* 2001;22:485-502.
33. Food and Drug Modernization Act of 1997, Public Law No. 105-115.
34. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonized Tripartite Guideline Clinical Investigation for medicinal products in the pediatric population. 2000. Brussels, Belgium.
35. Simar MR. Pediatric drug development: the international conference on harmonization focus on clinical investigations in children. *Drug Information Journal* 2000;34:809-819.
36. Best Pharmaceuticals for Children Act 2002, Public Law No. 107-109.
37. Pediatric Research Equity Act 2003, Public Law No. 108-155.
38. Mccune SK, Mathis LL, Cocchetta DM, et al. Safer, better more appropriate: clinical trial design for pediatric drug labels. *Drug Information Journal* 2006;40:185-195.
39. Food and Drug Administration. Critical Path Opportunities Reports. 2006. US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
40. Food and Drug Administration. Draft guidance for industry: pediatric oncology studies in response to a written request 2000. US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
41. Food and Drug Administration. Guidance for industry: antiretroviral drugs using plasma HIV RNA measurements — Clinical Considerations for Accelerated and Traditional Approval. 2002 US Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
42. Hughes, MD. The evaluation of surrogate endpoints in practice: experience in HIV. In: Burzykowski T, Molenberghs G, Buyse M (Eds.), *The Evaluation of Surrogate Endpoints* 2000, New York: Springer: 295-321.
43. Giacoia, GP and Mattison DR. Newborns and drug studies: the NICHD/FDA newborn drug development initiative. *Clinical Therapeutics* 2005;27:796-813.

44. Deykin A. Biomarker-driven care in asthma: are we there? *Journal Allergy and Clinical Immunology*. 2006;118:565-568.
45. Price J, Hindmarsh P, Hughes S, Efthimiou J. Evaluating the effects of asthma therapy on childhood growth: principles of study design. *European Respiratory Journal* 2002;19:1167-1178.
46. Sagel SD. Noninvasive biomarkers of airway inflammation in cystic fibrosis. *Current Opinions Pulmonary Medicine*. 2003;516-521.
47. Ljung T, Axelsson L-G, Heruf M, et al. Early changes in rectal nitric oxide and mucosal inflammatory mediators in Crohn's colitis in response to infliximab treatment. *Alimentary Pharmacology & Therapeutics*. 2007;25:925-932.
48. Vermeire S, van Assche G, and P. Rutgeerts. C-reactive protein as a marker for inflammatory bowel disease. *Inflammatory Bowel Disease* 2004; 10: 661–665.
49. Bush G, Spencer T, Holmes J, et al. Functional magnetic resonance imaging of methylphenidate and placebo in attention-deficit/hyperactivity disorder during the multi-source interference task. *Archives of General Psychiatry* 2008; 65:102-114.
50. Boyle EM, Freer Y, Wong CM, et al. Assessment of persistent pain or distress and adequacy of analgesia in preterm ventilated infants. *Pain* 2006; 124: 87-91.