Surrogate Endpoints: How Comprehensive Do They Need To Be?

Ross Prentice
Fred Hutchinson Cancer Research Center

The stringent criteria (e.g. Prentice, 1989, *Statist in Med* 431-440) that surrogate outcomes need to satisfy if they are to convey useful information about intervention effects on a 'true' endpoint will be reviewed. Essentially these criteria require the surrogate variables to be risk factors for the true endpoint that are comprehensive enough to capture all intervention effects that are pertinent to the true endpoint occurrence. In general, many body organs and systems may by affected, beneficially or adversely, by an intervention, and the histories of an ambitious set of markers may be required to capture pertinent intervention effects. Hence, to establish surrogate outcomes for a true outcome in relation to an intervention or class of interventions, it may typically be necessary to list all plausible effects of the intervention, to determine whether a suitable set of markers of each such effect is available, and to devise an analysis scheme to allow the possibly high dimensional surrogate outcome variables to be compared in a meaningful fashion between intervention groups. Well-developed methods for the analysis of censored multivariate failure time data and flexible methods for accommodating measurement errors in surrogate outcome measurement may be needed for such analyses.

One might hope to 'validate' a (multivariate) surrogate outcome on purely biological or mechanistic grounds. However, it seems unlikely that knowledge of pathways to the true endpoint and of the possible intermediate effects of the interventions being compared would ever be sufficient to do so in a clinical trial setting where one aims to glean knowledge about treatment effects on the true endpoint. Statistical validation, on the other hand, requires both surrogate and true outcome data collection in a single clinical trial or cohort setting, and will tend to require large sample sizes to rule out practically important departures from surrogate outcome criteria.

Some disease preventions and mortality reduction special cases will be used to illustrate the challenges associated with the use of intermediate markers as replacements for true outcomes in clinical trials.

Limitations of Surrogate Endpoints in Cancer Prevention Research

Arthur Schatzkin
National Cancer Institute

Because cancer occurs relatively infrequently, prevention trials with cancer endpoints have to be large, long, and costly. Studies with surrogate cancer endpoints may be smaller, shorter, and cheaper. Given that carcinogenesis is a process at multiple biologic levels (organism, tissue, cell, molecule), a broad range of potential surrogate endpoint markers is available. The key question is whether studies with surrogate endpoints give us the right answer about cancer. Using colorectal cancer as an illustration, I will discuss two potential surrogate endpoints: epithelial cell hyperproliferation and adenoma formation. Studies with colorectal mucosal proliferation as an endpoint have been influential (e.g., the calcium chemoprevention story). Although hyperproliferation has been postulated as a relatively early event in carcinogenesis, whether it is a necessary step on the pathway to cancer is uncertain. Alternative pathways to cancer bypassing (and possibly offsetting) hyperproliferation are plausible, which casts doubt on using hyperproliferation as a surrogate endpoint. The most definitive way to address this uncertainty is to integrate proliferation markers in prevention trials with cancer endpoints, the very studies we were trying to avoid. Less persuasive evidence can be gleaned from observational cohort studies of proliferation vs. cancer. Furthermore, even if we establish that hyperproliferation is a good surrogate endpoint in a prevention trial with treatment A (a chemopreventive agent or dietary modulation), there is no guarantee this marker is a valid surrogate for treatment B. Similar problems are likely to attend the use of such potential surrogates as growth factors, bile acid profiles, or gene expression products. In contrast, adenomatous polyp recurrence in the large bowel is widely regarded as a strong surrogate endpoint for intervention studies. Adenoma formation, a relatively late event in carcinogenesis, appears to be a necessary step in the development of most colorectal cancers (the adenoma-carcinoma sequence). Even for adenomas, however, inferences to cancer are not ironclad: (1) A small proportion of cancers may arise from areas of flat dysplasia not readily observable through the colonoscope. (2) It is theorized that only a small proportion of adenomas (the 'bad' ones) progress to cancer, whereas most (the 'innocent' ones) do not. Preventive interventions may affect the pathways to these heterogeneous lesions differentially, complicating the interpretation of polyp trials. (3) Because recurrent adenomas tend to be small, if an intervention operates primarily in the transition from small to large adenoma, or large adenoma to cancer, then this effect will be largely missed in polyp trials. Cervical intraepithelial neoplasia type 3 (CIN3) may be one of the strongest surrogates for cancer, but this state of severe dysplasia/carcinoma in situ is obviously very close to being invasive cervical cancer. The analogous state for large bowel might be the so-called advanced adenoma (either 1+ cm in size, villous elements, high grade dysplasia). An adenoma recurrence trial with large or advanced adenomas as endpoint, however, would have to be substantially larger and more expensive than the current generation of polyp trials. For surrogate markers in cancer prevention research, study cost and inferential certainty appear to be directly related. There is no free lunch.

Using Surrogates To Select Treatments for Further Study

Richard Simon
National Cancer Institute

The past 50 years produced the methodology of the modern clinical trial. After centuries of enduring a variety of erroneous methods for evaluating treatments, resulting in persistence of ineffective and morbid treatments, a methodology was developed that produced consistently reliable answers that avoided the biases of the medical authorities. The essence of the clinical trial is to ask an important, well-defined question and get a reliable answer. The important question is usually whether administration of a well-defined treatment to a well-defined population of patients produces a better distribution of a medically important outcome than does a placebo or control treatment or no treatment. A key component of the clinical trial is to evaluate the treatment with regard to a medically important endpoint that is relevant to the patient.

The past 50 years has also produced a wealth of information about the genetic and molecular basis of normal and disease processes. The pace of research accomplishments has accelerated in the laboratory, in population studies, and in the genome sequencing center. Increased understanding of the molecular and genetic basis for disease, availability of technologies to simultaneously monitor the levels of expression of tens of thousands of genes, and the development of enormously high through-put assays and diverse combinatorial libraries of compounds is likely to result in the development of new classes of drugs directed at specific molecular targets identified as important in diseases.

The future holds great opportunities and great challenges for therapeutic development and evaluation. Better understanding of the molecular basis for disease will in many cases show that current diagnostic classification systems are not valid, and will lead to stratification of patient populations for therapeutics development and evaluation. This may make it more difficult to obtain large numbers of patients for clinical trials but will offer the possibility that treatment benefits will be more uniform for the more molecularly homogeneous population studied in a clinical trial.

The clinical evaluation of treatments may become a bottleneck for the development of promising compounds. For many diseases in the past, highly promising compounds were rare and clinical trial methodology was better than the drugs available for test. This may now change and new pressure will be placed on clinical trialists to attempt to develop more efficient methods for evaluation of new treatments while retaining the fundamental reliability of clinical trials for offering physicians unbiased and reliable information about whether a treatment provides meaningful medical benefit to patients. My presentation will explore some aspects of the use of intermediate endpoints in meeting these challenges.

Joint Modeling of Longitudinal Measurements and Event Histories

Peter Diggle
Lancaster University

A number of approaches to the joint modeling of longitudinal measurements and event histories have been proposed. We describe an approach based on a bivariate latent Gaussian process whose margins respectively influence the mean of the measurement process and the rate of the event process.

One of the purposes of joint analysis is to assess whether the longitudinal measurement can be used as a surrogate marker for progression towards a single, terminating event. We develop a simple score test for association which is based on separate standard analyses of the measurement and event history components and is therefore easy to implement routinely.

Data from a longitudinal trial of schizophrenia patients, in which the measurements are depression scores and the terminating event is withdrawal due to inadequate response, illustrate the methodology.

Prevention Models for Alcoholism

Bengt Muthen
University of California, Los Angeles

This presentation discusses a statistical method for modeling the relationship between an endpoint outcome and repeated measures of intermediate outcomes. A finite mixture random coefficient growth model is used to assess the influence of growth trajectory class membership on the probability of a binary disease outcome. The method is illustrated by the prediction of alcohol dependence from several latent classes of heavy alcohol use trajectories among young adults. Implications for studying surrogate endpoints in clinical trials are discussed.

Coronary Flow as a Surrogate Endpoint in Phase II Studies of Myocardial Infarction

Keaven Anderson
Centocor

Dose-selection studies of medical therapy to lyse clots associated with myocardial infarctions often use angiographic examinations of coronary flow 90 minutes after initiation of therapy as a measure of effectiveness. A surrogate endpoint such as coronary flow is desirable in dose selection because the ultimate goal of demonstrating a mortality benefit can be reached only through study of a large number of patients. Better flow has been associated with reduced mortality in a variety of studies; however, regimens shown to improve flow have not always subsequently demonstrated improvements in mortality in Phase III trials. Intracranial hemorrhage has occurred in less than 1 percent of patients studied with currently used thrombolytic regimens, but an excess of intracranial hemorrhage or other catastrophic bleeding with a regimen has the potential of offseting any mortality benefits of successfully reestablishing coronary flow. Major bleeding complications may be considered a surrogate endpoint for increased bleeding-associated mortality. An ongoing Phase II program to select doses for a combination regimen of a fibrinolytic agent, an antiplatelet agent, and an antithrombin agent for a Phase III study will be discussed. Dose regimen selection, sample size for dose regimens, sequencing of dose regimens, confirmation of successful dose regimens within the Phase II program, appropriate selection of flow measurements for study, and balancing of coronary flow measurements with bleeding complications have been key challenges.

Equal Risk for Equal Measures? Cardiovascular Disease, Blood Pressure, and Lipids

Barry Davis
University of Texas School of Public Health

Hypertension and hypercholesterolemia have been identified as risk factors for coronary heart disease, stroke, and total mortality. Many clinical trials have been conducted showing that lowering SBP and/or DBP and lowering total cholesterol and/or LDL cholesterol can decrease the incidence of the aforementioned clinical outcomes in the primary or secondary setting. Can these measures serve as surrogate endpoints for these clinical outcomes? I will review the epidemiological, clinical trial, and statistical background for evaluating this issue. Questions to be considered: Are all blood pressure-lowering and lipid-lowering treatments equivalent? Are they equivalent in selected subgroups? Is the relationship between the surrogate and clinical endpoints linear? How much change is needed and for how long in the surrogate measure to affect the clinical outcome? Is a measure's surrogate status treatment dependent?

Evaluation of Surrogate Endpoints in HIV Clinical Trials

Michael Hughes
Harvard School of Public Health

Two meta-analyses have been undertaken to evaluate the value of CD4 cell count and HIV-1 RNA as (1) prognostic markers for the development of AIDS or death and (2) surrogate endpoints for these clinical outcomes in HIV trials. To evaluate surrogacy, a regression analysis conducted in a Bayesian framework was used to model the association of the difference between a pair of treatments in clinical outcome with the difference between the same treatments in change in marker values, across treatment comparisons from a number of clinical trials. Meta-analyses using the so-called proportion of treatment effect explained were also undertaken. The talk will describe the meta-analyses, including the methods used and the results obtained. Based upon these experiences, some of the practical challenges and methodological challenges confronting the evaluation of surrogate endpoints will be described.

Vaccine Correlates of Immunity

Robert Kohberger
Wyeth-Lederle Vaccines

In typical vaccine efficacy trials, subjects are given a small number of doses (from one to four) then followed over time to determine whether the treatment has prevented disease. The mechanism of action of vaccines is to modify the immune system so that when a foreign organism enters the body, the immune system mounts a defense to prevent undesirable effects of the foreign organism. The modification of the immune system is characterized by a variety of measurements. Most often, systemic IgG antibodies in sera are measured. It is of interest to determine how these immunogenicity measurements are related to the chance of acquiring disease, e.g., 'correlates of immunity.'

A simple logistic regression model will be used to demonstrate how immunogenicity, in a population treated with a vaccine, and efficacy can be linked together. This linkage is important for the development of vaccines. When there are changes to the product that was used in the efficacy trial, it is critical that immunogenicity be examined rather than clinical efficacy because of the size and time required for the completion of clinical efficacy trials.

These simple models have many serious defects, which will be discussed in detail. Most significantly they (1) ignore the time-varying changes in the measurement, (2) require some assumptions about exposure status, (3) are difficult to expand to multiple measurements in the typical observational setting, and (4) ignore anamnestic and cellular responses.

It is important to clearly understand the desired use of the relationship between an immunologic measurement and protection. If groups are to be compared on the basis of a measurement taken shortly after the dose series (as is usually the case), then a correlate relationship based on measurements taken at the time of exposure must be adjusted to account for the different times of measurement.

The difficulties in designing and analyzing a trial to determine the correlates of immunity have led some to propose an approximate method that could be called a population-based approach. In this approach, if a vaccine is 80 percent effective, the protective level of the immunologic measurement is defined as that level reached by 80 percent of the population at any given point in time.

The practical difficulties in designing and executing trials to address these questions will be discussed as well as the statistical methods and assumptions required for their analysis. Examples of attempts to characterize correlates of immunity will be discussed starting with Greenwood and Yule (1915) and including varicella vaccine (White et al., *Pediatr Infect Dis J*, 1992;11:19-23) and pertussis vaccines (Cherry et al., *Vaccine*, 1998;16:1901-1906 and Storsaeter et al., *Vaccine*, 1998;16:1907-1916).

Modeling of the Pathogenisis of HIV and HCV and the Selection of Surrogate Endpoints

Alan Perelson
Los Alamos National Laboratory

Marie Davidian
North Carolina State University

The use of simple models of viral infection to gain insights into the processes underlying viral pathogenesis in AIDS and hepatitis C infection will be described. The work done to date relies on perturbing the dynamic equilibrium that exists in most patients between virus production and clearance with antiviral drugs. Analyzing the resulting decline in viral load with the use of dynamic models and parameter estimation techniques has revealed information about the half-lives of both virus and productively infected cells, the daily rate of viral production, and the efficacy of the drug therapy. One important aspect is that studying the dynamic response of patients to therapy over short periods of time can yield information that may be predictive of long-term outcome. These results suggest that pathophysiological modeling, when incorporated into an appropriate statistical framework, may hold greater promise for evaluation of dynamic response as a surrogate endpoint than does more traditional empirical modeling. Approaches to such statistical modeling and the potential challenges and limitations will be discussed.

Diabetes Models Based on the DCCT: Research and Health Policy Implications

Richard Eastman
National Institute of Diabetes and Digestive and Kidney Diseases

A health state transition model of diabetes was developed from the results of the DCCT and for clinical trials and observational studies for outcomes not observed during the trial. The model uses Monte Carlo simulation techniques to progress patients through multiple health states leading to diabetes-related blindness, renal failure, amputation, and cardiovascular disease. Weibull probability distributions were developed from the DCCT that describe the risk of developing complications based on duration of diabetes and HbA_{1C}. The model yields results similar to those of a model developed independently from the epidemiology of type II Diabetes (Eastman et al., *Diabetes Care* 20:725-734, 1997). The DCCT-based model was validated by the ability to predict outcomes observed in patients with type II diabetes in the Wisconsin Epidemiologic Study of Diabetic Retinopathy, the Rochester Diabetes Study, and the United Kingdom Prospective Diabetes Study. The results indicate a universal relationship between glycemia and complications that applies to type I and type II diabetes. With the use of modeling techniques, the results obtained during the DCCT can be used to predict diabetes complications in diverse populations.

Risk of Microvascular Complications of Type I Diabetes Mellitus as a Function of Glycemic Exposure (HbA_{1C}) in the Diabetes Control and Complications Trial

John M. Lachin
George Washington University

The Diabetes Control and Complications Trial (DCCT) was a long-term (1983-1993) randomized multicenter clinical trial that demonstrated that intensive therapy aimed at near-normal levels of glycemia, when compared to conventional therapy aimed at maintenance of clinical well-being, markedly reduced the risks of the early microvascular complications of diabetes: retinopathy, nephropathy, and neuropathy (DCCT Research Group, *NEJM*, 1993). Intensive therapy was associated also with a threefold increase in the risks of severe hypoglycemia. Subsequent epidemiologic analyses (DCCT Research Group, *Diabetes*, 1995, 1996, 1997) employed Poisson, proportional hazards and multiplicative intensity regression models with fixed and time-dependent covariates to assess the relationship between these outcomes and the lifetime history of hyperglycemia as reflected by the glycosylated hemoglobin percentage (HbA_{1C}). These analyses showed that a history of hyperglycemia was the dominant determinant of the risk of complications and that virtually all of the beneficial effects of intensive therapy are attributable to the reductions in glycemia. Conversely, the risk of hypoglycemia was more strongly associated with intensive versus conventional therapy, less so with the levels of glycemia achieved. These analyses establish the HbA_{1C} as a meaningful outcome for therapies aimed at improvements in blood glucose control in type I diabetes with the objective of reducing the long-term risks of complications. The recently reported United Kingdom Prospective Diabetes Study does likewise for type II diabetes.