INTRODUCTION

Over the last several decades, epidemiology has incorporated into its practice a constellation of new methods that measure biologic attributes (1). The impact of these new technologies is evident in epidemiologic studies directed at the full spectrum of human disease. We can now provide information at the molecular level and have the capability of determining the biologically effective dose for some compounds, and can quantify the biologic consequences of exposure at sites of injury. These advances in biologic markers (biomarkers) have facilitated assessment of associations between environmental exposures and disease with greater precision and at more biologically relevant levels than previously (2, 3). For biomarkers linked to incidence of a particular disease, the biomarker itself becomes a useful outcome and it can be studied as a primary response variable, thereby shortening investigations. The focus of research may shift from risk factors for the ultimate disease outcome to determinants of the biomarkers of interest as an intermediate outcome. Prevention interventions can also be targeted to modify the levels of the biomarkers that mark progression of disease, thereby shortening intervention trials (4, 5). Furthermore, measures of biologic exposures provide a new level of resolution that enhances the adjustment of confounding and the assessment of effect-modifying variables.

Figure 1 shows a paradigm for biomarkers proposed by a joint committee of the National Academy of Sciences and the National Research Council (6). The paradigm is demonstrated through the relation of environmental aflatoxin exposure and hepatocellular carcinoma. The first two boxes in figure 1 after "exposure" relate to the use of biomarkers as measures of exposure. By measuring indicators in samples of blood, urine, or other biologic materials, markers can provide measures of not only the exposure itself, but of the internal dose and of the biologically effective dose. These biomarkers of exposure provide more biologically appropriate variables for the assessment of risk of disease. In turn, attributes in biologic specimens collected from individuals could provide signals of disease occurrence or progression. The two boxes in figure 1 to the left of "cancer" relate to the use of biomarkers as measures of disease; namely, markers at the molecular level in specimens from individuals could identify early biologic effects associated with disease and/or they could detect alteration of structure/functions leading to disease onset. Such biomarkers of disease can be used as intermediate outcomes in the sequence from normalcy to disease. Lastly, the "susceptibility" box at the top of figure 1 reflects factors that affect the rate of transition from one stage to the next. Presence of the factors will facilitate (i.e., synergistic factors) or slow (i.e., antagonistic factors) the transition from one stage to the next in the chain of events from exposure to disease.

In this context, data from cohort studies offer a platform for 1) assessing the relations between environmental exposures and biomarkers of internal and biologically effective dose, 2) describing the mean trajectories and within-individual correlation of biomarkers as levels are measured repeatedly, 3) evaluating the relation between level or trajectory of biomarkers and the incidence of disease over time, 4) determining the factors that induce different trajectories of the biomarkers and in studies considering the biomarker as an intermediate outcome, and 5) designing experimental studies (clinical trials) to test the efficacy of interventions on modifying the trajectories of biomarkers and the relevance of that modification to the occurrence of clinical disease (i.e., the degree by which a biomarker is surrogate of disease).

In this presentation we provide a review of the methodological approaches for the use of biomarkers as measures of exposure and as measures of disease...
BIOMARKERS AS MEASURES OF EXPOSURE

The general topic of using exposure biomarkers has previously been reviewed, as have methods for addressing intensity and duration of exposure (1, 2). In this section, we will review new uses of biomarkers as measures of exposure. First, we will link the use of biomarkers as measures of exposure to the mechanism underlying the exposure-effect relation. Second, we will indicate how biomarkers that measure duration of exposure can be useful for completion of prevalent cohorts that attempt to describe the timing of outcomes from the beginning of an exposure. Third, we will discuss how the composition of the biomarker is a dimension of interest after the two fundamental axes of intensity and duration. Fourth, we will illustrate how, in spite of the abundance of biomarkers, there are and will be instances in which more traditional measures of exposure will remain the preferred and only choice for large cohort studies.

In studies where the aim is to determine acute effects of exposures, the use of biomarkers that accurately and precisely measure dose is most crucial. In contrast, studies directed at chronic effects generally focus on cumulative exposures. For some exposures and outcomes, it may be necessary to consider the most recent dose, or maximum dose, or dose at a certain critical time, or some other indicator of exposure. Methods to summarize the duration and intensity of long-term exposure have been extensively studied in occupational epidemiology (14). Investigators conducting longitudinal studies for the investigation of acute effects need to use analytical methods that focus on the quantification of the changes in outcome associated with changes in exposure. If the effects are purely acute (e.g., reduction of viral load resulting in arrest of immune deficiency), then the comparison of an exposed individual with another unexposed indi-
individually does not need the incorporation of the previous history of the exposure of those individuals. In this case, a cross-sectional design or a cohort study with a short follow-up is sufficient, and if repeated outcomes are available, the task of the longitudinal analysis consists of combining the different cross-sections or visits into an overall estimate of the effect of the exposure (15). However, for many biologic processes, the effect of an exposure is not just acute, but reflective of past exposures (e.g., smoking and pulmonary function). An exposure may need to take place for a minimum time for the outcome of interest to occur, and once the exposure is absent, it may take a certain amount of time for the outcome to reverse if the effect is not permanent. Statistical procedures allowing past exposures to have an effect have been proposed and should be explored when analyzing longitudinal data (16). Biomarkers that provide measures of internal dose and/or biologically effective dose provide the basis for the epidemiologist to explore and elucidate how these measures of exposure relate to the outcome of interest.

Biomarkers measured in specimens collected in individuals could provide not only measures of internal dose, but, in the case of markers which have a monotonic change due to the exposure (e.g., increasing viral load in infected individuals), the biomarker can be used as a marker of duration of exposure in cohorts whose members were not observed from the beginning of exposure. For example, in the context of infectious diseases (e.g., hepatitis B), a proportion of participants in a cohort study will test positive for the infection at entry (i.e., prevalent subcohort). In order to establish the time between infection and outcome, it is crucial to have biomarkers that can estimate how long these individuals have been infected.

In order to make use of prevalent cohorts for the estimation of the incubation period of infectious diseases, procedures are necessary to determine the missing information on duration of infection prior to study entry. In the early 1990s, methods to estimate duration of exposure (e.g., viral infection) were developed and applied in large cohort studies, mainly in response to the AIDS epidemic. In a first approach, unknown infection durations were imputed based on the characteristics of the infection curve of the population from which the study was drawn (17–20). A second approach uses data at the individual level to determine how long the individual has been infected prior to entry into the study. The data used in this approach might include the history of behaviors likely to result in infection (21) and the observed times among an incident subcohort (i.e., individuals with known dates of infection) when certain biomarker changes after infection have been observed. Under the assumption that the trajectories of the biomarkers are similar to those observed in the incident subcohort, the duration of infection in members of the prevalent subcohort can be determined using the marker values on enrollment (22, 23). Evidence from the incident and the (completed) prevalent subcohorts is complementary, and by extending follow-up through imputation of the missing durations, information on the incubation period, at times longer than the duration of study, can be attained (24).

For many epidemiologic studies, emphasis is on intensity and duration of exposure, as measured by biomarkers, and the risk for the disease of interest (14). However, for many exposures, biomarkers may have several elements, and the composition of the biomarker itself may have relevance to the disease risk. For example, Donahue et al. (25) investigated the odds of post-transfusion hepatitis C infection in relation to the total number of units transfused and the composition of the transfusions in terms of the proportions of red cells, platelets, and plasma. In a regression analysis, covariates included the number of units transfused on each individual and the percentages of the units in two of the three categories. The authors found that risk was associated with the number of units transfused, not their composition. However, in other areas of epidemiologic research, the variable that matters is the composition, more than the total level. For example, it is the type of fat consumed in a diet, as opposed to the total fat intake, that constitutes an exposure carrying high risk for cardiovascular disease (26).

Major technical advances now facilitate measuring some molecular markers in large cohort studies (27, 28). There are, however, and always will be, parameters for which biomarkers are very difficult to develop. In instances where direct markers are difficult to obtain, the use of more traditional epidemiologic markers may provide an appropriate proxy for the variable of interest. An example of such a parameter is the measure of immune function in individuals infected with HIV. Measures of function, such as cytotoxic lymphocytes, are complex, but the clinical symptoms of thrush and fever are easy to diagnose and may provide a useful and simple indication of immune function for individuals with no signs of immune deficiency. HIV-infected individuals who develop an opportunistic infection in the absence of immune deficiency suggest poor immunologic function (29). The use of clinical parameters has, of course, been widely used in epidemiologic research and is complemented by the use of molecular biomarkers.
Design aspects of cohort studies measuring biomarkers

In order to incorporate all sources of variability, including those resulting from measuring duplicate specimens, reproducibility studies should always be considered as part of data collected prospectively in a cohort study. While longitudinal data collected over time provide the opportunity to quantify the between- and within-individual components of variance, reproducibility studies provide the elements to estimate the variability of a specimen. This component of variance is useful for the proper calibration of regression coefficients summarizing the association between a biomarker and the occurrence of disease (30, 31). Furthermore, these components of variation provide the ingredients to assess the degree of within-individual correlation (or tracking) of biomarkers measured over time. This description of biomarkers is an important complement to understanding the changes in biomarker level over time.

In multicenter studies, it is important to establish standardized procedures for collection, processing, and measurement of biomarkers in samples collected from participants in cohort studies. In addition to adhering to common protocols, cohort studies need to prospectively conduct programs in which the agreement of measurements obtained on the same samples is determined. Furthermore, in cohort studies that are active for relatively long periods of time, technical advances impose changes of methods to measure the same biomarker (e.g., assays with lower limits of detectability). In order to develop conversion formulas among old and new assays, it is important that duplicate samples are measured using the old and the new method so that proper standardization can be accomplished throughout the full length of the study. The simplest approach for these conversion formulas is to regress the values obtained using the new method on the values obtained using the old method. Once the regression is obtained, old values are converted to the expected value under the new method. A more rigorous approach uses the standard methods of multiple imputation that incorporate the uncertainty of the conversion from the old to the new method (32).

Another aspect of the design of cohort studies measuring biomarkers is the need to repeatedly measure a group of individuals on whom the biomarkers are not expected to change. For example, in the context of HIV infection, for which repeated assessments of CD4+ cell count are used to quantify the level of immune deficiency caused by infection with HIV, it is important to repeatedly measure seronegative individuals so that if drifts or unexpected changes at a given visit occur, then proper adjustments of the values observed among seropositive individuals can be made (33). If a standard or expected result is already available prior to the conduct of a cohort study, the biomarkers measured in the study could be checked against the standard.

Analytic methods for biomarkers as measures of exposure

In cohort studies, a primary outcome is the time individuals are disease-free. Ideally, the times are measured from when individuals are first placed at risk for the disease of interest. For example, in diseases caused by an infectious agent, time at risk is naturally measured from the date of infection; in occupational studies, time is to be measured from the date of employment or start of exposure, if different; in pediatric epidemiology, time is to be measured from birth; and in survival studies, time is to be measured from the date of diagnosis of diseases; in studies of the elderly, time might be measured from age 65 years. Often the date of origin is not available. From the clinical perspective, a typical situation is to have a constellation of markers measured at a single time point, and there is need to predict prognosis based on those markers. In this case, follow-up time begins from the point when markers were quantified, which typically represents entry into a cohort study. The now abundant procedures for survival analysis provide the appropriate methods for the analysis of the relation between levels of the biomarkers and the hazard of developing the disease under study (34, 35).

Biomarkers as fixed exposures. If biomarkers are only measured at a single time point, or if they are markers that are invariant over time (e.g., heritable genetic changes), the primary objective is to determine the weights (i.e., regression coefficients) that each marker is to be given to optimize the prediction of the prognosis of individuals. The primary epidemiologic objective is to determine which factors are important (i.e., adjusted by confounders) and to characterize effect modification (i.e., interactions). In epidemiologic studies with a constellation of biomarkers as measures of exposure or as prognostic variables, the investigator faces the decision of how to rank the contribution of different markers that have different scales and are interrelated. A commonly used procedure is to provide Kaplan-Meier estimates and relative hazards via a proportional hazards regression of the quartiles of the study population for each one of the markers of interest. The separation of the Kaplan-Meier curves and the values of the likelihood ratio test relevant to the null hypothesis of no differences between the quartiles can be used to compare the prognostic value of different markers.

Epidemiol Rev Vol. 20, No. 1, 1998
It is possible that other partitions different from the quartiles are more relevant to a particular problem. For example, cut-off values for a biomarker might be preestablished (e.g., systolic blood pressure, ≤120, 121–140, >140 mmHg). The cut-off values of a given biomarker \( (X) \) may not result in groups of individuals with equal sizes but with sizes \( n_1, n_2, \ldots, n_c \) for the \( c \) categories using prespecified values in a biomarker. To contrast the prognostic value of another marker \( (X') \), the cohort should be partitioned into the \( n_1 \) with the lowest values in \( X_1 \), the \( n_2 \) with the next set of values in \( X_1 \), until \( n_c \) with the highest values in \( X_1 \). Comparison of the risk discrimination between the \( c \) categories of same sizes of each of the biomarkers serves to estimate which is a better predictor of outcome (36).

An alternative graphic approach to depict the ability of a biomarker to distinguish those who develop the disease from those who did not is a direct extension of \( q-q \) plots, whereby the percentile values of each marker in the group who develop the disease are compared with the percentile values in the group who remained disease-free (37). Specifically, if the values \( x \) and \( y \) of a marker represent the \( p \)th percentile of the marker in the disease-free and diseased groups, a graph of the natural logarithm of the ratio of the percentile values \( \ln(y/x) \) versus \( x \) shows the difference in the distribution of the marker between the two groups according to the occurrence of disease. No differences in the distribution of a marker between the groups corresponds to a horizontal line at zero; markers that are higher (or lower) in those who developed the disease will have \( \ln(y/x) \) values that are above (or below) zero. The further away the values of \( \ln(y/x) \) are from zero, the stronger the prognostic value of a biomarker.

Figure 2 shows an example of these plots for four biomarkers comparing individuals who developed AIDS to members of a cohort study who remained AIDS-free. The biomarkers were measured early in the study to predict the onset of AIDS over the following 10 years (9). The graph clearly illustrates that viral load yields the greatest discrimination, followed by CD4+ count and neopterin, and that \( \beta_2 \) microglobulin offers the least (although consistent) discrimination.

The advantage of this extension of \( q-q \) plots is that comparison of markers that have different scales can be made because \( \ln(y/x) \) is unitless. The methods presented here are an alternative to the \( z \)-scores obtained by normalizing the biomarkers. Needless to say, the biomarkers need to be positive continuous numbers for the \( \ln(y/x) \) to be defined, but this is the case in the great majority of biologic markers. Actually, many biomarkers are concentrations of a compound and, thus, stochastically well described by extreme distributions (e.g., lognormal, Weibull, gamma). Therefore, formal testing of the relative percentiles depicted in the graphic procedures described above can be accomplished by the use of regression methods for accelerated failure-time distributions (38). A disadvantage of the graphic procedure is that it compares the groups defined by the occurrence of disease at any time, not the survival time, among different biomarkers.

The epidemiologic literature has amply seen the use of proportional hazards regression for the summarization of the effect of multiple biomarkers and exposures on the incidence of disease. Software is widely available for the implementation of analysis including interactions (via covariates as products of biomarkers appropriately centered (39)) of which the particular case of interactions with time (or its logarithmic transformation) allows for nonproportional hazards. An alternative method that canonically incorporates interaction and does not require presetting categories of the biomarkers is that of regression trees (40, 41).

Regression trees partition the population in groups at significantly different risks without the assumptions of generalized linear models, and regression trees are closer to the way biomarkers are used in clinical medicine for the monitoring of patients. Furthermore, they naturally incorporate the hierarchy of the prognostic information in different biomarkers and avoid making extrapolations beyond the scope of data observed in populations. However, the flexibility of regression trees carries the disadvantage of overfitting to the data. It is important to implement sensitivity analysis and to provide measures of uncertainty using resampling methods (e.g., cross-validation and bootstrap) (40). An application of regression trees to summarize the joint effect of HIV load and enumeration of the CD4+ lymphocytes on the times to AIDS provided the basis for the 1997 US Public Health Service guidelines on treating HIV-infected individuals in the United States (9).

Proportional hazard regression for the analyses of cohort studies is geared toward the estimation of relative hazards, that formally is a complicated measure of how the conditional chance of failure is different among groups. These methods have a limited ability to provide inferences about relative percentiles that are a more intuitive measure, since they compare disease-free times directly. Specifically, relative percentiles are the ratios of the times in the exposed group to the times in the unexposed group associated with a specified disease rate. Parametric (e.g., lognormal, Weibull) regression models provide direct measures of relative percentiles as well as estimates and confidence.
FIGURE 2. Natural logarithm of the ratio of the percentile values of the markers in those who developed acquired immunodeficiency syndrome (AIDS) compared with those who remained AIDS-free during follow-up. The data points show the natural log values of the ratios. The distance of the data points from 0 (broken line) indicates the strength of the association of the marker with AIDS development. The values of the markers in the AIDS-free group are shown on the abscissa: "X" marks the median (50th percentile) value of the marker and "O" marks the 25th and 75th percentiles. From Albert et al. (37); reprinted with permission from John Wiley and Sons.

Biomarkers as time-dependent exposures. Since, in most cases, individuals are subject to different exposures at different times, the unit of analysis for epidemiologic inferences in cohort studies is not the
individual, but "individual-periods" constructed while exposures are constant. The changes in exposure recorded by cohort studies offer the opportunity to properly classify all individual-periods that have similar exposures. The ratio of the number of events occurring among the total individual-periods provides a direct estimate of the incidence of disease under a given exposure. Figure 3 depicts an example of an individual having exposure \( E_0 \) from 0 to \( t_1 \), and at \( t_1 \) the exposure changes to \( E_1 \), and remains so until \( t_2 \) when the exposure changes to \( E_2 \), and remains so until \( t_3 \) when the individual developed the disease. The contributions of that individual are as follows: for the incidence of the disease under \( E_0 \), the individual contributes to the denominator from 0 to \( t_1 \); for the incidence of disease under \( E_1 \), the individual enters (late) at \( t_1 \) and remains disease-free until \( t_2 \) (i.e., contributes to the denominator from \( t_1 \) to \( t_2 \)); for the incidence of disease under \( E_2 \), the individual enters (late) at \( t_2 \) and develops disease at \( t_3 \) (i.e., contributes to the denominator from \( t_2 \) to \( t_3 \) and to the numerator as one case).

There are four major cases of time-dependent exposures. The simplest case corresponds to that of an intermediate event (e.g., secondary diagnosis) before the endpoint of interest, which typically is death. In this case, there are four types of observations: 1) the individual has an intermediate event at time \( r \) and dies at time \( r \); 2) the individual has an intermediate event at time \( r \) and is alive and censored at time \( r \); 3) the individual dies at time \( t \) free of an intermediate event; and 4) the individual is alive at time \( t \) and is free of an intermediate event. The time-dependent covariate takes value 0 before the intermediate event and 1 afterwards. For cases 1 and 2, the covariate is 0 up to time \( r \) and 1 from \( r \) to \( r \); and for cases 3 and 4, the covariate remains as 0 for all the time from 0 to \( t \). The hazard of death before an intermediate event is determined by all the periods from 0 to \( r \) of the first two types, and by the full follow-up from 0 to \( t \) of the last two types. The hazard of death after an intermediate event is determined by all periods from \( r \) (i.e., late entries) to \( t \) from the first two types. It is inappropriate to simply compare the Kaplan-Meier curves of the full follow-up times from 0 to \( r \) of the two strata defined by ever intermediate event (i.e., first two types) versus never intermediate event (i.e., last two types).

The second type of time-dependent exposure is that of interventions or treatments in cohort studies. Treatments in observational studies are not administered in a randomized fashion, as is the case in clinical trials (45, 46). Those who receive treatments need to survive long enough to make it likely for them to receive treatments. Univariate comparisons of ever receiving treatment versus never receiving treatment may tend to overestimate the putative beneficial effect of treatments. On the other hand, it is often the case that those who receive treatments are those who have the greatest need due to conditions placing them at high risk of developing disease. Therefore, cohort analyses that use data at the individual level require a tight adjustment of different sources of confounding and the use of multivariate methods that fully utilize the comprehensive data collected. As previously mentioned, the initiation of therapy is often made based on biomarker levels. Thus, the use of biomarkers as controlling for confounding is very important.

The third type of time-dependent exposure is that of markers and symptoms that are measured repeatedly over time, and thus offer the data to determine how the hazard of the event of interest relates to the most recent exposure and/or the history of different exposures over time. The ideal situation for this analysis to be valid is in natural history studies in which markers and symptoms are observed in the absence of interventions, and the investigator aims to describe how the hazard of disease relates to natural changes in markers and symptoms. Importantly, trajectories of biomarkers can be of particular interest when biomarkers are repeatedly measured. Caution needs to be exercised in situations where exposures are changed due to interventions that are triggered by the imminent risk of disease (e.g., renal transplant patients are moved to hemodialysis if the transplant fails; or administration of antiretroviral therapy is triggered by severe immunodeficiency in HIV-infected individuals).

The fourth type of time-dependent exposure corresponds to the evaluation of effectiveness of interventions and therapies at the population level. The impact of therapies that may be expected to occur outside of the rigorous clinical trial setting may best be approximated by studying the effect of the incidence of
of an individual according to exposures that define subperiods, so that the unit of analysis is a record defined by entry time, exit time, status at exit, and the characteristics present during the subperiod from entry to exit. The first three variables describe the outcome on a time scale; the fourth set of variables are the covariates for risk factors for which regression coefficients will quantify the magnitude and significance for the hazard of the event of interest in a cohort study. The risk factors may be quantified by levels of biomarkers measured repeatedly over time. In this analysis, changes in covariates can be incorporated to estimate relative hazards and/or relative percentiles (times), and, more importantly, the analyst controls for survival bias in cohort studies (i.e., the longer the survival, the higher the chance of changing exposure). Software is amply available for the incorporation of late or staggered entries into the analysis of cohort studies (see the appendix of Lamarca et al. (47) for specific codes in several software packages).

BIOMARKERS AS MEASURES OF DISEASE: INTERMEDIATE OUTCOMES

In the previous section we described methodological aspects appropriate for the quantification of exposure by a marker for assessing relations between biomarkers and disease outcomes. In this section, we treat the biomarker itself as the measure of disease outcome, and we are interested in determining the factors that relate to different levels of the biomarker (cross-sectional analysis) and different trajectories of the biomarker (cohort studies measuring the biomarker repeatedly). This type of marker is generally referred to as an intermediate outcome in observational studies, while the term surrogate marker is often used in clinical trials; both refer to a biologic event or measurement that can be assessed or observed after exposure and prior to the clinical appearance of disease and that is associated with the occurrence of disease. In the context of the paradigm depicted in figure 1, the biomarkers quantify early biologic effects and/or altered structure/function prior to the occurrence of disease.

The identification and use of markers to be used as intermediate outcomes is as old as the science of epidemiology. Indeed, the elucidation of the causal pathway between exposure and disease consists of identifying factors that precede the disease outcome. Examples of intermediate endpoints abound in all disease fields. In the natural history of HIV infection, for example, CD4+ lymphocyte counts were identified early in the AIDS epidemic as a continuous marker of disease progression (50). A rapid decline in CD4+ cell counts shortly after infection (11), followed by a
gradual decline over time (12), can be observed years before the occurrence of clinical disease and death. In cardiovascular studies, serum cholesterol and lipoproteins have become intermediate endpoints (51). In respiratory disease, changes in volumes measured after 1 second of forced expiration have been used extensively to elucidate the natural history of chronic obstructive pulmonary disease (52).

The advances in molecular biology have greatly contributed to the identification, characterization, and use of intermediate endpoints. Indeed, many authors and two books have chronicled the rise of 'molecular epidemiology' (1, 2).

Once a biomarker has been identified as a measure of disease progression, cohort studies offer the opportunity to estimate and compare the changes of a biomarker in different subgroups. More importantly, cohort studies provide the data to relate changes in exposure (e.g., cigarette smoking) and changes in outcome (e.g., forced expiratory volume in 1 second). In this issue of Epidemiologic Reviews, Samet and Muñoz have delineated the advances of analytic methods for biomarkers measured repeatedly in a cohort study (53). Methods range from the ability to combine the cross-sectional relations provided by each of the visits in a cohort study to random effects models which, in a unified framework, estimate and compare rates of change of biomarkers in different groups according to a specific exposure (e.g., decline of pulmonary function measured by forced expiratory volume in 1 second and use and amount of cigarette smoking). A thorough discussion of all aspects of using intermediate endpoints would exceed the scope of this paper. Instead, we consider two important issues with intermediate outcomes in the context of cohort studies. First, we describe the role cohort studies play in validating intermediate outcomes. Second, we describe how intermediate outcomes can be creatively used to nest studies for elucidating disease etiology.

Development and validation of intermediate endpoints

Cohort studies provide a setting for identifying intermediate endpoints. For example, in the context of HIV infection, two prospective cohort studies, the Multicenter AIDS Cohort Study (54) and the Women’s Interagency HIV Study (55), provide excellent examples of studies that have enhanced biomarker research. At baseline and at regular semiannual follow-up visits, extensive information is gathered from interview-administered questionnaires and through medical examinations. Blood and other specimens are concurrently analyzed using a battery of tests but, perhaps more importantly, samples are frozen and stored at a central repository for future analyses. After a period of time when disease outcomes are recorded, nested case-control studies can be used to investigate the association of biomarkers and disease outcomes. Newly developed assays can take advantage of the specimens in the repository to measure biomarkers that might have been unmeasurable when the samples were collected.

Obviously, the choice of which markers to investigate requires an intimate familiarity with the disease process. The majority of long-term cohort studies invest substantial resources compiling repositories of specimens that are key for intensive laboratory pathogenesis studies. However, documentation of the quality of those repositories has seldom been reported. If biomedical research will use samples from repositories for important societal issues, such as genetic research and other surrogate markers of disease, measures of quality and validity of the samples are imperative (C.A. Kleeberger, et al. Viability and recovery of peripheral blood mononuclear cells cryopreserved for up to 12 years in a multicenter study. Unpublished manuscript).

Once a candidate biomarker has been identified, it is important that it be validated as a true intermediate endpoint before being used for investigating disease etiology. It has been noted that “...most molecular epidemiologic research has involved validating biologic markers (as opposed to using them for etiologic and intervention research or risk assessment)” (1, p. 79). These types of studies (referred to as transi-tional studies by Hulka et al. (2)) are ideal for conducting within prospective cohort studies when exposure, biomarkers, and disease outcomes have been assessed within the same participant.

Perhaps the most critical assessment of the validity of a biomarker as an intermediate endpoint is the degree to which a biomarker captures an exposure-disease relation. While this idea has been prevalent for a long time, it is perhaps best explained using the statistical formalization of Freedman et al. (56). This approach builds upon the work of Prentice (57), who defines operational criteria for assessing whether a biomarker can be considered as a surrogate endpoint in a clinical trial. In simplest terms, consider a binary disease outcome (D or D̅), a binary exposure variable (E or E̅), and a binary biomarker variable (B or B̅) under evaluation. The probability of disease given exposure can be expressed as $Pr(D|E)$. Similarly, the probability of disease among those who are exposed and who have the biomarker can be expressed as $Pr(D|E, B)$, while those exposed without the biomarker have probability of disease $Pr(D|E, B̅)$. The operational criteria described by Freedman et al. (56) for...
biomarker \( B \) to be an intermediate endpoint of \( D \) for assessing exposure \( E \) can be shown to be equal to:

\[
Pr(D|E,B) = Pr(D|B),
\]

that is, once the biomarker is given, the investigator does not need to know the exposure to determine the probability of disease.

Freedman et al. (56) noted the difficulties in conducting formal statistical tests of this hypothesis. For example, the hypothesis could be rejected because of the use of an inappropriate statistical model. They alternatively suggest estimating the degree to which the exposure is explained by the intermediate endpoint. That is, one can fit two models. The first model predicts the outcome with only the exposure and confounding terms as independent predictors to obtain the coefficient for the exposure \( \beta \). The second model includes both the exposure and the intermediate outcome (and confounders) as independent predictors to obtain the adjusted coefficient for exposure as \( \gamma \). If the exposure is related to the outcome in a pathway independent from the intermediate outcome, then \( \beta \) should approximately equal \( \gamma \). However, if the exposure is related to the outcome completely via the intermediate outcome, then \( \beta \) should be of much greater magnitude than \( \gamma \), which should be near zero. Hence, the ratio \( \gamma/\beta \) estimates the degree that the exposure is explained by the intermediate endpoints, and confidence intervals for this ratio can be constructed.

Freedman et al. (56) noted that the statistical power and confidence interval width for this ratio is an important design consideration. As with any examination of statistical power, either the exposure association needs to be strong or the sample size needs to be large for the confidence intervals to be of moderate length. Freedman et al. suggest that the "...best opportunities for validating [intermediate endpoints] are likely to be through incorporating measurements of [intermediate endpoints] in prospective cohort studies" (56, p. 175).

An obvious practical consideration for measuring intermediate outcomes among large numbers of individuals is the need for laboratory methods that can be used widely. Intermediate outcomes that can be assessed using methods that are very expensive, labor intensive, or that require special resources are less likely to be practically useful. For example, in HIV disease, the use of measures of viral burden (e.g., serum HIV RNA quantification) did not become widely considered for use as an intermediate outcome for AIDS until low-cost, semiautomated kits became available in the mid 1990s. Epidemiologists and laboratory scientists need to continue their dialogue for expanding promising methods to wide populations for proper validation.

The validation of intermediate outcomes in a cohort study as described above requires the observation of the exposure, intermediate outcome, and clinical outcome of interest. The usefulness of the intermediate outcome to the cohort study in which the validation has been completed is limited, however, since the exposure-outcome relation can be directly estimated. The more important use of the intermediate outcome is for future studies, such as clinical trials or prevention trials, where the results can be used as justification for designing a study with the intermediate outcome as the primary study outcome instead of the actual disease outcome. The use of intermediate endpoints in clinical trials has been vigorously debated (see Fleming and DeMets (58) for a review) in light of the few true validation studies.

**Use of intermediate outcomes for etiologic research in cohort studies**

Cohort studies using biomarkers as intermediate outcomes can be viewed as a collection of cross-sectional studies performed in the same group of individuals. Differences between designs of cohort studies can be characterized by three variables: the number of individuals, the number of visits provided by each individual, and the time lag between baseline and last visit. The design of cohort studies requires the determination of appropriate sample size as well as the specification of the frequency of visits and the time lag between them. In cross-sectional studies, sample size calculation is simple and, in most cases, involves comparing the prevalences of disease in exposed and unexposed individuals. In cohort studies using biomarkers as intermediate outcomes, it is important to estimate sample size by incorporating the intrinsic within-individual correlation of the biomarker. The simplest measure of change of a biomarker is the difference of the levels between the first and last visit. Intermediate outcomes can be viewed as a collection of cross-sectional studies performed in the same group of individuals. Differences between designs of cohort studies can be characterized by three variables: the number of individuals, the number of visits provided by each individual, and the time lag between baseline and last visit. The design of cohort studies requires the determination of appropriate sample size as well as the specification of the frequency of visits and the time lag between them. In cross-sectional studies, sample size calculation is simple and, in most cases, involves comparing the prevalences of disease in exposed and unexposed individuals. In cohort studies using biomarkers as intermediate outcomes, it is important to estimate sample size by incorporating the intrinsic within-individual correlation of the biomarker. The simplest measure of change of a biomarker is the difference of the levels between the first and last visit, but this approach does not use the information collected in all visits. The natural extension of the outcome measure is the slope from individual regression of the biomarker on time. Methods have been developed for estimation of sample size using slope or rate of change on biomarkers (59).

In classic investigations of a disease etiology, individuals can often be classified as diseased or disease-free, and exposures can be assessed in terms of their ability to predict the probability of predicting disease or the time of disease onset. Intermediate outcomes, however, can often be measured repeatedly in a prospective cohort study. For example, to describe the progressive immunosuppression resulting from HIV infection, repeated CD4+ lymphocyte counts are being collected over semiannual follow-up visits in the
Multicenter AIDS Cohort Study and in the Women’s Interagency HIV Study. The measurement of repeated biomarker values can lead to creative exposure-biomarker investigations.

There are other uses of biomarkers as intermediate outcomes in cohort studies. Below we describe examples of incorporating repeated biomarkers in nested case-control designs and in the determination of disease etiology.

**Use of biomarkers with nested case-control studies.** The classic nested case-control study within a cohort relies on measurements of individuals with disease for comparison with those free of disease. Nested studies have been widely used because of the efficiency in testing hypotheses relating to the clinical course of disease (60). With intermediate outcomes, cases can be selected using the biomarker instead of the disease. With repeated biomarker measurements, a function of all biomarker measurements (e.g., slope), or some summary measure, can be used to define the cases. These designs are most useful when new data need to be obtained through laboratory testing of stored specimens because of the cost and inefficiency associated with retesting all study samples may not be justifiable. Cases can be matched to controls either at a single (e.g., baseline) visit or can be matched to their longitudinal data. The matched study design contrasts with the typical analysis of longitudinal cohort data that use as much data as possible while “adjusting” out any differences using regression models. While this approach may have more statistical power, it is dependent upon a correct regression model. Thus, the tradeoff in precision (variance) using the matched approach with smaller numbers may be justified to obtain a more valid (unbiased) measure.

An example of this approach for investigating factors related to late HIV-1 disease progression is described by Muñoz et al. (61) and Gange et al. (62). Longitudinal CD4+ lymphocyte counts collected over 7 years were used to characterize individuals as those with either “stable” (cases), “moderate” (controls), or “rapid” (controls) disease progression (61). Cases were matched to controls at baseline by their age, race, center of enrollment, and baseline CD4+ lymphocyte count. Comparisons of cross-sectional variables had demonstrated statistically significant differences among cases and controls in certain variables, such as antibody-dependent cell-mediated cytolysis (63) and tumor necrosis factor-α (64). Furthermore, those with “stable” disease were followed for an additional 6 years and split into those with late disease progression or those with continued stability (62). The result was a study that allowed comparisons of longitudinal patterns of the level of plasma HIV RNA (65) among a group that demonstrated homogeneous stability in CD4+ lymphocyte markers over 7 years.

**Evaluating disease etiology.** Elucidating whether an exposure is associated with an intermediate outcome has been the object of the analytical methods intensively developed in the 1980s and 1990s for the analysis of measurements collected repeatedly from the same individual over time (15, 66). One analysis, of course, is to summarize the longitudinal measurements into a single summary measure, such as the overall mean or the trajectory (slope) of the biomarker over time. Data collected repeatedly on cohort participants are substantially more versatile, however, and compel investigations using the association of repeated observations over time. Analyses can make use of the association for different scientific questions in different ways. One approach is to treat the within-individual correlation of the repeated measurements as a nuisance parameter when the scientific questions are focused only on the exposure-intermediate outcome relation. When the primary objective is to determine whether individuals, who at a given time have similar levels of the intermediate outcome but are subject to different exposures, have different levels of the intermediate outcome in future times, the analysis would be best accomplished by conditional regression models (e.g., Markovian) (16).

Alternatively, the analyst may opt to dedicate a few parameters in the model to extract within-individual correlation information regarding the intermediate outcome. In Gaussian data cases this can be accomplished somewhat more easily (67) than with non-Gaussian intermediate outcomes (68, 69). The purpose of this modeling may be only to obtain elements for predicting future outcomes. However, this modeling might also be undertaken because the correlation of intermediate endpoints is of primary interest. Epidemiologically, the correlation or tracking of variables over time is important for studying the stability versus fluctuations between individuals over time (70). Consequently, the search for variables that affect the tracking of intermediate endpoints may be of primary importance. In Gaussian models, between- and within-variance components can be computed separately for different strata. For categorical intermediate outcomes, where the prevalence is steady over time, binomial models (68) or their extensions (71) provide convenient models in which the tracking of intermediate outcomes has a simple form and is easily modeled. For nonstationary situations in which the prevalence changes over time, more complicated models are needed (69).

Statistical procedures and approaches are now well developed and widely available in several statistical...
software packages. It is imperative that the epidemiologist conceptually understands the differences in the approaches and ensures that the epidemiologic aims guide the analysis of data in cohort studies. The greatest challenge is to provide a biologic foundation for the nature of the effect of an exposure—from very acute to long persistence—on intermediate outcome. In other words, how the history of exposure (not only the current one) affects the current intermediate outcome and its future trajectory or its degree of tracking.

BIOMARKERS AS MEASURES OF SUSCEPTIBILITY

The previous two sections described the use of biomarkers for measuring the degree of exposure or for assessing a disease outcome. A third type of biomarker is that which measures susceptibility (risk) of disease for a given exposure—that is, a marker for a variable that modifies the exposure-disease association. The classic example of susceptibility is due to genetic components of host factors. In HIV research, for example, resistance to infection by HIV-1 in exposed individuals, as determined by their sexual behavior (e.g., anal receptive intercourse) or use of injecting drugs, has been documented to occur with a mutation of the CC chemokine receptor 5 CD4+ receptor (13). Genes themselves can be considered as biomarkers since only their protein products, and not the genetic material itself, can effect cellular changes.

The issues surrounding the use of biomarkers for susceptibility variables in cohort studies are generally similar to the use of biomarkers for measuring exposure. Genetic susceptibility factors have been investigated in cohort studies using family (proband) studies, which require special analytical methods to account for the lack of independence among family members (72). The expression of a susceptibility factor is, in statistical terms, an effect modifier/interaction. That is, a binary biomarker B would be considered a susceptibility factor of exposure E to disease D if the risk ratio Pr(D|E,B)/Pr(D|E) was different from the risk ratio Pr(D|E,B)/Pr(D|E). Testing for interaction effects can be substantially underpowered even for cohort studies with sufficient power to evaluate the unmodified exposure-disease relation Pr(D|E) (73), unless the interaction effect is substantial, such as the difference in the association of aflatoxin B₃ and hepatocellular carcinoma by status with respect to hepatitis B virus surface antigen (74).

EPILOGUE

Changes in molecular biology are occurring with unprecedented speed. The impact of these advances will be to further the use of markers of disease, exposure, and susceptibility. Cohort studies should embrace these changes, since they promise the ability to better understand the events occurring at the cellular and molecular scale. The use of validated intermediate outcomes, in lieu of disease outcomes, could reduce the amount of follow-up needed to identify risk factors, thereby making cohort studies of shorter duration. Nested case-control studies and case-cohort studies will be more commonly designed with intermediate outcomes for assessing biomarkers of exposure. Indeed, even confounding and effect-modifying variables collected over time will be based on biologic markers. Despite these changes in measurement, the principles of good study design, conduct, and analysis remain the paramount goal of epidemiologic research.

ACKNOWLEDGMENTS

This work was supported by grants U01-AI-35043, U01-AI-42590, P03-ES03819, and P01-ES060502, funded by the US National Institutes of Health.

The authors are grateful to the reviewers for suggestions that substantially improved this presentation.

REFERENCES

54. Mels M, Phair JP, Saah AJ, et al. Recent scientific contribu-