ENVIRONMENTAL RESEARCH 48, 129–144 (1989)

REVIEW

A Conceptual Framework for the Validation and Use of Biologic Markers

PAUL A. SCHULTE

Industrywide Studies Branch, Division of Surveillance, Hazard Evaluations and Field Studies, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226-1998

Received July 14, 1988

Biologic markers have been discussed extensively in the scientific literature in the past 5 years. That literature generally has focused on the promise and limitations of markers. Currently, a great amount of effort is under way in government, academia, and the private sector to move the field forward. This effort may be characterized by the inventory and review of potential markers and their use. The next requirement is to add a consideration of research and design strategies for the validation and use of biologic markers, especially as they pertain to the assessment of xenobiotic exposures and resultant health impairments.

This paper delineates a conceptual framework for the validation and use of biologic markers. It expands on the concept of a continuum of events between ambient exposure to a xenobiotic substance and resultant clinical disease. Strategies for research and marker validation are presented. Biologic markers are considered useful in etiologic and mechanistic research, in secondary prevention of disease, in risk assessment, and in assessing the effectiveness of environmental controls.

Biologic markers have been discussed extensively in the scientific literature in the past 5 years (Perera and Weinstein, 1982; Fowle, 1984; Council on Environmental Quality, 1985; Underhill and Radford, 1986; NIH, 1986; NRC, 1987; Schulte, 1987; Perera, 1987; Harris *et al.*, 1987; Hatch and Stein, 1987; Hulka and Wilcosky, 1988). That literature generally has focused on the promise and limitations of markers. Currently, a great amount of effort is under way in government, academia, and the private sector to move the field forward. This effort may be characterized by the inventory and review of potential markers and their use. The next requirement is to add a consideration of research and design strategies for the validation and use of biologic markers, especially as they pertain to the assessment of xenobiotic exposures and resultant health impairments.

Biologic markers may represent signals in a continuum of events between a causal exposure and resultant disease (NRC, 1987). Although biologic markers have been used for decades, current technological advances and developments in basic sciences allow for detection of smaller signals at diverse points in the continuum. These markers are generally biochemical, molecular, genetic, immunologic, or physiologic signals of an event. The current method for estimating risks by relating exposure to clinical disease (morbidity and mortality) can now be supplemented by a fuller method, one that identifies intervening relationships more precisely or with greater detail than in the past. As a result, health events are less likely to be viewed as binary phenomena (presence or absence of disease) but

0013-9351/89 $3.00

rather as a series of changes—through homeostatic adaptation, dysfunction, to disease and death.

The continuum between exposure and disease has been characterized by a number of authors and scientific committees (NRC, 1987; Perera, 1987; Hatch and Stein, 1987) and is shown in Fig. 1. Between exposure (E) in the ambient environment and the development of clinical disease (CD), four generic component classes of biological markers have been identified: (1) internal dose (ID); (2) biologically effective dose (BED); (3) early biologic effects (EBE); and (4) altered structure and function (ASF). Clinical disease also may be represented by biologic markers for the disease as well as markers for prognostic significance (PS). The relationship between all the markers in the continuum is influenced by various factors that reflect susceptibility for occurrence (such as genetic or other host characteristics). The definition of all the marker components has been elaborated elsewhere (NRC, 1987; Hulka and Wilcosky, 1988). In brief: ID is the amount of xenobiotic substance found in a biological medium; BED is the amount of xenobiotic material interacting with critical subcellular, cellular and tissue targets, or with an established surrogate tissue; a marker of early biologic effect (EBE) is an event, correlated with, and possibly predictive of, health impairment; ASF is a prodromal biologic change more closely related to the development of disease. Markers of disease (CD) and of prognostic significance (PS) reflect the presence and future of developed disease, respectively. Markers of susceptibility are indicators of increased (or decreased) risk for any component in the continuum.

RESEARCH STRATEGIES

In this paper a conceptual framework for research strategies for the use of biologic markers in toxicologic and epidemiologic research is presented. The point of departure from current state-of-the-art is the need for studies to validate and characterize relationships between the classes of markers shown in Fig. 1. Figure 2 shows the number of distinct relationships between any two components in the exposure–disease continuum (including markers of prognostic significance). There are at least seven generic components in the continuum: exposure; internal dose; biologically effective dose; early biological effect; altered structure/function; clinical disease; and prognostic significance. Hence, there are a maximum of $n(n - 1)/2 = 21$ nominal relationships that may be studied in any single exposure–disease continuum with $n = 7$ components. This number could, at least, be doubled if the role of susceptibility between two components in the continuum was evaluated.

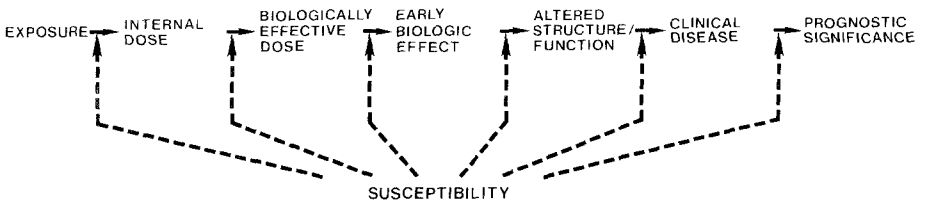

FIG. 1. Biologic marker components in sequential progression between exposure and disease.

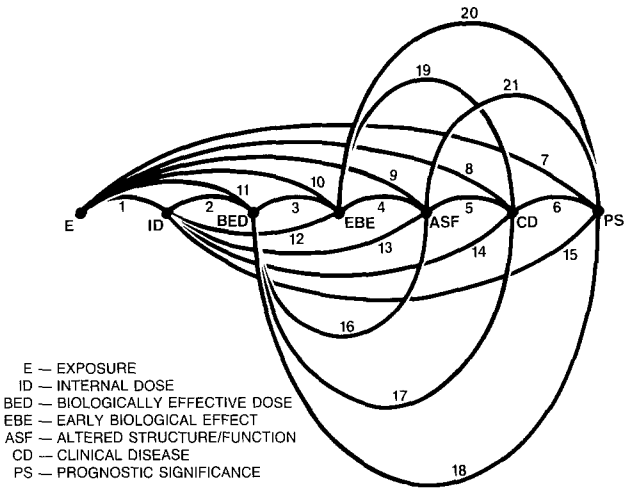

FIG. 2. Schematic representation of research possibilities using biologic markers.

It is important to emphasize that the continuum is a conceptual temporal sequence; the left-most component of the continuum generally precedes any component to the right of it. The direction is arbitrary. It is possible to identify for a given exposure and disease most of the components (or signals that represent them) and to postulate a sequential progression. For example, Table 1 shows the continuum for exposure to various xenobiotic agents. Whether the progression is exactly linear or some other form, such as a multidimensional network, is debatable, but the linear paradigm is useful for research planning purposes. Use of the heuristic linear sequential model should not preclude efforts to explore more complex relationships between markers. For example, multiple markers may be more efficacious than a single marker for characterizing a component of the continuum. In most exposure–disease relationships, the linear causal sequence is an implied framework. Appraisal of the validity of components of the sequence requires that the framework be made explicit and that the existence of causal relationships be tested. The paradigm of a continuum is only meant to elucidate a single pathway among many pathways to a disease from a given exposure. To better model the situation one would consider that there may be multiple pathways leading to a given disease. The contribution of one due to a specific exposure needs to be evaluated in light of contributions from other pathways from other exposures.

Figure 2 shows 21 possible relationships that may be evaluated along the continuum between exposure and disease. The importance of each of these will vary depending on the priorities and objectives of investigators and funding institutions. For example, Table 2 shows how some objectives can be met by studying the associated relationships. These are not the only relationships that can be studied to meet those objectives, but they represent possible initial approaches. Research planners might consider research in light of these possible studies and develop priorities and program plans accordingly.

TABLE 1
EXAMPLES OF CONFIRMED OR HYPOTHETICAL BIOLOGIC MARKERS FOR VARIOUS XENOBIOTIC EXPOSURES[a]

Exposure	Internal dose	Biologically effective dose	Early biological effect	Altered structure/function	Clinical disease[b]	Prognostic significance
Lead	Blood lead levels	Lead level in bone marrow cells	Inhibition of *d*-aminolevulinic acid dehydratase	Accumulation of Zn protoporphyrin	Anemia	Rate of lead decrease on removal from exposure
Ethylene dioxide	Hemoglobin adducts	DNA adducts	HPRT mutation[c]	Sister chromatid exchange	Leukemia	?
Benzidine	Urinary benzidine	DNA adducts	Activated H-ras oncogene	DNA Hyperploidy	Bladder cancer	GAG[d]
Ionizing radiation	Inhaled Radionuclides	HPRT mutation	Chromosomal micronuclei	Hyperplasia	Lung cancer	Tumor antigens
Dioxin	TCDD[e] in blood	Urinary porphyrins	Hyperkeratinization of sebaceous gland	?	Chloracne	?
Fatty food	Serum cholesterol	HDL/LDL[f]	Chylomicrons in blood	Serum enzymes	Myocardial infarction	Serum enzymes
Dibromo-chloropropane	DBCP in blood	?	Mean plasma FSH[g]	Sperm count	Oligospermia	Sperm motility

[a] The order of specific components in each continuum may be speculative and subject to other interpretations.
[b] This component can be represented by markers but also be represented by a constellation of signs and symptoms.
[c] HPRT, hypoxanthinequanine phosphoribosyl transferase.
[d] GAG, glycosaminoglycans.
[e] TCDD, 2,3,7,8-tetrachlordibenzo-*p*-dioxin.
[f] HDL/LDL, high density lipoprotein/low density lipoprotein.
[g] FSH, follicle stimulating hormone.

TABLE 2
POTENTIAL RELATIONSHIPS BETWEEN COMPONENTS IN AN EXPOSURE–DISEASE CONTINUUM THAT CAN BE TESTED TO SUPPORT OBJECTIVES OF INVESTIGATORS AND FUNDING AGENCIES

Objective	Sequential relationships[a] between component markers
Etiology	7, 8, 14, 15, 17
Mechanism	1, 2, 3, 4, 5, 6
Secondary prevention	5, 18, 19, 21
Risk assessment[b]	1, 9, 10, 11, 14
Assessment of environmental controls	1, 2, 3, 4, 9, 10

[a] Numbers refer to pathways in Fig. 2. The numbers have no particular order and are for purposes of identification. This is not an exhaustive list.

[b] Although "risk assessment" requires both exposure and disease data, some of these represent important intervening relationships useful in risk assessment.

For research planning purposes, one way to conceptualize these possibilities is to consider E, ID, or BED in one category as markers of exposure, and ASF, CD, and PS as markers of effect. EBE is an ambiguous marker prior to study and may be used as either a marker of exposure or a marker of effect. Hence, a broad goal would be to initiate research that examines the possible link between a marker of exposure and a marker of effect and, more generally, to determine the entire continuum for a certain exposure.

Biologic markers in the continuum can be treated as analogous to "exposure" or "disease" in classical experimental or epidemiologic research. This analogous treatment is not without its pitfalls. In some cases, it cannot occur without covariate adjustments and a stipulation of an independent determination of each variable.

These adjustments notwithstanding, the conventional techniques for assessing exposure–disease associations, for screening for disease in populations, and for handling multiple variables can be practiced for any two or more components in the continuum. The major assumption that permits this analogous approach is that there is an association between component markers.

Of great importance in relating variables in the continuum is that they be "critical effects." A critical effect is that biologic marker deemed most representative of a particular component in the continuum and ultimately most pathognomonic. It is important, therefore, to attempt to determine which is the critical effect among the range of effects in a component category of candidate markers (Ashby, 1987; Hernberg, 1987). This requires a series of independent studies primarily toxicologic, and then clinical and epidemiologic. It is necessary to develop a hypothesis of the role of the marker in the development of disease. As more causal component associations are identified, it becomes necessary to elucidate quantitative relationships of the kinetics, natural history, and rates of transition along the continuum.

Following determination of critical effects, it is necessary to relate critical effects to estimates of amount (dose) of preceding components in the continuum and the amount of the succeeding ones (that is, for whatever marker is the indepen-

dent variable). The conventional approach to validation is to relate a critical effect to exposure or dose, or to toxic effects. It is suggested here that an additional aspect of validation might include testing the association between a critical effect at any point in the continuum and any other critical effect elsewhere in the continuum. Finally, it is necessary to develop precise, accurate, sensitive, specific, and reliable assays for each component estimate and to determine factors that influence them (Griffith *et al.*, 1988; Schulte, 1987; Ashby, 1987).

RELIABILITY

A biologic marker is a measurement and hence can be considered as having two components frequently referred to as "signal" and "noise" or "true effect" and "random error." Measurement errors need to be acknowledged and controlled. Failure to control these errors can lead to a decreased sensitivity due to the lack of reliability in measurements. Additionally, failure to control the unreliability may lead to the following untoward consequences described by Fleiss (1986): (1) need for increased sample size; (2) systematic bias of correlations in the direction of underestimating them; and (3) biased selection of cases in case–control studies. To control unreliability in measurement, Fleiss (1986) recommends that it be standard practice to conduct a pilot reliability study before embarking on a major research undertaking with measures known to be unreliable. Reliability may also be improved by replicating measurement procedures on each study subject (Fleiss, 1986).

VALIDATION

The validity of a biologic marker can be viewed in terms of the definitions of "measurement validity" used in epidemiology (Last, 1983). Three aspects of validity have been defined: construct validity, content validity, and criterion validity. The construct validity of a biologic marker is its ability to correspond to theoretical constructs under study. For example, if some event such as kidney function would change with age, then a biological marker with construct validity should change as well. A biologic marker would have content validity if it incorporates the domain of the phenomenon under study. For example, a DNA adduct for aromatic amines will integrate exposure from various routes and from occupational and life-style exposures. A biologic marker will have criterion validity according to the extent to which the measurement correlates with an external criterion of the phenomenon under study. The two types of criterion validity that have been distinguished are concurrent validity and predictive validity (Last, 1983). Concurrent validity is when the marker and the criterion refer to the same point in time. For example, ambient air measures of occupational exposure to trichloroethylene could be validated against breath analysis of trichloroethylene. Predictive validity indicates the ability of a marker to predict a criterion. For example, detection of (HLA) B27 can be validated against the appearance of ankylosing spondylitis.

Validation of the relationship between various components of the continuum from exposure to disease involves four levels of effort [as adapted from Gann (1986)]: (1) the determination of an association between a marker and a preceding

exposure or subsequent effect; (2) the location, shape, and slope of the exposure–marker, or marker–effect relationship; (3) the threshold of "no observed effect" level; and (4) the positive predictive value of the marker for exposure or disease.

The validity of biologic markers may be assessed in terms of sensitivity, specificity, disease frequency, and predictive value. The relationship between these parameters and marker frequency is exhibited by the equation: $m = bp + (1 - a)(1 - p)$, where a is the specificity, b is the sensitivity, m is the marker frequency, and p is the disease frequency (Khoury *et al.*, 1985). The positive predictive value (ppv) of a marker is the conditional probability of developing a subsequent component in the continuum given the presence of the marker. The ultimate criterion of a marker is whether it has a strong positive predictive value.

The ppv can be calculated accordingly:

$$\text{ppv} = 1 / \frac{1 + (1 - a)(1 - p)}{bp}$$

(Adapted from Khoury *et al.*, 1985; Arezzo *et al.*, 1983; and Griffith *et al.*, 1988.)

A qualitative rating scale for the validity of biologic markers can be adapted from the work of Busch *et al.* (1988) in reviewing tests for mutagens. Eight levels, increasing in validity, have been delineated as follows: (1) the marker is "totally experimental" with complete uncertainty about health or exposure significance of results; (2) the marker is experimental but theoretical reasons exist to suggest that the marker will correlate with exposure or disease; (3) the marker may be found to correlate with exposure or disease but significance of the data is still uncertain; (4) the marker probably correlates well with exposure or disease but truly conclusive data are not available; (5) the marker has been extensively studied and has been validated as a useful tool for monitoring exposure or diseases, but gives an unexpected positive response in 10% of people screened; (6) the marker has been extensively studied and has been validated as a useful tool for monitoring exposure or disease but gives an unexpected negative response in 10% of people screened who have a history of chronic abnormal exposure; (7) the marker has been extensively studied and has been validated as a useful tool for monitoring exposure or disease, with no or very rare false positives and negatives; and (8) the marker has been validated and is completely predictive of exposure or disease (Busch *et al.*, 1988).

Ultimately, as discussed by Hatch and Stein (1987), an essential requirement for a successful biologic marker of effect is that it should identify from among all exposed individuals those most likely to become diseased. Ideally, we should be able to observe a considerable gradient proceeding from left to right in the continuum between exposure and early biologic change. If the numbers scored positive for left-most markers do not include those positive for ASF or CD, then the marker is of dubious value (Hatch and Stein, 1987).

MULTIPLE MARKERS

In some cases, it is possible that more than one marker will be required to accurately represent a component in the continuum because a single marker will

provide only a weak signal of events for a given component. Multiple markers may offer a better possibility in representing a component. The major issues that arise are (1) how to select the best markers from a group of candidates, and (2) how to combine them into a useful index. The ultimate test for a group of markers is how well they can discriminate between individuals with and without a particular exposure or outcome. The work of Gail *et al.* (1986) on multiple markers for lung cancer diagnosis is illustrative with regard to combining markers in general. They demonstrated that various discrimination rules, such as the Fisher linear discriminant, the logistic discriminant, quadratic discrimination, and recursive partitioning may be useful for attempting to combine multiple markers for any given component in the continuum or for combining markers from multiple components. These discrimination rules do not preclude the use of markers that are correlated, although correlation may have an effect on whether a group of markers can predict an exposure outcome.

SUSCEPTIBILITY

To adequately evaluate the association between two components in the continuum it is necessary to account for differences in or manifestations of susceptibility. Susceptibility can depend on a variety of inherent or acquired host factors ranging from genetic and demographic to behavioral. As depicted in Fig. 1, susceptibility can impinge on the occurrence of each component in the continuum following exposure. The most systematic efforts at evaluating susceptibility have involved genetic factors. Khoury *et al.* (1988) concluded that relative risk of disease for the genetic markers is a function of the frequency of exposure to the environmental agent, the strength of interaction between the genotype, and the specificity of the environmental effect in relation to the genotype. They demonstrated six patterns of genetic and environmental interactions that could affect an association between a genetic marker and disease. These patterns reflect the relative risk of disease for the exposure given that the relative risk for the genetic factor equals one, or is greater or less than one. A relative risk less than one is indicative of a protective factor which, in terms of this discussion, may be considered a lesser degree of susceptibility.

These patterns may apply generically to associations between any two components in the continuum. Hence, the relative risk of the association between dependent and independent variables can be assessed in terms of the interaction between the dependent variable and the predisposing genetic factor. The failure to consider the role of susceptibility in the association between two markers in a causal sequence can lead to erroneous inferences of association.

EPIDEMIOLOGICAL DESIGNS

Longitudinal Cohort Studies

Studies of the relationship between biologic markers will be highly informative when performed longitudinally—that is, according to the temporal sequence, the left-most component in Fig. 1 occurs before any component to the right of it. Longitudinal cohort studies are often expensive and require time to allow for the

period between appearance of the first component and the second component (dependent variable). The approach, however, is generally best for ascertaining the predictive nature of one marker component for another.

In a cohort study, marker components in the continuum can be independent variables for any component marker (dependent variable) to its right in Fig. 1. This is due to the fact that there is an implied causality or positive association between a preceding and any marker following in the sequence. The appearance of any marker can be estimated by the standard hazard rate function $\lambda [t_i Z(t)]$, where t is the failure time (i.e., when the outcome occurs, and Z is the standard covariate history) (Prentice *et al.*, 1986). These can be analyzed by relative risk regression methods. Traditionally, in epidemiologic research, most estimates of regression parameters are based on the assumption of exponential relative risk (Prentice *et al.*, 1986). There appears to be no reason that this would not pertain to evaluations of markers in the exposure–disease continuum, where one marker would be a covariate risk factor for a succeeding marker which would be the outcome or failure.

A pressing need in epidemiologic research in general and with biological markers in particular concerns the impact of covariate measurement errors on relative risk estimation (Prentice *et al.*, 1986). It also has been demonstrated that when laboratory data are collected over an extended period, trends in the data may be misleading. Often, these trends are the result of long-term drifts in values differentially weighted toward people who have been recruited near the beginning of the study rather than those recruited near the end (Thompson, 1983). It is, therefore, necessary to adjust for temporal variation.

Case–Control Studies

The traditional case–control approach can be modified slightly for the purpose of studying the relationship between any two biologic marker components in the continuum. Hence, "case" status, while traditionally based on clinical disease, also may refer to a marker component which will serve as the dependent variable (i.e., the disease surrogate). For example, cases of ASF and controls without ASF may be identified from a population, then both may be evaluated for BED (given the appropriate half-life), or the presence of other exposure markers and for other various risk factors. One may still use the standard risk estimates such as the odds ratio and various statistical tests of association. The interpretation of such research will need to be qualified according to the types of markers and the extent to which they are generalizable to other primary exposure situations and to ultimate disease states. In other words, while an association may exist between intermediary components of the exposure–disease continuum, until the predictive value of those components for exposure and disease is assessed, the results will be germane only to a specific part of the continuum.

The case–control design may be described in terms of incidence density sampling. It involves the selection of a random sample of cases (failures) occurring in a population during some specified case accession period. Corresponding comparison individuals (the controls) are randomly selected from the population without failure during a specified subset of the case accession period (Prentice *et al.*,

1986). Both cases and controls can be defined in terms of presence and absence of biologic markers. Covariate histories pertaining to times up to case or control ascertainment then are obtained retrospectively. The composition of such histories between cases and controls can then be used in the estimation of relative risks and odds ratios. Relative risk regression models are appropriate for evaluating these data and can be used without introduction of a rare disease assumption (Prentice *et al.*, 1986). Similarly, cases and controls can be defined conventionally, and covariate risk factors can be obtained using biological markers of exposure or early effects.

In estimating the sample size for a case–control study involving biologic markers, the data for the independent variable are often continuous. This has received little attention. What attention there has been involved a probit risk model with the assumption of multivariate normal distribution for exposure variables (Lubin *et al.*, 1988). Since probit parameters are not directly interpretable as odds ratios and because exposures may have skewed distributions, these formula have not had wide application (Lubin *et al.*, 1988). An alternative has been developed for case–control designs. The approach uses a score statistic, flexible for any form of exposure data—continuous, ordinal, or dichotomous (Lubin *et al.*, 1988). For hypothesis testing purposes, any component in the exposure–disease continuum can be considered an "exposure" for a succeeding component.

Another important issue to be considered is the use of historical control data in studies involving biologic markers. The following criteria for judging the acceptability of historical control data from observational studies using biologic marker can be synthesized from recommendations by Pocock (1976) for clinical trials and Margolin and Risko (1984) for laboratory experiments: (1) the historical data must have been gathered by same the research team conducting the current study; (2) the study protocol must have remained fixed throughout the period covering the historical studies and the current one; (3) the historical and concurrent control groups must be comparable with regard to demographic factors; and (4) there must be no detectable systematic differences in response between various control groups (Margolin, 1988).

Cross-sectional Studies

Prior to initiating longitudinal or case–control studies, it is often preferable to perform a cross-sectional study. The cross-sectional study is the easiest to perform but often yields the least amount of information. This is because the temporal relationship between the dependent and independent variables cannot be examined. Thus it may not be possible to establish whether one factor is the cause of an effect. However, if there is a mechanistic model developed in animals or for an *in vitro* test system, more can be inferred from a cross-sectional design. Still, such studies are quite valuable either as developing leads for further studies or for determining feasibility for longitudinal or case–control studies.

Cross-sectional studies will yield the most information when participants are selected by random (probability) sampling (Gail *et al.*, 1986). This is not always possible when the focus of the study is in a previously targeted group, such as workers in a particular department or residents living near a landfill. Without

random sampling, however, the cross-sectional study will have serious limitations for making causal inference or generalizations. Cross-sectional studies with biologic markers are best conceived if there is a causal hypothesis for the relationship between markers.

Hybrid Studies

Various hybrid designs for epidemiologic studies have been identified (Kleinbaum *et al.*, 1982). These combine elements of two basic designs, extend a design through repetition, or combine elements of a basic observational design with elements of non-observational design (experiment or quasi-experiment) (Kleinbaum *et al.*, 1982). The same approach as discussed previously (using marker surrogates for exposure or disease) can be used here as well, with the same precautions mentioned. Biologic markers also lend themselves for use in non-observational studies because of their ready accessibility and often shorter time-to-appearance than clinical disease.

An advantageous hybrid design for the use of biologic markers involves case–control sampling within a cohort. One such design has been referred to as a synthetic case–control design, time-matched, or "nested" case–control design. It involves matching each case (failure) in the cohort to a random sample of subjects at risk and without failure at the time of failure in the stratum of the case (Gail *et al.*, 1986; Breslow and Patton, 1979). Another approach, the "case–cohort design" for sampling within a cohort, involves selection of a randomly selected subcohort from the entire cohort, which then serves as a comparison group for all cases arising during cohort follow-up. Covariate histories need to be assembled only for subcohort members and cases. Further development of designs for sampling within a cohort has been recommended and urged as a high priority (Gail *et al.*, 1986).

META-ANALYSIS

Due to the cost and logistical considerations, studies using biologic markers often will be relatively small. In some situations, there may be benefit in a formal evaluation of information from a series of comparable but not truly replicate studies that share a common response variable. This approach, often credited as arising in the social science literature, has been called meta-analysis (Margolin, 1988; Light and Pillemer, 1984). The use of a meta-analytic approach is not recommended to supplant the performance of studies with adequate sample size. This is because the collection of a series of small studies that reinforce each other may produce a false negative result due to the lack of power or a false positive result due to an illusion of comparability which is actually not present.

The method of meta-analysis may be pertinent to studies of biologic markers which often utilize continuous outcomes and show a wide range of variation. One benefit of meta-analysis is the analysis of heterogeneity, which is often more important than computing some fictional common "average" effect (Light and Pillemer, 1984; Greenland, 1987).

A meta-analysis of a number of studies on the same continuum should be

viewed with caution. However, it may provide an overview which will foster the development of "new" hypotheses, which can be tested in subsequent studies.

ANALYTICAL ISSUES

The component markers in the continuum shown in Fig. 1 are not necessarily discrete or the only events in the continuum. There may be a series of other components (steps or stages) between (or in parallel with) these components that have yet to be discovered. In part, the definition of particular markers depends on the current state of knowledge about the relationship between a xenobiotic exposure and a disease. Without knowledge of mechanism, it is not possible, ultimately, to effectively differentiate between a marker of exposure and a marker of effect (Harris *et al.*, 1987; Lucier and Thompson, 1987).

Research designs involving more than two components in a continuum will, by definition, make some of the components intervening variables. At issue is whether the variable is truly an intervening variable or a confounding factor. In a causal progression, it may be a risk factor for disease but it may be correlated also with exposure (since it presumably results from exposure). In this case it should not be considered as a confounding factor, since the effect of exposure is mediated through the effect of altered biochemistry or physiology. Any factor that represents a step in the causal progression between exposure and disease is not a confounding factor; rather, it is an intervening variable (Rothman, 1986). Rothman (1986) has discussed how an investigator can decide if a factor is confounding or not. The answer is seasoned expert judgment of the best available information. When there is uncertainty about the mechanism, handling a potential confounding factor as both confounding and not confounding in different analyses is justified (Rothman, 1986). If the variable is believed to be an intervening variable, then controlling for it in the analysis as a confounding factor may lead to a serious underestimation of effect.

The effective use of various markers as dependent or independent variables in epidemiologic studies must resolve problems of overlap between subjects with and without the marker of interest. The classification of subjects with regard to a marker is analogous to the traditional evaluation of screening tests to detect high risk individuals. Ideally in a disease screening context, a marker would predict without error every individual with disease (Makuch and Muenz, 1987). In a more generic sense regarding the continuum between exposure and disease, a marker would ideally predict every individual with a subsequent marker component. Due to the complexity of human biologic phenomena, the refinement of marker detection and measurement techniques, the absence of knowledge of the transition rates of particular markers, and unreliability in measurements, there can be overlap and misclassification (Makuch and Muenz, 1987). Not all individuals with a particular marker will develop a subsequent marker. In fact, some markers may be a reverse indicator of a subsequent marker or individual disease risk (Ashby, 1987).

Quite often more than one marker will be a candidate to be a component in the continuum. Most of the candidates could be spurious. With some techniques, for example, such as two-dimensional electrophoresis and image analysis, dozens or

even hundreds of gene expression products (proteins) can be candidate markers for single or multiple components in the continuum (Anderson and Anderson, 1984). In other cases, a component will be best defined by several markers rather than an individual one. In both of these instances it is necessary to have a strategy for (1) how to screen a group of several markers and determine which ones should be submitted for formal statistical analysis; (2) the prerequisites to studying several markers simultaneously, including what kinds of statistical models are appropriate and how the markers should be entered into the model; and (3) methods to determine if one model is superior to another in its ability to discriminate between two populations (Makuch and Muenz, 1987). Makuch and Muenz (1987) have used a technique for graphical analysis of tumor markers that may also be applicable generically for the evaluation of multiple biologic markers to adjust for chance findings due to simultaneous inferences, and to enter the markers in regression models to allow for further investigation of the relation between markers and other independent variables.

Analysis of epidemiologic data for studies involving biologic markers can utilize a rich diversity of statistical models. Breslow and Storer (1985) have proposed a family of relative risk functions so that by varying the exponent of power transforms, the effects of individual risk variables are made to combine in a fashion that ranges from subadditive to supramultiplicative. These risk functions are applicable to studies using biologic markers. The choice of models can produce large quantitative differences in risk that can have profound implications for causal interpretation and for public health practice.

The use of biologic markers should allow for a realistic representation of variation in host and dose factors in a heterogenous population. Selection of appropriate statistical models for analysis of biologic marker data should be part of a staged and iterative process that draws upon developments in the understanding of toxicokinetic and pathologic mechanisms from laboratory studies. Further, as larger segments of the continuum are modeled, there will be an increased opportunity for an impact of host and dose factor variation. Modeling of intermediate steps will help control for some of this variation. Biologic marker data from future epidemiologic studies should stimulate the development of improved mathematical models (Alavanja *et al.*, 1987).

Relationships between components in the continuum can be modeled by two broad approaches: empirical and process modeling. The empirical approach can be used when there are no explicit hypotheses about components. The approach is to use statistical techniques to find the combination of descriptors that "best" explain the observed effects (Checkoway, 1986; Smith, 1987). The process modeling approach uses quantitative toxicologic models to estimate concentrations in biological compartments and temporal patterns of occurrence. It requires explicit hypotheses. Process modeling should be the goal as more is learned about the continuum in Fig. 1.

DISCUSSION AND CONCLUSION

The next stage in the development of biologic markers for use in toxicologic and epidemiologic research can be stimulated by a strategy for validating markers and

the development of a framework for their application. The use of the "continuum" paradigm provides the opportunity for wholly modeling the intervening relationships between exposures and disease outcomes. Heretofore, the detailed steps between exposure and disease have been unknown for many diseases, or not understood enough for use in modeling.

It is possible to critique the proposed exposure–disease continuum in several ways. First, the effort to fit a model of biologic markers into a causal pathway for disease assumes that the continuum for exposure and the continuum for disease are parallel and continuous. This is the paradigm, but it does not preclude other possibilities of multiple pathways and multiple markers for each component. Some components will be represented by markers or surrogates that are in the causal pathway between an exposure and disease. Other markers may represent a component in the pathway of interest but are indicative of the causal pathway for another disease (Fig. 3). There is still a great deal to learn about the design impact of using different types of markers in studies. What is the impact of using stochastic and nonstochastic markers, surrogate versus correlative markers, or markers as both independent and dependent variables?

The second issue is that testing causation has been problematic for epidemiologists and depends on views of inductive and deductive logic and models of causation (Susser, 1973; Weed *et al.*, 1988; Rothman, 1986). Historically, observational epidemiology only has been able to infer causality. There is an attempt, in this paper, to consider that between ambient environmental exposure and clinical disease there is a continuum or a progression of factors which have a causal association. Using biologic markers allows for a refined view of the continuum.

In addition to design considerations, there are various other methodologic issues that need to be considered in research involving biologic markers, not the least of which are issues of human subjects protection, interpretation and communication of results, and risk management (Schulte, 1987; Ashby, 1987). These may be limiting in future research (Higginson, 1988).

This paper describes and advocates the use of biologic markers. This should not be taken as a recommendation to dismiss consideration of the social, psychological, cultural, political, or economic factors that interact in the development of

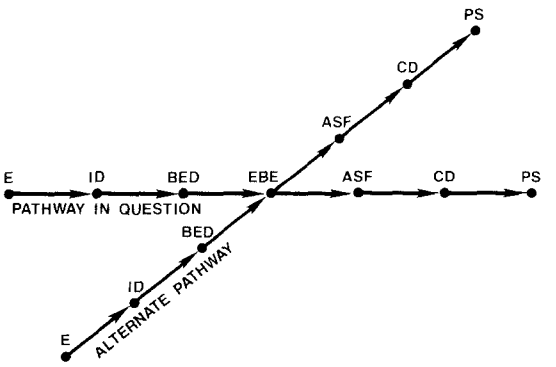

FIG. 3. A common biologic marker in two distinct casual pathways.

disease. Nor is this a prescription for shifting the emphasis of epidemiologic research toward toxicologic poles. Rather, this paper advocates a conceptual framework that can enrich both toxicologic and epidemiologic research.

REFERENCES

Alavanja, M., Aron, J., Brown, C., and Chandler, J. (1987). Cancer risk assessment models: Anticipated contributions from biochemical epidemiology. *J. Natl. Cancer Inst.* **78**, 633–643.

Anderson, L., and Anderson, N. (1984). Some perspectives on two-dimensional protein mapping. *Clin. Chem.* **30**, 1898–1905.

Arezzo, J. C., Schaumberg, H. H., and Petersen, C. A. (1983). Rapid screening for peripheral neuropathy: A field study with Optacon. *Neurology* **33**, 626–629.

Ashby, J. (1987). Monitoring human exposure to genotoxic chemicals. *In* "Occupational and Environmental Chemical Hazards: Cellular and Biochemical Indices for Monitoring Toxicity" (V. Foa, E. A. Emmett, M. Maroni, *et al.*, Eds.), pp. 121–241. Ellis Horwood Limited, Chichester, England.

Breslow, N. E., and Patton, J. (1979). Case–control analysis of cohort studies. *In* "Energy and Health" (N. E. Breslow and A. Whittemore, Eds.), Society for Industrial and Applied Mathematics, Philadelphia.

Breslow, N. E., and Storer, B. E. (1985). General relative risk functions for case–control studies. *Amer. J. Epidemiol.* **122**, 149–162.

Busch, D. B., Easterling, D., Leventhal, H., and Bryan, G. (1988). Recontacting subjects in mutagen-exposure-monitoring studies. II. Results of a questionnaire study of mutagenesis investigators with review of pertinent literature. *Environ. Mutagen.* **11**, 333–358.

Checkoway, H. (1986). Methods of treatment of exposure data in occupational epidemiology. *Med. Lav.* **77**, 148–173.

Council on Environmental Quality, Executive Office of the President (1985). "Report on Long-Term Environmental Research and Development." Washington, DC.

Fleiss, J. (1986). Statistical factors in early detection of health effects. *In* "New and Sensitive Indicators of Health Impacts of Environmental Agents" (D. M. Underhill and E. D. Radford, Eds.), pp. 9–16. University of Pittsburgh.

Fowle, J. R., III. (1984). "Workshop Proceedings: Approaches to Improving the Assessment of Human Genetic Risk—Human Biomonitoring." Report No. EPA-60019-84-016. Office of Health and Environment Assessment, U.S. EPA, Washington, DC.

Gail, M. H., Muenz, L., McIntire, K. K., *et al.* (1986). Multiple markers for lung cancer: Validation of models for advanced lung cancer. *J. Natl. Cancer Inst.* **76**, 805–816.

Gann, P. (1986). Use and misuse of existing data bases in environmental epidemiology: The case of air pollution. *In* "Environmental Epidemiology" (F. C. Topfler and G. F. Craun, Eds.). Lewis, Chelsea.

Greenland, S. (1987). Quantitative methods in the review of epidemiologic literature. *Epidemiol. Rev.* **9**, 1–30.

Griffith, J., Duncan, R. C., Goldsmith, J. R., and Hulka, B. S. (1988). Biochemical and biological markers: Implications for epidemiologic studies. (Unpublished manuscript)

Harris, C. C., Weston, A., Willey, J. C., Trivers, G. E., and Mann, D. L. (1987). Biochemical and molecular epidemiology of human cancer: Indicators of carcinogen exposure, DNA damage and genetic predisposition. *Environ. Health Perspect.* **75**, 109–119.

Hatch, M. C., and Stein, Z. A. (1987). The role of epidemiology in assessing chemical-induced disease. *In* "Mechanisms of Cell Injury: Implications for Human Health" (B. A. Fowler, Ed.), pp. 303–314. Wiley, New York.

Hernberg, S. (1987). Report of Session 1. *In* "Occupational and Environmental Chemical Hazards: Cellular and Biochemical Indices for Monitoring Toxicity" (V. Foa, E. A. Emmett, M. Maroni, *et al.*, Eds.), pp. 121–241. Ellis Horwood Ltd., Chichester, England.

Higginson, J. (1988). Changing concepts in cancer prevention: Limitations for future research in environmental carcinogenesis. *Cancer Res.* **48**, 1381–1389.

Hulka, B. S., and Wilcosky, T. (1988). Biological markers in epidemiologic research. *Arch. Environ. Health* **43**, 83–89.

Khoury, M. J., Adams, M. J., Jr., and Flanders, W. D. (1988). An epidemiologic approach to ecogenetics. *Amer. J. Hum. Genet.* **42**, 89–95.

Khoury, M. J., Newill, C. A., and Chase, G. A. (1985). Epidemiologic evaluation of screening for risk factors: Application to genetic screening. *Amer. J. Public Health* **75**, 1204–1208.

Kleinbaum, D. G., Kupper, L. L., and Morganstern, H. (1982). "Epidemiologic Research: Principles and Quantitative Methods." Wadsworth, Belmont, CA.

Last, J. M. (Ed.). (1983). "A Dictionary of Epidemiology," p. 107. Oxford Univ. Press, New York.

Light, R. J., and Pillemer, D. B. (1984). "Summing Up: The Science of Reviewing Research." Harvard Univ. Press, Cambridge, MA.

Lubin, J. H., Gail, M. H., and Ershow, A. G. (1988). Sample size and power for case–control studies when exposures are continuous. *Stat. Med.* **7**, 363–376.

Lucier, G. W., and Thompson, C. L. (1987). Issues in biochemical applications to risk assessment: When can lymphocytes be used as surrogate markers. *Environ. Health Perspect.* **76**, 187–191.

Makuch, R. W., and Muenz, L. R. (1987). Evaluating the adequacy of tumor markers to discriminate among distinct populations. *Semin. Oncol.* **14**, 89–101.

Margolin, B. H. (1988). Statistical aspects of using biological markers. *Stat. Sci.* **3**, 351–357.

Margolin, B. H., and Risko, K. J. (1984). The use of historical data in laboratory studies. *Proc. Int. Biometric Conf.* **12**, 21–30.

National Institutes of Health (1986). "Third Task Force for Research Planning in Environmental Health Science: Biochemical and cellular markers of chemical exposure and preclinical indicators of exposure. Human Health and the Environmental—Some Research Needs." NIH (86-1277), pp. 207–267.

National Research Council (1987). Biological markers in environmental health research. *Environ. Health Perspect.* **74**, 1–191.

Perera, F. (1987). Molecular epidemiology: A novel approach to the investigation of pollutant-related chronic disease. *In* "Environmental Impacts on Human Health: The Agenda for Long-Term Research and Development" (S. Draggan, J. J. Cohrssen, and R. C. Morrison, Eds.), pp. 61–88. Praeger, New York.

Perera, F. P., and Weinstein, I. B. (1982). Molecular epidemiology and carcinogen-DNA adduct detection. New approaches to studies of human cancer causation. *J. Chronic Dis.* **35**, 581–600.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *J. Chronic Dis.* **29**, 175–188.

Prentice, R. L., Moolgavkar, S. H., and Farewell, V. T. (1986). Biostatistical issues and concepts in epidemiologic research. *J. Chronic Dis.* **39**, 1169–1183.

Rothman, K. J. (1986). *In* "Modern Epidemiology," pp. 89–91. Little, Brown, Boston.

Schulte, P. A. (1987). Methodologic issues in the use of biologic markers in epidemiologic research. *Amer. J. Epidemiol.* **126**, 1006–1016.

Smith, T. J. (1987). Exposure assessment for occupational epidemiology. *Amer. J. Ind. Med.* **12**, 249–268.

Susser, M. (1973). "Causal Thinking in the Health Sciences." Oxford Univ. Press, New York.

Thompson, S. G. (1983). A method of analysis of laboratory data in an epidemiological study where time trends are present. *Stat. Med.* **2**, 147–153.

Underhill, D. W., and Radford, E. P. (Eds.). (1986). "New and Sensitive Indicators of Health Impacts of Environmental Agents." Center for Environmental Epidemiology, University of Pittsburgh, Pittsburgh, PA.

Weed, D. L., Selmon, M., and Sinks, T. (1988). Links between categories of interaction. *Amer. J. Epidemiol.* **127**, 17–27.