MOLECULAR EPIDEMIOLOGY

EDITED BY

Molecular Epidemiology

Principles and Practices

Edited by

Paul A. Schulte

Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health
Robert A. Taft Laboratories
Cincinnati, Ohio

Frederica P. Perera

Department of Epidemiology
Columbia University
New York, New York

Academic Press, Inc.
Harcourt Brace & Company

San Diego New York Boston London Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

Academic Press, Inc.
1250 Sixth Avenue, San Diego, California 92101-4311

United Kingdom Edition published by
Academic Press Limited
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Molecular epidemiology : principles and practices / edited by Paul A. Schulte, Frederica Perera.
p. cm.
ISBN 0-12-632345-3
1. Molecular epidemiology. 2. Biochemical markers. I. Schulte, Paul A. II. Perera, Frederica P.
[DNLM: 1. Biological Markers. 2. Epidemiologic Factors. 3. Molecular Biology. QH 506 M71925]
RA652.5.M65 1993
614.4–dc20
DNLM/DLC
for Library of Congress 92-49193
CIP

PRINTED IN THE UNITED STATES OF AMERICA
93 94 95 96 97 MP 9 8 7 6 5 4 3 2 1

3

Validation

Paul A. Schulte
and Frederica P. Perera

The contribution of molecular epidemiology to etiologic research, risk assessment, or disease prevention and control depends on the use of valid biologic markers. The use of invalid markers can lead to wrong conclusions and costly programs. Validity is the best approximation of the truth or falsehood of a marker (Cook and Campbell, 1979). Validity is a sense of degree rather than an all-or-none state. To validate the use of biologic measurement as a marker, it is necessary to understand the relationship between the marker and the event or condition of interest. Biologic markers can be validated against exposure, disease, or susceptibility events. To date, most molecular epidemiologic research has involved validating biologic markers (as opposed to using them for etiologic and intervention research or risk assessment). When the validity, reliability, and practicality of a marker have not been demonstrated, pilot studies are useful. Perera (1987), among others, has demonstrated a strategy and approaches for these pilot studies: characterize the marker in known high-dose groups such as chemotherapy patients, proceed to highly exposed groups such as occupational groups, then use less exposed occupational and environmental groups. The goal of these studies is to determine characteristics of markers that must be known prior to their use in large population studies. These characteristics include a dose–response relationship, marker persistence, inter- and intraperson variation, correlation between markers, and correlation with clinical response. For example, Perera *et al.* (1992) studied cancer patients treated with *cis*-platinum (*cis*-DDP)-based chemotherapy and found posttreatment differences in a battery of biologic markers, including increased levels of platinum–protein and platinum–DNA adducts and increased incidence of sister chromatid exchanges, micronuclei, and gene mutation at the glycophorin A locus.

In many, but not all, of the studies (conducted in the 1980s) using biomarkers, particularly genetic and molecular markers, the validity of findings

Molecular Epidemiology: Principles and Practices

was limited or in question because adequate attention was not given to subject selection, control of confounding, and choice of statistical analyses. Subjects often appeared to be selected with no appreciation of the impact of bias, attention to confounding factors, or attention to sample size, statistical power, or other design features. Granted, many of the early studies were conducted to see if an assay "worked" or to examine how the assay performed under a range of conditions. In most of these cases, investigators had the good sense not to include statistical analyses, since such assessments generally were not appropriate. In other cases, however, studies included practically no discussion of statistical design features or evaluation of the underlying assumptions for the statistical tests employed.

Often in these studies, attention was paid only to one definition of validity of a marker, which has different meanings to laboratory scientists and epidemiologists. To the laboratory scientist, validity generally means the ability of a test to respond in the presence of a marker and to not respond in its absence. To the epidemiologist, validity pertains to predictive value, that is, the probability that a person who has a marker actually experiences the event being indicated. Ultimately, from an epidemiologic viewpoint, a marker will be valid and useful if it reduces misclassification, provides better interpretation of exposure–disease associations, or is useful in prevention or control of disease.

Research involving biologic markers first must determine whether such markers are valid—whether they measure what they are believed to measure—and what they mean with respect to the risk of disease. Only after validation can markers be used effectively for etiologic and disease-control research and in risk assessments. The validation of biologic markers for use in molecular epidemiologic research requires that extensive laboratory work be performed prior to testing in humans. Key to the validation procedure for markers of effect is agreement on what constitutes a "critical effect" (Ashby, 1987; Hernberg, 1987). Exposure to xenobiotic substances results in a range of perturbations to biologic systems. To determine which effects are critical (that is, indicate some aspect of a disease response) and which are merely adaptive, it is necessary to relate critical effects to dose estimates, to determine what factors affect dose, and to define a no-effect level (Ashby, 1987; Hernberg, 1987). A similar but less familiar process is required to validate markers of exposure and markers of susceptibility.

Although homage often is paid to the concept of validity, little attention has been given to its meaning and how to evaluate it (Schulte and Mazzuckelli, 1991). The objective of this chapter is to identify and explore the range of considerations that constitute the concept of validity and to address how validity pertains to the use of biologic markers in epidemiologic risk research and quantitative health risk assessments.

Three broad categories of validity can be distinguished: measurement validity, internal study validity, and external validity (Figure 3.1). Measure-

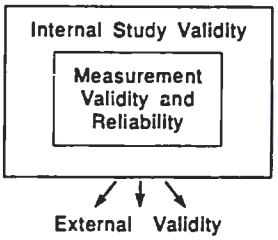


FIGURE 3.1 Three categories of validity.

ment validity has been defined as an expression of the degree to which "... a measurement measures what it purports to measure" (Last, 1988). Internal study validity is the degree to which inferences drawn from a study pertain to the actual subjects in the study (Rothman, 1986). External study validity is the extent to which the findings of a study can be generalized to apply to other populations (Last, 1988).

Biologic markers and the studies that include them need to be shown to have measurement, internal, and external validity before they can be used accurately in etiologic research and quantitative risk assessment. The use of invalid markers can result in nondifferential misclassifications of exposure or outcome, which can lead to underestimation of a true effect (Hogue and Brewster, 1988). Risk assessments based on studies that underestimate a true effect can lead to regulations that provide exposure limits that are thought to be "safe" but, in fact, are not. Conversely, a differential misclassification bias, depending on the direction of the bias, can lead to regulations that provide exposure limits that are either too high or too low. In quantitative risk assessment, the inferences derived from small study groups are generalized to larger populations. The strength of those inferences depends on the methodology of the study, including measurements and other design factors that lead to the results. Invalid measurements, inferences, or generalizations may lead to erroneous risk assessments. In this chapter, the three categories of validity are discussed in terms of how they apply to biologic markers for epidemiologic research and quantitative risk assessment.

Measurement Validity

Biologic measurements are among the principal building blocks of molecular epidemiology. If measurements are invalid, the research and risk assessments constructed from those measurements are likely also to be invalid. Researchers can use three aspects of measurement validity to characterize the markers they intend to employ in a study. Measurement validity reflects the extent to which a marker of a phenomenon has (1) content validity, that is, pertains to

the underlying biologic phenomenon; (2) construct validity, that is, correlates with other relevant characteristics of the underlying phenomenon; and (3) criterion validity, that is, predicts some aspect of the underlying phenomenon. In general, these three components of measurement validity are best assessed in terms of the extent or degree to which they apply to the underlying phenomenon, rather than as an all-or-none condition (Nunnaly, 1967). The precision and reliability of measurements of biologic markers also have an effect on their validity and use in research. These issues will be discussed in Chapter 4.

Content Validity

"Content validity is the extent to which a marker represents the underlying biological phenomenon being studied" (Last, 1988). For example, a marker of internal dose will have content validity if it reflects the dose contributed by all routes of exposure. A marker of effect will have content validity if it encompasses the essential characteristics of the disease it represents. In other words, the marker must pertain to the appropriate target organ or the relationship of the marker to the natural history of the disease in question must be unambiguous. For example, a DNA adduct of benzo[*a*]pyrene [B(*a*)P] will have content validity as a marker of exposure in a study of B(*a*)P exposure and lung cancer, since the involvement of DNA in B(*a*)P-induced carcinogenesis is well documented. In contrast, the development of DNA adducts at the N7 position might not have content validity as a marker of biologically effective dose if the O6 methylguanine adduct is shown to be related more clearly to the carcinogenic process. However, the N7 adducts might be reasonably valid markers of the biologically effective dose of B(*a*)P if (1) the production of O6 and N7 adducts is directly proportional, as would be expected if they were produced by the same activated B(*a*)P metabolite; and (2) if relatively little time is allowed for possible differential repair, or the likely effect of differential repair on the measurement is removed during extrapolation of the data to zero time (Schulte and Mazzuckelli, 1991).

A proper assessment of content validity considers the extent to which the marker pertains to the phenomenon of interest (exposure, effect) or the extent to which the marker represents a relevant feature of that phenomenon (Schulte and Mazzuckelli, 1991). For example, to assume that hydroxyethyl histidine adducts of hemoglobin are valid markers of the internal dose of occupational exposure to ethylene oxide would be erroneous. The presumed marker would lack complete content validity, since hydroxyethyl histidine adducts of hemoglobin can result from exposure to other substances that contain ethyl groups. Further, populations with no known exposure to ethylene oxide have been shown to form hydroxyethyl histidine adducts of hemoglobin. Without considering content validity, one might reach erroneous conclusions if only occupational ethylene oxide exposure is assumed to be

responsible for the observed adducts. By subtracting the number of adducts attributable to factors other than the exposure under study from the total number of adducts formed, valid measures might be developed. This manipulation requires the evaluation of a nonexposed comparison group and attention to other sources of ethyl groups that generate hemoglobin adducts.

Since content validity is assessed by professional judgment and logical consensus, there are no universally accepted criteria for its determination (Zeller and Carmines, 1980). However, it is possible to strengthen determinations of content validity if judgments are made by a group of experts or specialists with understanding of a particular marker. The focus of such judgments should be the degree to which the marker represents the underlying phenomenon. Establishing content validity is especially difficult in situations in which it is needed most, that is, when there is an incomplete understanding of the underlying characteristics of the exposure–disease process. Content validity should be established before a marker is used in intermediate or large scale human studies.

Construct Validity

Construct validity describes the extent to which a marker corresponds with other relevant characteristics of the underlying phenomenon, that is, the theoretical concepts or constructs concerning the phenomenon under study (Last, 1988). This correspondence is exhibited in part by association of the subject marker with other markers or covariates of the phenomenon. For example, if the characteristics of a phenomenon change with age, a marker with construct validity will change accordingly (Last, 1988). Further, if the marker shows no associations with other variables that are reasonably expected to be linked with the phenomenon under study, then the marker may be of questionable relevance in an epidemiologic study or subsequent risk assessment.

Construct validity is sometimes difficult to distinguish from content validity when describing biologic markers, but should be evaluated whenever general understanding of the underlying phenomenon is not clear. Hence, if a marker is a candidate for inclusion in an epidemiologic study of exposure or outcome and the actual role of the marker in the exposure–outcome continuum has not been established (i.e., its content validity has not been established), it still may be useful as a covariate if it can be shown to have construct validity. For example, if one wants to study the relationship between exposure to a dietary toxicant and some immunologic outcome, a marker of DNA repair could be useful as an independent variable. A marker of DNA repair may not have content validity for a study of immunologic changes after exposure to a dietary toxicant but still might be useful as a marker of biologic age because of its construct validity and, thus, could be used as a covariate in a multivariate model of dietary exposure and immunologic outcome.

Criterion Validity

Criterion validity describes the extent to which a marker correlates with the phenomenon being studied (Last, 1988). The criterion validity of a marker is assessed in terms of its sensitivity, specificity, and predictive value. Griffith *et al.* (1989) have distinguished the terms sensitivity and specificity as they refer to laboratory methods to detect a marker. The terms also are used to describe the ability of a marker to detect an exposure, or to detect or predict an event in a population.

> Laboratory sensitivity to detect a marker refers to the ability of a detection system to respond in the presence of the markers. Population sensitivity, in contrast, is the ratio of numbers of subjects positive for both the marker and the event to the number of subjects with the event.
>
> Laboratory specificity refers to the detection system's ability to fail to respond in the absence of the marker. Population specificity is the ratio of the number of subjects negative for both marker and event to the number of subjects that are negative for the event. (Griffith *et al.*, 1989)

Perhaps the most valuable indicator of whether a marker is valid is the predictive value. Predictive value for a marker of disease is the proportion of people studied with a particular disease among all the people who have the marker. Thus, in terms of whether the marker reflects disease, *predictive value* is

$$\frac{\text{true positives}}{\text{true positives and false positives}}$$

Predictive value can be calculated in terms of those with (predictive value positive) or without (predictive value negative) a marker. The relationship between sensitivity, specificity, and predictive value is shown in Figure 3.2.

Ideally, a marker assay should have high sensitivity and specificity but, in reality, one can increase the value of one aspect at the expense of the other. Thus, it is necessary for the investigator to decide where between a positive and negative test to place the assay result cut off point (Hulka, 1990). The complex relationship between population sensitivity, specificity, and the choice of threshold point is perhaps best depicted through a receiver–operator characteristic (ROC) curve. ROC analysis began during the 1940s as an approach to optimizing signal detection and has become a tool increasingly used to evaluate screening and diagnostic tests and algorithms (Swets *et al.*, 1979; Hanley, 1989; Thompson and Zucchini, 1989). The most common form of the ROC curve depicts the true-positive fraction (TPF) on the vertical axis and the false-positive fraction (FPF) on the horizontal axis (Figure 3.3). The curve is constructed by plotting the TPF and FPF for each decision threshold (or "cut off") value for which data are available, then connecting those points through an appropriate regression methodology (e.g., England, 1988; Tosteson and Begg, 1988). Both linear and probability co-

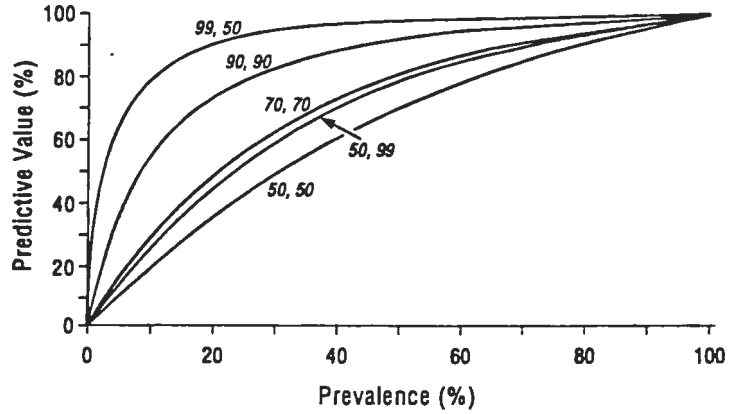


FIGURE 3.2 Relationship between predictive value and prevalence of marker. Numbers on lines represent percentage sensitivity and percentage specificity.

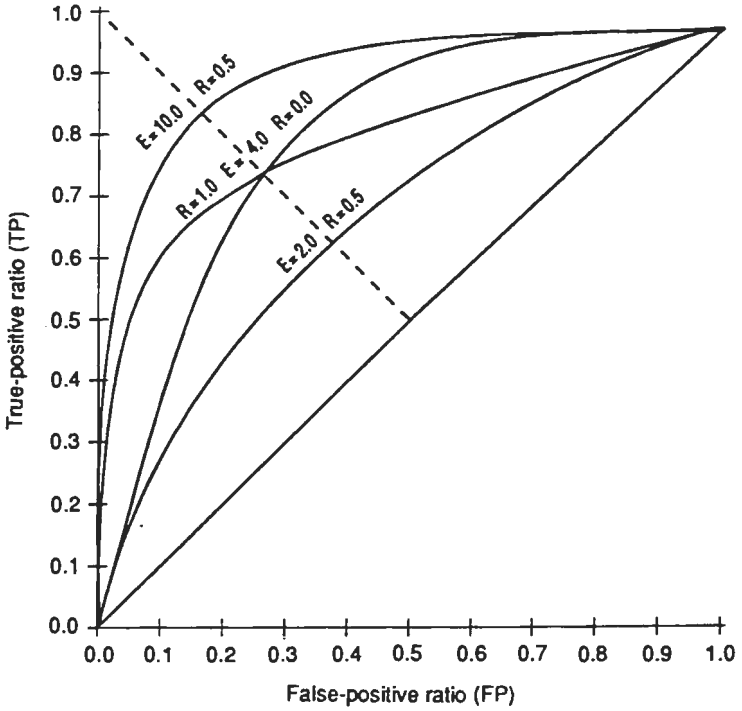


FIGURE 3.3 Data generated from receiver-operator characteristic (ROC) curves can take different shapes depending on sensitivity and specificity at different cut-off points. A general model for fitting these curves includes the parameter E (reflecting the height of the curve) and the parameter R (reflecting the skewness of the curve). (Reprinted with permission from England, 1988.)

ordinates have been used for TPF and FPF axes (Swets *et al.,* 1979; England, 1988). The decision threshold values may be continuous (as with the concentration of a chemical analyte) or categorical (as with radiologic or histopathologic assessment).

ROC curves plotted on linear scales allow a quick visual assessment of assay performance. In general, the more nearly the curve approaches the left corner of the coordinate system, the better the test performance. The point of closest proximity to the left corner represents the optimal decision threshold value, that is, the greatest sensitivity with the least nonspecificity. When comparing multiple assays depicted by ROC curves on the same scale, a simple assessment is provided by noting that the test with the greatest area under the curve provides the best overall combination of sensitivity and specificity (Hanley and McNeil, 1982). However, not all regions of the curve are necessarily of equal interest; more advanced techniques have been applied to comparisons of tests by ROC analysis (McClish, 1987, 1989).

In biomedical science, ROC analyses have been applied extensively to diagnostic imaging, clinical laboratory testing (e.g., Hanequin *et al.,* 1987; Hunink *et al.,* 1990), and the use of biologic markers in psychiatric diagnosis (Somoza and Mossman, 1991). These tests also have been used to address more general questions, such as the gains achieved through replicate analysis (Metz and Shen, 1992) and the efficiency of training neural networks for artificial intelligence (Meistrell, 1990). Of course, like any other method of test evaluation, conventional ROC analysis is entirely dependent on the accuracy of the health end point assessment to obtain true values for sensitivity and specificity. Some investigators have suggested that the use of ROC analysis with correlative information can avoid the need for such independent assessment of health end points (Henkelman *et al.,* 1990); however, great caution is required with such approaches.

Key in the validation and use of biologic markers for studies of human populations is the prevalence of the underlying condition being represented. Figure 3.2 shows how the positive predictive value of a marker is reflective of the underlying prevalence of the marker. Thus, for example, a marker assay that is 90% sensitive and 90% specific will still only have a predictive value of 50% when the prevalence of the underlying event is 10%. Field studies that do not incorporate prevalence considerations in planning are likely not to be able to detect an association, even if one exists.

In general, the sensitivity and positive predictive value of a marker can be increased by studying susceptible populations. For example, as Wilcosky (1992) describes, exposure to acid aerosols/oxidants can exacerbate symptoms of sensitized asthmatics, whereas it will likely generate an inflammatory response in nonasthmatics at lower exposure levels; thus, markers of inflammation will have greater sensitivity in studies of exposed and unexposed asthmatics.

Strategy for Validating Intermediate Markers

The strategy for validating markers that are intermediate between an exogenous exposure and disease involves two steps: (1) selection of candidate markers and (2) identification and quantification of the association between the markers and the disease (Schatzkin *et al.*, 1990). Selection of candidate markers from the large pool of available markers can be accomplished using case series, ecologic studies, and animal or *in vitro* experiments. In some instances, these markers can be used in case-control and cohort study designs (see Chapter 5) to identify an association with disease. This association can be quantified using the attributable proportion (AP) (Cole and MacMahon, 1971; Miettinen, 1974; Schatzkin *et al.*, 1990) or other measures of association such as odds ratio or relative risk. AP is defined as the proportion of cases of disease that is attributable to the marker. The AP can be determined directly from the sensitivity (S) and the relative risk (R), as defined in the formula from the data array in Table 3.1. Schatzkin *et al.* (1990) have developed examples (Table 3.2) of how AP varies with different levels of relative risk and sensitivity. Table 3.2 shows the AP for various levels of S and R. Levels of low S and high R have an AP much less than those with high S and low R.

To assess the validity of an intermediate marker of disease, it is useful to consider the simplified framework in Figure 3.4. In Figure 3.4A, a single marker is shown to be linked causally and, hence, necessary and sufficient for disease. In this case, AP approaches 1.0 and S equals 1.0. This marker would have a large degree of content and criterion validity.

In most instances, more than one pathway is involved; this is shown in Figure 3.4B. Here, neither pathway is necessary for disease but either is sufficient. The content validity may be judged to be relatively high, but the criterion validity is reduced (as evidenced by S, which is less than 1.0). The AP is also less than 1.0. The closer the AP is to 0, the less a marker can be considered a valid surrogate for disease.

TABLE 3.1 Calculation of Attributable Proportion (AP)[a]

	Disease	
Marker	Yes	No
+	A	B
−	C	D

Source: Shatzkin *et al.* (1990).

[a] Sensitivity (S) = A/A + C; relative risk (R) = [A/ (A + B)] / [C/ (C + D)]; using these values, AP = S (1 − 1/R).

TABLE 3.2 **Relationship between Attributable Proportion, Relative Risk, and Sensitivity**[a]

Sensitivity of intermediate marker	Relative risk for intermediate marker positives									
	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	
0.1	0.03	0.06	0.07	0.08	0.08	0.09	0.09	0.09	0.09	0.09
0.3	0.10	0.18	0.21	0.23	0.25	0.25	0.26	0.26	0.27	0.27
0.5	0.17	0.30	0.36	0.39	0.41	0.42	0.43	0.44	0.45	0.45
0.7	0.23	0.42	0.50	0.54	0.57	0.59	0.61	0.62	0.63	0.63
0.9	0.30	0.54	0.64	0.70	0.74	0.76	0.78	0.79	0.81	0.81

Source: Adapted from Schatzkin *et al.* (1990).

[a] $[AP = S(1 - 1/R)]$

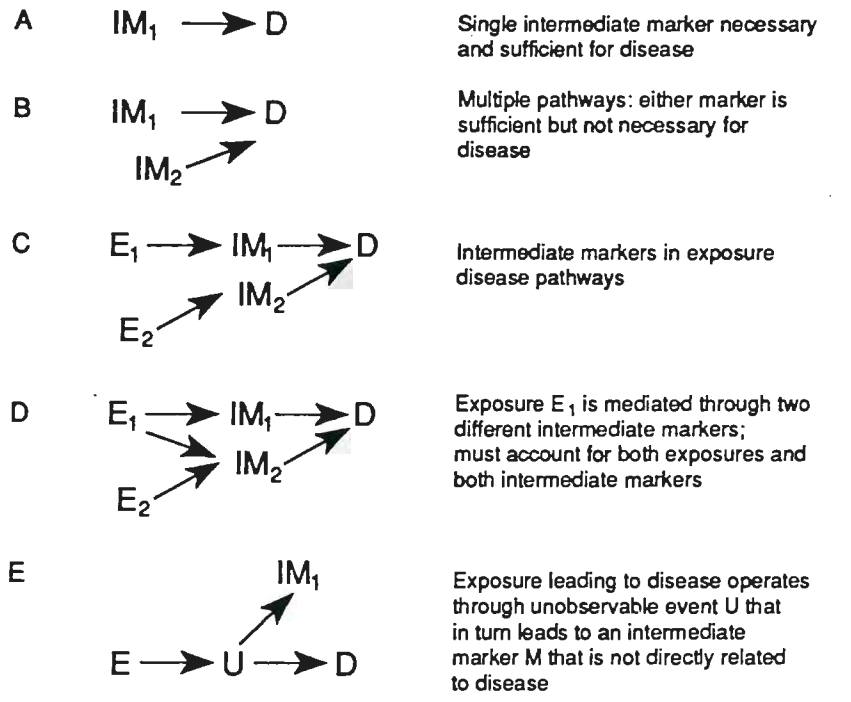


FIGURE 3.4 Framework for considering exposure–marker–disease relationships. (Adapted from Schatzkin *et al.*, 1990.)

In Figure 3.4C, separate exposure–marker–disease pathways are shown. The effects of each exposure are mediated entirely through their intermediate markers. In Figure 3.4D, a different situation is depicted. The effect of exposure E_1 is mediated partially through IM_2. The AP for neither of these markers will be 1.0. The marker that has the highest AP is the one with the best content and criterion validity (Schatzkin *et al.*, 1990).

Another situation is shown in Figure 3.4E. An exposure leading to disease operates through an unobservable event U that, in turn, leads to an intermediate marker that is not directly on the disease pathway. This process can be illustrated, for example, with chromosomal micronuclei, which, although they occur in nonviable cells, reflect genotoxic events similar to ones that may be necessary for carcinogenesis (Schatzkin *et al.*, 1990). In this case, the marker may have some content validity, but not of the greatest degree. The criterion validity of the marker would depend on its relationship to U and the relationship of U to disease. If the latter is not high, the former will be meaningless and the marker cannot be used as a surrogate.

In summary, the quality of epidemiologic studies and risk assessments depends on the quality and validity of measurements. Assessment of validity

involves the empirical process of determining that the candidate biomarker is indeed measuring what it intends to measure. To achieve that goal, a range of validation studies should be performed prior to use of the marker in etiologic research or intervention.

A practical example of the employment of the concepts of content, construct, and criterion validity is shown in the scheme in Figure 3.5 and in Table

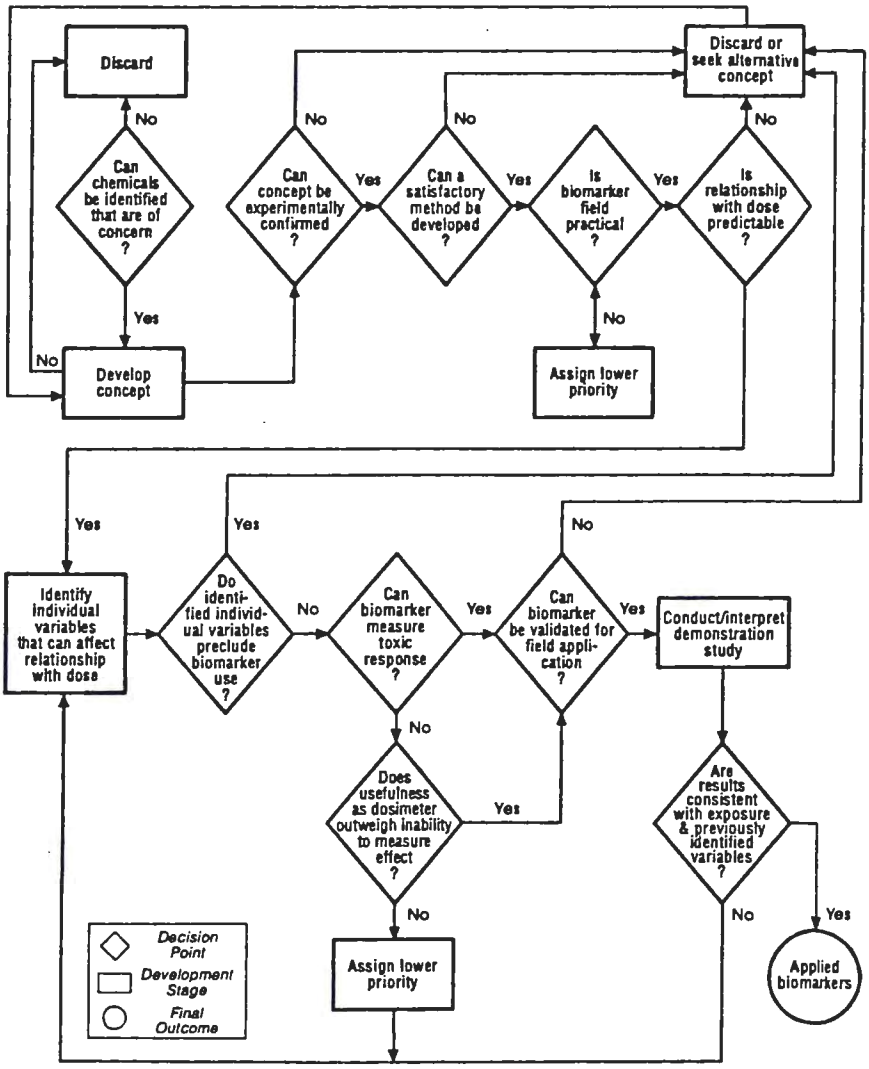


FIGURE 3.5 Decision model for development of biomarkers of exposure. (Bull, 1989; Stevens *et al.*, 1991.)

TABLE 3.3 Steps in the Development of a Biomarker

Step	Action required	Relative importance[a]
1. Chemical selection	Prioritize based on occurrence, significant human exposure, potential for adverse human health effects	C
2. Conceptualization	Identify logical consequence of chemical exposure that might serve as a useful measure of exposure	C
3. Confirmation of concept	Experimentally confirm validity of basic concept	C
4. Develop method of measurement	Identify method for detecting changes in biomarker at doses at or below those producing toxic effects	C
5. Biomarker practical for field?	Develop plausible field methodology and develop sufficient sensitivity of biomarker to monitor existing exposures	L
6. Establish dose–response relationship	Characterize pharmacokinetics and metabolism of chemical (consistent relationship to systematic dose is critical; knowledge of effective dose is limiting)	C,L
7. Identify variables affecting relationship with dose	Establish specificity of response and identify lifestyle, genetic, disease state, therapeutic, or occupational variables that modify the response	C,L
8. Measures toxic effect?	Provides advantage only among biomarkers of equal ability as measures of exposure	N
9. Validation of applicability to humans	Conduct pilot study in small groups of humans with defined exposure gradients to chemical of interest	C
10. Conduct demonstration study	Determine whether variation in response in larger population can be accounted for by known variables	C

[a] C, critical to the application of the biomarker; L, limiting to the application of the biomarker, i.e., places limits on interpretation of results for secondary purposes, e.g., risk assessment; N, nice to have, but not essential to the application of the biomarker (Bull, 1989).

3.3. This example is a decision algorithm developed to assess biologic markers of exposure. An analogous approach could be developed for markers of effect and susceptibility.

Internal Validity

Proper use of biologic markers can enhance the internal validity of epidemiologic studies. The internal validity of a study is the degree to which index and comparison groups are selected and compared so, apart from sampling errors, the observed differences between the dependent variables are attributed

only to the hypothesized effect (Last, 1988). Such a result is validity in the estimation of effect, and is dependent on the ability to control bias. Internal study validity has been discussed widely in epidemiology textbooks. In this section, we discuss some issues of internal validity that pertain to the use of biologic markers. Although some of this discussion is specific to markers, other more general issues also merit comment.

Statistical bias is a distortion that may result during evaluation of an association and can occur when subject selection is unequal according to disease or exposure status (Ozonoff and Wartenberg, 1991). When selecting subjects for studies, the investigator must identify factors such as background rates of markers and the range of normal so classification and subject selection are equal for the groups being compared. These issues have been discussed elsewhere (Hulka and Wilcosky, 1988; Schulte, 1989). Bias can lead to misclassification of subjects based on exposure or disease, and may result from failure to adjust for other variables that are also predictive of the disease of interest.

Misclassification

Since the basis of the epidemiologic method is the comparison of rates of disease in exposed versus unexposed populations, a mix of exposed and unexposed or disease and nondisease characteristics (or both) clearly will harm the chances of seeing a difference (Ozonoff and Wartenberg, 1991). This misclassification can be "differential," that is, systematically wrong in one aspect of classifying exposure or disease, or nondifferential, that is, nonsystematic. Differential misclassification of exposure or disease can reduce the validity of a study (Hogue and Brewster, 1988; Hulka and Wilcosky, 1988). Biologic markers that reduce misclassification may enhance study validity. Similarly, biologic markers can contribute to the reduction of nondifferential misclassification. This type of misclassification, which has been considered a lesser threat to validity, can result in bias toward the null value.

The key to valid epidemiologic studies is a strong rationale for selection of the exposure (dose) variables. The choice of exposure variables for individuals exposed to toxic substances can range from anamnestic information gathered by questionnaire to detailed measurement of biologic markers (Rogan, 1988). However, as Rogan (1988) notes, ". . . in the strict sense, any exposure information other than biological effective dose is a surrogate." Thus, the question is, "How closely does the exposure surrogate, used to derive a model, resemble the actual exposure under study?" Valid biologic markers can provide empirical data that is preferable to deductively derived estimates (Rogan, 1988).

For example, Lawrence and Taylor (1985) demonstrated the value of empirical exposure measurements when they were confronted with the problem of assessing historical polychlorinated biphenyl (PCB) exposures among

women who manufactured electrical capacitors. The purpose of their investigation was to determine the effects of PCB exposure on reproductive outcomes from 1979 to 1983. Although the investigators did not have actual serum PCB measurements for that period, they did have a complete work history for each subject. The industrial hygiene data allowed classification of each job in terms of a low, medium, or high PCB concentration. The challenge was to choose a surrogate that best approximated the true exposure. The investigators also had sera that had been gathered in 1976 from a sample of workers as part of a general company survey. Using those data, the investigators developed a regression model to estimate the explicit serum PCB concentration as a continuous variable level for each woman during each of her pregnancies between 1979 and 1983. Hence, the serum PCB concentration, derived from a sample of subjects, was used as a biologic marker to construct a more accurate estimate of the true exposure than was available using job-classification data (Lawrence and Taylor, 1985).

Analytical Adjustment for Other Variables

Proper data analysis depends on the choice of the correct mathematical model, especially when multiple variables exist. The strongest models take into account *a priori* hypotheses specific to the topic under study. The incorporation of biologic markers into study designs and mathematical models also implies an understanding of the direction and mechanism of action. For example, in a study of hospital workers exposed to low levels of ethylene oxide (Schulte *et al.*, 1992), the association between hydroxy ethyl hemoglobin adducts and ethylene oxide exposure was assessed using multiple linear regression based on two assumptions: (1) exposure at low doses was linear and (2) the impact of confounding and intervening variables could be evaluated in this approach. Thus, as shown in Table 3.4, the level of hydroxy ethyl adducts showed an exposure–response relationship when group means were adjusted for confounding factors such as age, race, cigarette smoking, and education. Additionally, controlling measurement validity makes it possible partially to control study validity, since measurement errors can produce biased estimates of regression coefficients used in models (Louis, 1988).

Longitudinal studies that employ biologic markers will be used increasingly in molecular epidemiologic studies and quantitative risk assessments. The validity of those study results will depend in part on the analytical approach that is selected (Dwyer *et al.*, 1992). Such studies may involve repeated measures of a continuous random variable; thus, there may be measurement errors that are considered random among persons but are autocorrelated within persons. The use of autoregressive modeling by epidemiologists for the analysis of longitudinal data is increasing, and will be more frequent in studies involving biologic markers. These models permit consideration of the time course of change of a variable (Rosner *et al.*, 1985). Other

TABLE 3.4 Relationship between Ethylene Oxide Exposure and Hemoglobin Adducts in Hospital Workers

Ethylene oxide exposure (ppm-hr)	Mean[a]	*N*	Hemoglobin adducts[b] (pmol/mg Hb)
0	0 (0)	8	0.06 (0.02)
>0–32	12.8 (11.0)	32	0.09 (0.01)
>32	105.2 (45.7)	11	0.16 (0.02)
Comparison	0 vs > 0–32		p = 0.27
	0 vs > 32		p = 0.002
	>0–32 vs >32		p = 0.004

[a] Means were adjusted for age, race, cigarette smoking, and education.
[b] Values are means. Standard deviations given in parentheses.

methods for analyzing repetitive measures with a Gaussian error structure have been reviewed by Louis (1988), who concluded that this area needs continued statistical, numerical, and interpretive research and development.

External Validity

Application of molecular epidemiologic research often will be in the assessment of health risks to groups not included in the studies. Risk assessment is an effort to address a condition of incomplete data (Erdreich, 1988) and involves the extrapolation (or generalization) of known exposure–response data to ill-defined risk situations in target populations. External validity is the degree to which a study can produce unbiased inferences about those target populations. For risk assessment, external validity involves the appropriateness of the following extrapolations: within or between populations or species, from high doses to low doses, and between different organs within a species. All these efforts can be enhanced using biologic markers common to each population or species. Allometric assessment of effects in different species can be determined by observing how the same marker varies with similar exposures. Valid extrapolation requires an understanding of the major events that can cause such inter- and intraspecies differences. For example, in chemical carcinogenesis, the following factors appear to play a critical role in species and organ differences: the overall balance of metabolic activation and detoxification, the balance of DNA damage and repair, the persistence of DNA damage, and tumor formation (Slaga, 1988).

Many uncertainties are attendant to extrapolation of data from an epi-
emiologic study of a smaller group to a large population. The characteris-
cs that make a study internally valid are often barriers to extrapolation.
evertheless, extrapolation is a current practice in risk assessment. The use
valid biologic markers may allow some evaluation of whether a particu-
r extrapolation is warranted, whether the variability is too extreme, or
hether differences in susceptibility have resulted in sensitive subgroups
rdreich, 1988).

Extrapolation to low doses (or exposures) involves determining (or as-
ming) the shape of the dose–response curve. In some instances, establish-
g a dose–response relationship in a risk assessment might be considered
meta-analytic procedure, that is, results from different studies might be
mbined to provide a larger sample size or a broader range of dose esti-
ates. The validity of this effort can be enhanced if the same markers are
ed in different studies or if different markers have been shown to be corre-
ted, that is, to have construct validity.

The contribution of macromolecular adducts to low-dose extrapolation
as been heralded as a potential improvement to risk assessment. However,
e use of biologic markers also can be a source of confusion in risk assess-
ent. Most studies of adducts in humans have not yet demonstrated a clear
se response, perhaps because of the wide variability in human response
d the current inability to determine true individual exposures (Perera,
87, 1988). However, trends with exposure have been observed in occupa-
nal settings (Perera *et al.*, 1988). Until the sources of variability can be
entified and their impact evaluated, the absence, or faulty characterization,
a dose response will limit the usefulness of this class of biologic markers in
sk assessment (Brown, 1988; Motulsky, 1988). A potentially major source
differential susceptibility in dose response is the phenotypic variation of
etabolic parameters (Motulsky, 1988). This variation rarely has been con-
dered in epidemiologic studies and risk assessments, partially because such
ata have not been developed.

The effect of the choice of a dose variable on risk estimates can be severe,
pecially when the pattern of exposure that the estimates supposedly reflect
ffers from the predominant pattern experienced by a study cohort (Crump
d Allen, 1985). The use of a biologic marker of exposure can help reduce
e impact of using an ambiguous dose variable, because it can reflect the
ue dose more accurately, even in studies in which exposures are observed
have occurred over a wide range. For example, attempts have been made
compare biologically effective doses at high exposures at which tumors are
served to low exposures to determine whether linearity of the carcinogenic
fect is a valid assumption. Perera (1987, 1988) has concluded that extensive
ata on DNA, RNA, and protein binding indicate that macromolecular ef-
cts, at the lowest administered doses, generally follow first-order kinetics,

that is, the rate of binding in target organs *in vivo* is directly proportional to the administered dose. Since many carcinogens covalently bind to and structurally alter DNA, the adducts that are formed are conceptually valid markers of exposure and, possibly, of effects. Moreover, the ratio of surrogates for DNA adducts, for example, protein adducts, to dose have been shown to be constant over a dose range of 10^{-5} mol/kg to 10^{5} mol/kg. However, as Swenberg (1988) asked, ". . . What data-bases are available so that such a molecular dosimetry approach can be validated?" Few carcinogens have been evaluated for which the exposure range is greater than one order of magnitude.

Validation and Selection of Biomarkers

A determination of the validity of a biomarker for molecular epidemiologic studies should be made on the basis of six fundamental criteria. (See Perera, 1987, for review.)

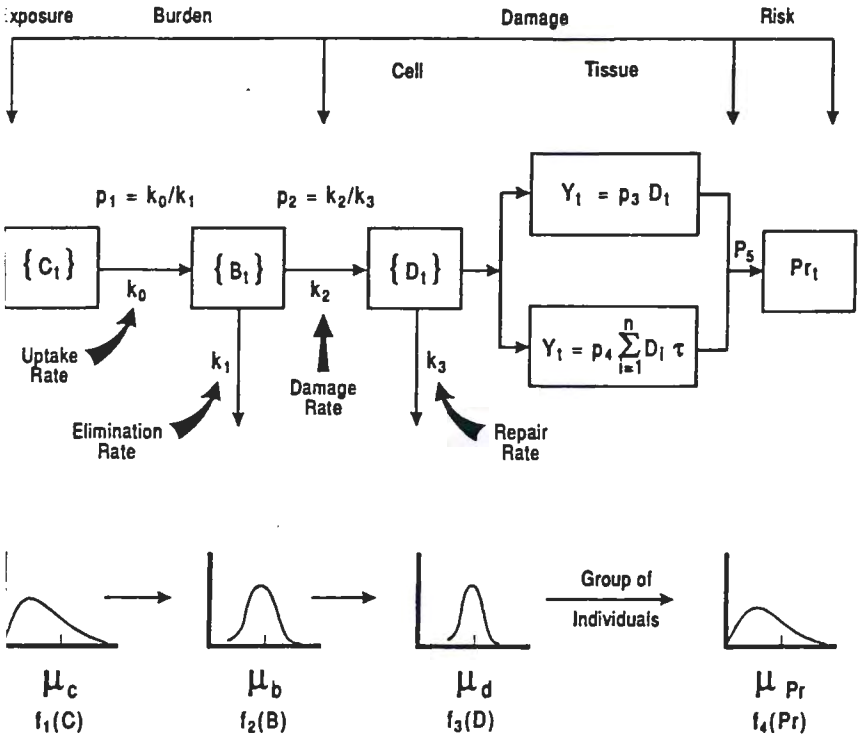
Biologic Relevance

A clear hypothesis or model of exposure–dose or exposure–response relationships must be available that defines the role of the specific marker in relation to other points in the continuum (Perera, 1987). This relationship is content validity. For example, the biologic relevance of carcinogen–DNA adducts as initiating events in carcinogenesis and carboxyhemoglobin levels as indicators of reduced oxygen-bearing capacity of erythrocytes are well established. The meaning of sister chromatid exchanges (SCEs) and H-*ras*-encoded p21 proteins is more ambiguous.

The biologic relevance of markers of exposure is generally easier to establish than that of effect markers. Determining the ability of markers of early biologic effect (such as SCEs or mutated oncogenes) to predict cancers or reproductive risk requires prospective or "nested" case-control studies. Circumstantial evidence could be provided through findings of increased biomarkers in groups (such as workers) with historically high cancer rates. Unfortunately, longitudinal studies have not yet been carried out to assess the predictive ability of effect markers. Indeed, only limited biomarker data are available from cross-sectional studies of workers exposed, in most cases by inhalation, to known occupational carcinogens.

Understanding Pharmacokinetic Aspects

Molecular epidemiology relies on a conceptual model that depicts the exposure–response continuum as a sequence of time series related to xenobiotic exposure, burden, damage, and risk (see Figure 3.6). Analysis of the link between these processes identifies two kinetic conditions that are neces-



IGURE 3.6 Conceptual model relating exposure, dose, damage, and risk of disease for an ıdividual worker or uniformly exposed group of workers. The symbols labelled k_0–k_3 represent ıte constants for uptake (k_0), elimination (k_1), cellular damage (k_2), and repair (k_3). The proɔrtionality constants labelled p_1–p_5 result from linear transfer functions between succeeding ries. The distribution functions shown at the bottom represent the following: exposure [$f_1(C)$], ırden [$f_2(B)$], cellular damage [$f_3(D)$], and population risk [$f_4(Pr)$]. (Rappaport, 1991.)

ıry for variability of exposure to affect individual risk of chronic disease ɔpreciably (Rappaport, 1991). To illustrate these issues, it is useful to conder occupational exposures to a toxicant. Rappaport (1988, 1991) has de:ribed the functional relationships between airborne exposure distributions ıd subsequent distributions of dose, damage, and risk of disease. Key to ıderstanding these relationships is the realization that environmental exɔsures vary greatly over time and among individuals. The validity of a arker will depend on its biologic relationship to the process being marked :s content and construct validity) and also on the way in which groups of ıbjects are sampled. Thus, collection of biologic specimens and attendant :posure assessment is inherently a statistical process. Sampling designs must :commodate the distribution of possible exposures as well as the purpose the study.

Two considerations are important in addressing kinetic conditions that affect an exposure–disease continuum. First, the variations of exposure from interval to interval must be translated efficiently into burden and damage; that is, there should be no physiologic damping of exposure variability. Second, during the period of intense exposure, the relationship between intense burden and damage must be nonlinear (Rappaport, 1991). Practically speaking, with respect to chronic diseases, the sampling strategy that generally will be useful is one that involves assessment of mean exposure with little regard for short-term fluctuations of exposure. However, if acute effects are the subject of investigation, it may be appropriate to estimate short-term peak exposures (Rappaport, 1991).

Thus, the proper selection and use of biologic markers depends on an understanding of the underlying pharmacokinetics and pharmacodynamics (World Health Organization (WHO), 1986; Andersen, 1987; Smith, 1987). Knowledge of pharmacokinetics is important to determine the frequency of sampling and the tissues or fluids that are most appropriate for study. It also guides the interpretation of the dosimetry and effect data obtained in the target tissue or in a surrogate.

Temporal Relevance

The temporal relationship of markers to external exposure or to disease end points must be clear. Markers have different half-lives in biologic media. Droz *et al.* (1991) have characterized the extent to which exposure history can be represented by a biologic marker. Droz *et al.* (1985) had demonstrated how the measured dose of organic solvents in workers was influenced by the following variables: interday and intraday fluctuation of exposure, physical workload, body build, and metabolism. Individuals classified as having the same exposure still can have a different dose because of the influence of these variables. The timing of marker measurement in relation to exposure influences the ability to detect response; measurements made too early or too late may underestimate the magnitude of the response (Wilcosky, 1992). For example, transient markers such as the influx of neutrophils or eosinophils into the respiratory tract will occur usually during the first 3–7 days of an inflammatory response. Hence, measurement of these markers of inflammation in bronchoalveolar lavage fluid immediately after exposure would underestimate the inflammatory response (Wilcosky, 1993). Whether an exposure marker reflects recent or cumulative exposures, peaks or averages, will depend on the pharmacokinetics of the toxicant and the persistence of the marker in the biologic sample being assayed (which, in large part, is a function of the turnover rate or half-life of the sample, but also is a reflection of repair rates).

An understanding of temporal relevance is essential to developing moni-

toring strategies and to interpreting results. Most measures of internal dose, for example, reflect recent exposures (ranging, for example, from hours for cotinine to days for benzene in exhaled air). An exception would be a substance that is fat soluble and is stored in adipose tissue. Hemoglobin is a good integrating dosimeter over the 4-month life span of the erythrocyte which, unlike lymphocytes, lacks repair systems. Human serum albumin has a 20- to 25-day half-life. Since it is synthesized in the liver where many carcinogens are metabolically activated, albumin might collect adducts not detected in hemoglobin (Perera, 1987).

The time period reflected by total white blood cell, or lymphocyte, markers is considerably more complex than that reflected by albumin (Perera, 1987; Carrano and Natarajan, 1988). For example, adducts on DNA from lymphocytes can reflect past as well as current exposure, since a subset (T-cells) is very long lived. A review of lymphocyte subpopulations shows that measurements of DNA adducts in these cells will be influenced by the longer lived T cells and will, therefore, reflect exposures that occurred both recently and several decades in the past. T cells make up 60–90% of lymphocytes, which in turn represent 22–28% of peripheral lymphocytes in circulating blood. Thus they constitute a maximum 25% of lymphocytes. The estimated half-life of T cells is 3 years. In contrast, B cells and monocytes constitute 1–2% and 1–7% of circulating lymphocytes, respectively, and have lifetimes ranging from days to weeks. Granulocytes represent the remaining 66–85% of lymphocytes and are short lived (hours to days). Thus, considering only DNA adducts in cells damaged while in circulation, and excluding consideration of adducts in circulating lymphocytes that might result from damage to stem or precursor cells in the bone marrow, when all DNA from a sample of peripheral blood is assayed for DNA adducts, the long-lived T cells will be the major contributors in cases of past discontinued exposure, largely because of their 100–1000 times longer life-span. In addition, lymphocytes have a lower repair activity than do cycling cells; thus, DNA lesions are likely to be more persistent in lymphocytes. In cases of current, recent, or long-term uninterrupted exposure to carcinogens, T cells will contribute less significantly to total adduct measurements. The preponderance of adducts will be measured in the shorter-lived granulocytes, B cells, and monocytes.

In retrospective case-control studies, a permanent marker left decades earlier by the initiating carcinogen would be ideal. Unfortunately, in the case of discontinued past exposures, even the most long-lived markers will be diluted by cell turnover, thereby underestimating the true past or cumulative dose. Only if exposure was continuous and had not changed significantly during the past decades (and only if the disease did not alter metabolism) would current levels of the marker directly represent critical prior exposure. At the very least, however, markers such as adducts reflect individual responsiveness to carcinogenic exposures, provided that metabolism is unchanged

by disease. Many other exposure patterns are relevant to case-control and cohort studies (current but interrupted, continuous but of varying magnitude, etc.). Each pattern leads to a different distribution of adducts among lymphocyte populations and, hence, to a varying pattern of persistence.

Understanding "Background" Variability and Confounding Variables

The variability of a biologic marker is a statistical characteristic of groups of subjects or populations (Janetos, 1988). Variability is the result of genetic and environmental factors, separately and interacting. Motulsky (1988) has described the pervasiveness of genetic variation that accounts for human individuality.

> Human physiognomy is unique and no two human beings except identical twins are alike. The involved genes remain unknown. Remarkable genetic individuality also exists for red cell and tissue cell (HLA) groups, in enzymes and proteins. Enzyme variation usually is associated with variable enzyme levels in the normal range, so a person's exact activity level for a given enzyme (i.e., high normal, average, low normal) may be the speed of breakdown of various substances. Protein variation may lead to differential binding of foreign substances. . . . Variability in receptor activity may cause differential metabolism of foreign or endogenous ligands. Various HLA and related markers may lead to differential immunologic reactions that predispose to certain autoimmune diseases. Variability at the DNA level is more striking. Frequent differences occur at the individual nucleotide level (every 500 nucleotides), as do size variations of longer stretches of DNA (minisatellites). Most such DNA variants are phenotypically silent but often can be used as "markers" for closely linked gene loci that specify proteins that have physiologic, biochemical, or immunologic effects (Motulsky, 1988).

This natural variability makes it essential to know the range of biomarker values in a "normal" population. Care must be taken not to be deceived by the extensive variation in biochemical individuality (a "healthy" level for some individuals may indicate a health risk for others) (Schulte, 1987). The range of normal can be quite extensive. For example, it is well known that the cholinesterase level in subjects not exposed to organophosphorus insecticides covers a wide range—a 25% change in the group may overshadow a decrease of 50% in a few subjects (WHO, 1975).

Interindividual and intraindividual variations are important contributors to "noise" or "background" in biomonitoring or epidemiologic studies and should be characterized prior to large-scale application of a particular biomarker. Such data can, however, be generated only through large-scale surveys that employ repeated sampling and efforts to control for confounding variables. Thus, a background study is a significant epidemiologic exercise in itself. With respect to carcinogen–DNA and –protein adducts, significant interindividual and intraindividual variation has been observed with polyaromatic hydrocarbons (PAHs), 4-aminobiphenyl (4–ABP), and nitrosamines (Harris, 1985; Umbenhauer *et al.*, 1985; Bryant *et al.*, 1987; Perera,

1987). SCEs also vary significantly within and among individuals (Carrano and Natarajan, 1988).

As discussed earlier, since biologic markers can be potentially more sensitive than indicators used in conventional epidemiologic methods, there is a greater need to control for confounding or mitigating factors [National Academy of Sciences (NAS), 1991]. Confounding variables in studies using biomarkers may include age, sex, race, cigarette smoking, alcohol consumption, diet, drugs or other environmental exposures, genetic factors, or preexisting health impairment.

Age, a potential confounding factor of exposure–effect associations, often is addressed in a facile manner. Increasing evidence shows that chronologic age is not an adequate descriptor of biologic age (Ingram, 1988). Biologic age may be more of an influence on marker frequency than chronologic age. However, the definition and measurement of aging is controversial. Efforts to identify and develop biologic markers of aging are underway but are in the early stages of progress (Hart and Turtarro, 1988). Molecular epidemiologic studies may be influenced significantly by the fact that DNA alterations (DNA chemical structure, DNA sequence organization, and gene expression) increase with age (Mullaart *et al.*, 1990). DNA repair also has been shown to correlate well with lifespan in various mammalian species. This correlation has been cited as an indication for the role of DNA repair in the aging process. Any conclusion is complicated by the large interindividual differences in DNA repair. (See Figure 3.7 for a schematic representation of various types of DNA damage).

A more vigorous appraisal of age beyond a chronologic one may be important to conducting effective molecular epidemiologic studies. Biomarkers of aging may be susceptibility markers or may be covariates in exposure–response studies. In designing molecular epidemiologic studies and validating a biologic marker, attention may need to be paid to the comparability of populations in terms of biologic and functional age as well as chronologic age.

Race may be another potentially confounding variable that generally is controlled in epidemiologic studies, but only categorically. Within apparently homogeneous racial categories may exist subcategories distinguished by molecular markers or DNA repair capacities (Weston *et al.*, 1991). For molecular epidemiologic research, classification by race may result in a net loss of information, since such categorization may be misleading and fail to define large categories of interperson variability (Cooper and David, 1986).

Many other potentially confounding factors also need consideration in molecular epidemiologic studies but often are neglected. Most reports on cytogenetic studies have provided no information on smoking and other exposures to carcinogens or mutagens, yet life-style factors (e.g., smoking, diet) and other chemical exposures (e.g., environmental, recreational, medicinal, and drug-related) are potential confounding factors with respect to SCEs

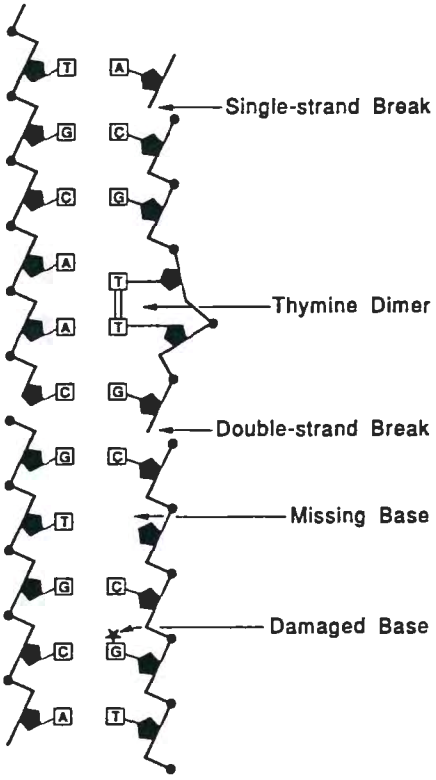


FIGURE 3.7 Schematic of DNA damage. (Reprinted with permission from Baan, 1987. Copyright © 1987 by Springer-Verlag.)

and most other markers. Other potential confounders are host factors (e.g., health and immune status) and exposures that influence the marker of interest. In a study of PAH–DNA adducts in workers, for example, it is necessary to account for all "background" exposures to the chemical and, ideally for factors that could induce or inhibit metabolism and binding of benzo(a)pyrene (BaP). These factors include inducers such as cigarette smoke, charbroiled meat, ethanol, sedatives, and PCBs, as well as inhibitors such as methylxanthines in foods, steroids, solvents, and spray paints. Thus, even pilot studies become fairly complex epidemiologic exercises that require careful interviewing (Perera *et al.*, 1982; NAS, 1991). For example, Figure 3.8 shows the magnitude of a molecular epidemiologic study of 33 mortuary science students evaluated for various biomarkers before and after exposure to formaldehyde during an embalming course. Despite the small number of subjects, many data points were collected (Suruda *et al.*, 1992).

Since most biomarkers are nonspecific, that is, different exposures may

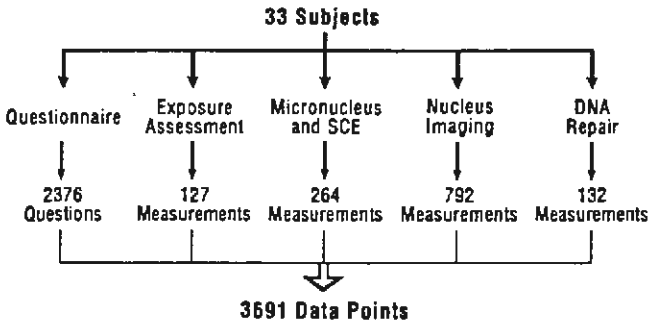


FIGURE 3.8 Magnitude of a molecular epidemiologic study.

cause the same marker response, attention should be paid to the impact of their use in studies. Nonspecific biomarkers may lead to dilution and confounding of exposure–response associations in the presence of other exposures (Weiss and Liff, 1983; Wilcosky, 1993). If two different exposures, E_1 and E_2, cause the same marker response through independent pathways, they increase the overall marker response rate in an additive manner. However, the assessment of the different exposures that cause the same response through independent pathways will be misleading in epidemiologic studies that use relative risk or odds ratio because these measures are based on the assumption of a multiplicative model of association, whereas the two exposures working independently increase overall response in an additive manner. Thus, the relative risk of response of one agent will be influenced by the background response of the other agent. Therefore, the use of risk difference, rather than relative risk, to compare responses will help avoid the problems of dilution from a high background incidence from one exposure (Wilcosky, 1993).

Reproducibility, Sensitivity, Specificity, and Predictive Value of Assays

Assays should be reproducible with limited variability that is attributable to laboratory personnel or assay method (Gann *et al.*, 1985). The same criteria of adequate specificity, sensitivity, and predictive value that apply to the validation of screening methods should be met by biomarkers. Markers of exposure should be "sensitive" and "specific" to toxic exposures, picking up a high percentage of individuals in the exposed group and attributing negative results to a high percentage of unexposed persons. Given interindividual variation, however, not all exposed persons would be expected to be positive. Markers of effect or response should detect a high number of individuals at elevated risk of adverse outcomes. Both types of markers should give a high proportion of "correct" answers.

Summary

Validation of candidate biomarkers for molecular epidemiologic studies is an empirical process that can be approached by producing several different, but convergent, lines of evidence. Key to future efforts in the use of a particular biomarker is the establishment of content, construct, and criterion validity, and attention to factors that can cause confounding and variability in an exposure–response association. To validate a biomarker adequately it is necessary to understand its natural history and temporal relevance to the event being marked. Epidemiologic methods for validation are discussed in Chapter 6.

Acknowledgment

The authors thank Robert F. Vogt, Jr., for the discussion of the ROC curves.

References

Andersen, M. E. (1987). Tissue dosimetry in risk assessment, or what's the problem here anyway? *In* "Pharmacokinetics in Risk Assessment," pp. 8–23. National Academy Press, Washington, D.C.

Ashby, J. (1987). Monitoring human exposure to genetoxic chemicals. *In* "Occupational and Environmental Chemical Hazards: Cellular and Biochemical Indices for Monitoring Toxicity (V. Foa, E. A. Emmett, M. Maroni, A. Colombi, eds.), pp. 411–423. Ellis Horwood, Chichester, England.

Baan, R. A. (1987). DNA damage and cytogenetic endpoint. *In* "Cytogenetics" (G. Obe and A. Basler, eds.), pp. 327–337. Springer-Verlag, Berlin.

Brown, S. L. (1988). Differential susceptibility: Implications for epidemiology, risk assessment, and public policy. *In* "Phenotypic Variation in Populations: Relevance to Risk Assessment" (A. D. Woodhead, M. A. Bender, and R. C. Leonard, eds.), pp. 255–269. Plenum Press, New York.

Bryant, M. S., Skipper P. L., Tannenbaum, S. R., and Maclure, K. M. (1987). Hemoglobin adducts of 4-aminobiphenyl in smokers and nonsmokers. *Cancer Res.* 47, 602–608.

Bull, R. J. (1989). Decision model for the development of biomarkers of exposure. U.S. Environmental Protection Agency (EPA 600/X-89/163), Las Vegas, Nevada, p. 12.

Carrano, A. V., and Natarajan, A. T. (1988). Considerations for population monitoring using cytogenetic techniques. *Mutat. Res.* 204, 379–406.

Cole, P., and MacMahon, B. (1971). Attributable risk percent in case-control studies. *Br. J. Prev. Soc. Med.* 25, 242–244.

Cook, T. D., and Campbell, D. T. (1979). "Quasi-experimentation: Design and Analyses Issues for Field Settings." Rand McNally, Chicago.

Cooper, R., and David, R. (1986). The biological concept of race and its application to public health and epidemiology. *J. Public Health. Policy Law* 11, 97–115.

Crump, K. S., and Allen, B. C. (1985). Methods for quantitative risk assessment using occupational studies. *Am. Stat.* 39, 442–450.

Droz, P. O., Berode, M., and Wu, M. M. (1991). Evaluation of concomitant biological and air monitoring results. *Appl. Occup. Environ. Hyg.* 6, 465–474.

Dwyer, J. H. Feinleib, M., Lippert, P., and Hoffmeister, H. (eds.) (1992). "Statistical Models for Longitudinal Studies of Health." Oxford University Press, New York.

England, W. L. (1988). An exponential model used for optimal threshold selection on ROC curves. *Med. Decis. Making* 8(2), 120–131.

Erdreich, L. S. (1988). Combining animal and human data: Resolving conflicts, summarizing the evidence. *In* "Epidemiology and Health Risk Assessment" (L. Gordis, ed.), pp. 197–207. Oxford University Press, New York.

Gann, P. H., Davis, D. L. and Perera, F. (1985). "Biological Markers in Environmental Epidemiology: Constraints and Opportunities." Paper presented at the 5th Workshop of the Scientific Group on Methodologies for the Safety and Evaluation of Chemicals, World Health Organization, 12–16 August, Mexico City.

Griffith, J., Duncan, R. C., and Hulka, B. S. (1989). Biochemical and biological markers: Implications for epidemiologic studies. *Arch. Environ. Health* 44, 375–381.

Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: The state of the art. *Crit. Rev. Diagn. Imaging* 29(3), 307–335.

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1), 29–36.

Hannequin, P., Liehn, J. C., Delisle, M. J., Deltour, G., and Valeyre, J. (1987). ROC analysis in radioimmunoassay: An application to the interpretation of thyroglobulin measurement in the follow-up of thyroid carcinoma. *Eur. J. Nucl. Med.* 13(4), 203–206.

Harris, C. C. (1985). Future directions in the use of DNA adducts as internal dosimeters for monitoring human exposure to environmental mutagens and carcinogens. *Environ. Health Perspect.* 62, 185–191.

Hart, R. W., and Turtarro, A. (1988). Biomarkers in aging and toxicity. *Environ. Gerontol.* 23, 241–43.

Henkelman, R. M., Kay, I., and Bronskill, M. J. (1990). Receiver operating characteristic (ROC) analysis without truth. *Med. Decis. Making* 10(1), 24–29.

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, A. Columbi, eds.), pp. 121–241. Ellis Horwood, Chichester, England.

Hogue, C. J. R., and Brewster, M. A. (1988). Developmental risks: Epidemiologic advances in health risk assessment. *In* "Epidemiology and Health Risk Assessment" (L. Gordis, ed.), pp. 61–80. Oxford University Press, New York.

Hulka, B. D. (1990). Principles of bladder cancer screening in an intervention trial. *J. Occup. Med.* 32, 812–816.

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

Hunink, M. G., Richardson, D. K., Doubilet, P. M., and Begg, C B. (1990). Testing for fetal pulmonary maturity: ROC analysis involving covariates, verification bias, and combination testing. *Med. Decis. Making* 10(3), 201–211.

Ingram, D. K. (1988). Key questions in developing biomarkers of aging. *Environ. Gerontol.* 23, 429–434.

International Programme on Chemical Safety (1983). Guidelines on studies in environmental epidemiology. *In* "Environmental Health Criteria" vol. 27, pp. 133–136. World Health Organization, Geneva.

Janetos, A. C. (1988). Biological variability. *In* "Variations in Susceptibility to Inhaled Pollutants" (J. D. Brain, B. D. Beck, A. J. Warren, and R. A. Shaikh, eds.), pp. 9–29. Johns Hopkins University Press, Baltimore.

Last, J. M. (ed.) (1988). "A Dictionary of Epidemiology." 2nd Edition, Oxford University Press, New York.

Lawrence, C. E., and Taylor, P. R. (1985). Empirical estimation of exposure in retrospective epidemiologic studies. *In* "Environmental Epidemiology" (F. C. Kopfler and G. F. Craun, eds.), pp. 239–246. Lewis Publishers, Chelsea, Michigan.

Louis, T. A. (1988). General methods for analyzing repeated measure. *Stat. Med.* 7, 39–45.

McClish, D. K. (1987). Comparing the areas under more than two independent ROC curves. *Med. Decis. Making* 7(3), 149–155.

McClish, D. K. (1989). Analyzing a portion of the ROC curve. *Med. Decis. Making* 9(3), 190–195.

Meistrell, M. L. (1990). Evaluation of neural network performance by receiver operating characteristic (ROC) analysis: Examples from the biotechnology domain. *Comput. Meth. Prog. Biomed.* 32(1), 73–80.

Metz, C. E., and Shen, J. H. (1992). Gains in accuracy from replicated readings of diagnostic images: Prediction and assessment in terms of ROC analysis. *Med. Decis. Making* 9(3), 190–195.

Miettinen, O. S. (1974). Proportion of disease caused or prevented by a given exposure, trait, or intervention. *Am. J. Epidemiol.* 99, 325–332.

Motulsky, A. G. (1988). Human genetic individuality and risk assessment. *In* "Phenotypic Variation in Populations: Relevance to Risk Assessment" (A. D. Woodhead, M. A. Bender, and R. C. Leonard, eds.), pp. 7–9. Plenum Press, New York.

Mullaart, E., Lohman, P. H. M., Berends, F., and Vijg, J. C. (1990). DNA damage, metabolism, and aging. *Mutat. Res.* 237, 189–210.

National Academy of Sciences (1991). "Human Exposure Assessment for Airborne Pollutants." National Academy Press, Washington, D.C.

Nunnaly, J. C. (1967). "Psychometric Theory." McGraw-Hill, New York.

Ozonoff, D. M., and Wartenberg, D. (1991). Toxic exposures in a community setting: The epidemiologic approach. *In* "Molecular Dosimetry and Human Cancer: Analytical, Epidemiological, and Social Considerations" (J. D. Groopman and P. L. Skipper, eds.), pp. 77–88. CRC Press, Boca Raton, Florida.

Perera, F. P. (1987). Biological markers in risk assessment. *Environ. Health Perspect.* 76, 141–146.

Perera, F. P. (1988). The significance of DNA and protein adducts in human biomonitoring studies. *Mutat. Res.* 205, 255–269.

Perera, F. P., Poirier, M. C., Yuspa, S. H., Nakayama, J., Jaretzki, A., Curnen, M. M., Knowles, D. M., and Weinstein, I. B. (1982). A pilot project in molecular cancer epidemiology: Determination of benzo[*a*]pyrene–DNA adducts in animal and human tissues by immunoassays. *Carcinogenesis* 3, 1405–1410.

Perera, F. P., Hemminki, K., Young, T. L., Brenner, D., Kelly, G., and Santella, R. M. (1988). Detection of polycyclic aromatic hydrocarbon–DNA adducts in white blood cells of foundry workers. *Cancer Res.* 48, 2288–2291.

Perera, F., Motzer, R. J., Tang, D., Reed, E., Parker, R., Warburton, D., O'Neill, P., Albertini, R., Bigbee, W. L., Jensen, R. H., Santella, R., Tsai, W. Y., Simon-Cereijido, G., Randall, C., and Bosl, G. (1991). Multiple biologic markers in germ cell tumor patients treated with platinum-based chemotherapy. *Cancer Res.* 52, 3558–3565.

Rappaport, S. M. (1988). Biological considerations for designing sampling strategies. *In* "Advances in Air Sampling Strategies" (W. John, ed.), pp. 337–352. Lewis Publishers, Chelsea, Michigan.

Rappaport, S. M. (1991). Assessment of long-term exposures to toxic substances in air. *Ann. Occup. Hyg.* 35, 61–121.

Rogan, W. J. (1988). Relation of surrogate measures to measures of exposure. *In* "Epidemiology and Health Risk Assessment" (L. Gordis, ed.), pp. 148–158. Oxford University Press, New York.

Rosner, B., Munoz, A., Tager, I., Speizer, R., and Weiss, S. (1985). The use of an autoregressive model for the analysis of longitudinal data in epidemiologic studies. *Stat. Med.* **4**, 457–467.

Rothman, K. J. (1986). "Modern Epidemiology." Little, Brown, Boston.

Schatzkin, A., Freedman, L. S., Schiffman, M. H., and Dawsey, S. J. (1990). Validation of intermediate endpoints in cancer research. *J. Natl. Cancer Inst.* **82**, 1746–1752.

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

Schulte, P. A. (1989). A conceptual framework for the validation and use of biological markers. *Environ. Res.* **48**, 129–144.

Schulte, P. A., and Mazzuckelli, L. F. (1991). Validation of biological markers for quantitive risk assessment. *Environ. Health Perspect.* **90**, 239–246.

Schulte, P. A., Boeniger, M., Walker J. T., Schober, S. E., Pereira, M. A., Gulati, D. K., Wojciechowski, J. P., Garza, A., Froelich, R., Strauss, G., Halperin, W. E., Herrick, R., Griffith, J. (1992). Biological markers in hospital workers exposed to low levels of ethylene oxide. *Mutat. Res.* **278**, 237–251.

Slaga, T. J. (1988). Interspecies comparisons of tissue DNA damage, repair, fixation, and replication. *Environ. Health Perspect.* **77**, 73–82.

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

Somoza, E., and Mossman, D. (1991). Biological markers and psychiatric diagnosis: Risk–benefit balancing using ROC analysis. *Biol. Psychiatry* **29**(8), 811–826.

Stevens, D. K., Bull, R. J., Nauman, C. H., and Blancato, J. N. (1991). Decision model for biomarkers of exposure. *Regulatory Toxicol. Pharmicol.* **14**, 286–296.

Suruda, A., Schulte, P., Boeniger, M., Hayes, R. B., Livingston, G. H., Steenland, K., Stewart, P., Herrick, R., Douthitt, D., and Fingerhut, M. A. (1993). Cytogenetic effects of formaldehyde exposure in students of mortuary science (submitted).

Swenberg, J. (1988). Banbury Center DNA Adduct Workshop (Commentary). *Mutat. Res.* **203**, 55–68.

Swets, J. A., Pickett, R. M., Whitehead, S. F., Getty, D. J., Schnur, J. A., Swets, J. B., and Freeman, B. A. (1979). Assessment of diagnostic technologies. *Science* **205**, 753–759.

Thompson, M. L., and Zucchini, W. (1989). On the statistical analysis of ROC curves. *Stat. Med.* **8**(10), 1277–1290.

Tosteson, A. N., and Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Med. Decis. Making* **8**(3), 204–215.

Umbenhauer, D., Wild, C. P., Montesano, R., Saffhill, R., Boyle, J. M., Hahn, K. Y., Thomale, J., Rajewsky, M. I., and Lu, S. H. (1985). O-6 Methyldeoxyguanosine in oesophageal DNA among individuals at high risk of oesophageal cancer. *Int. J. Cancer* **36**, 661–65.

Weiss, N. S., and Liff, J. M. (1983). Accounting for the multicausal nature of disease in the design and analysis of epidemiologic studies. *Am. J. Epidemiol.* **117**, 14–18.

Weston, A., Vineis, P., Caporaso, N. E., Krontiris, T. G., Lonergan, J. A., and Sugimura, H. (1991). Racial variations in the distribution of Ha-ras-1 alleles. *Mol. Carcin.* **4**, 265–268.

Wilcosky, T. C. (1993). Biological markers of intermediate outcomes in studies of indoor air and other complex mixtures, Health Effects Institute research report (*HEI*) (in press).

World Health Organization (1975). Early Detection of Health Impairment in Occupational Exposure to Health Hazards. WHO Technical Report No. 571. World Health Organization, Geneva.

World Health Organization (1986). Principles of toxicokinetic studies. *In* "Environmental Health Criteria" vol. 57, World Health Organization, Geneva.

Zeller, R. A., and Carmines, E. G. (1980). "Measurement in the Social Sciences." Cambridge University Press, Cambridge.