Validation of Biological Markers for Quantitative Risk Assessment

by Paul Schulte* and Lawrence F. Mazzuchelli*

The evaluation of biological markers is recognized as necessary to the future of toxicology, epidemiology, and quantitative risk assessment. For biological markers to become widely accepted, their validity must be ascertained. This paper explores the range of considerations that compose the concept of validity as it applies to the evaluation of biological markers. Three broad categories of validity (measurement, internal study, and external) are discussed in the context of evaluating data for use in quantitative risk assessment. Particular attention is given to the importance of measurement validity in the consideration of whether to use biological markers in epidemiologic studies. The concepts developed in this presentation are applied to examples derived from the occupational environment. In the first example, measurement of bromine release as a marker of ethylene dibromide toxicity is shown to be of limited use in constructing an accurate quantitative assessment of the risk of developing cancer as a result of long-term, low-level exposure. This example is compared to data obtained from studies of ethylene oxide, in which hemoglobin alkylation is shown to be a valid marker of both exposure and effect.

Introduction

It is generally accepted that valid biological markers can make an important contribution to toxicologic and epidemiologic research, and ultimately, to quantitative risk assessment (1-3). While obeisance is paid to the concept of validity, little attention has been given to what it means and how to evaluate it. The objective of this paper is to identify and explore the range of the concept of validity and to address how considerations that comprise the concept of validity and to address how validity pertains to the use of biological markers in quantitative risk assessment. The term “biological marker” has been defined as an indicator that signals events in biological systems or samples, and it is generally taken to be any biochemical, genetic, or immunologic indicator that can be measured in a biological specimen (4-7). Ascribed to the term biological marker is its role as an indicator of events in a continuum between exposure to a xenobiotic substance and resultant disease (4,6). Biologic markers can refer to any of three categories of events: exposure, effect, and susceptibility. Unless otherwise specified, in this discussion, a marker is considered to relate to an event in the exposure-disease continuum without further reference as to whether that event is exposure, effect, or susceptibility. Biological markers can contribute to quantitative risk assessment by helping to: determine the forms of dose-time-response relationships; assess the biologically effective dose; make interspecies comparison of effective dose, relative potency, and effects; resolve the quantitative relationships between human interindividual variability in susceptibility; and identify subpopulations that are at enhanced risk (2,8).

Three broad categories of validity can be distinguished: measurement validity, internal study validity, and external validity. Measurement validity has been defined as an expression of the degree to which “…a measurement measures what it purports to measure” (9). Internal study validity is the degree to which inferences drawn from a sample are warranted when account is taken of the study methods, the representativeness of the study sample, and the nature of the population from which the sample is drawn (9). External study validity is the extent to which the findings of a study can be generalized to other populations (9).

Biological markers and the studies that include them need to be shown to have measurement, internal, and external validity before they can be accurately used in 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 (3). Risk assessments based on studies that underestimate a true effect can lead to regulations that contain exposure limits 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 containing 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 depend on the methodology of the study, including the measurements and other design factors that lead to the results. Invalid measurements, inferences, or generalizations may lead to erroneous risk assessments. In this paper, the three categories of validity are discussed in terms of how they apply to biological markers for research and quantitative risk assessment. These theoretical considerations of validity are illustrated by examples of risk assessments involving ethylene dibromide and ethylene oxide.

*National Institute for Occupational Safety and Health, Cincinnati, OH 45226. Address reprint requests to P. A. Schulte, Industrywide Studies Branch, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, OH 45226.
Measurement Validity

Measurements are one of the principal building blocks of quantitative risk assessment. If measurements are invalid, it is likely that the risk assessments constructed from those measurements will also be invalid. Measurement validity characterizes the extent to which a marker of a phenomenon has content validity (i.e., pertains to the underlying phenomenon); construct validity (i.e., correlates with other relevant characteristics of the underlying phenomenon); and criterion validity (i.e., predicts some component 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 (10).

Content Validity

Content validity is the extent to which a marker "incorporates the domain of the phenomenon under study" (9). 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 its relationship to the natural history of the disease in question must be unambiguous. For example, a DNA adduct of benzo(a)pyrene (BaP) will have content validity as a marker of exposure in a study of BaP-induced lung cancer, since the involvement of DNA in BaP-induced carcinogenesis is well documented. In contrast, the development of DNA adducts in the N7 position might not have content validity as a marker of biologically effective dose if the O6 methylguanine adduct is shown to be that which is most clearly related to the carcinogenic process. However, the N7 adducts might be reasonably valid markers of BaP biologically effective dose if the production of O6 and N7 adducts are directly proportional (as would be expected if they were produced by the same activated BaP metabolite), and 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 0 time).

To properly assess content validity, one must consider the extent to which the marker pertains to the phenomenon (exposure, effect) of interest or, the extent to which the marker represents a relevant feature of that phenomenon. For example, if it were assumed that hydroxymethyl histidine adducts of hemoglobin were markers of the internal dose of ethylene oxide, that marker would lack complete content validity since hydroxymethyl histidine adducts of hemoglobin can result from exposure to other substances that contain ethyl groups. Furthermore, populations with no known exposure to ethylene oxide have been shown to form hydroxymethyl histidine adducts of hemoglobin. Without considering content validity, one might reach erroneous conclusions if it were assumed that only ethylene oxide exposure was responsible for the observed adducts. Valid measures might be developed by subtracting the amount of adducts attributable to factors other than the exposure under study from the total amounts of adducts formed. This requires the evaluation of a nonexposed comparison group.

Because content validity is assessed by professional judgment, there are no universally accepted criteria for its determination (11). However, it is possible to strengthen determinations of content validity if judgments are made by a group of experts. 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 where it is most needed, i.e., where there is an incomplete understanding of the domain of underlying characteristics of the exposure-disease process.

Construct Validity

Construct validity describes the extent to which a marker corresponds to other relevant characteristics of the underlying phenomenon, that is, the theoretical concepts or constructs concerning the phenomenon under study (9). This correspondence is exhibited in part by association of the subject marker with other markers or variables of the phenomenon (12,13). For example, if the characteristics of a phenomenon change with age, a marker with construct validity will change accordingly (9). Furthermore, if there are no associations with other variables that would reasonably be expected to be linked with the phenomenon under study, then the marker may be of questionable relevance in a study or subsequent risk assessment.

Construct validity is sometimes difficult to distinguish from content validity when describing biological markers, but it should be evaluated whenever general understanding of the underlying phenomenon is not clear. Hence, if a marker is a candidate for inclusion in a study of an exposure or outcome, and the actual role of the marker in the exposure-outcome continuum has not been established (that is, its content validity has not been established), it still may be useful as a covariate if it can be shown to have construct validity.

Criterion Validity

Criterion validity describes the extent to which a marker correlates with the phenomenon being studied (9). For example, the criterion validity of a marker of disease is the extent to which people who have the marker already have or will develop the disease. The criterion is what is being marked or indicated by the marker; generally this is a disease, but it could also be an exposure.

Two aspects of criterion validity have been distinguished, concurrent validity and predictive validity (9). When a marker and its criterion refer to the same point in time, they have concurrent validity. For example, a biological marker of exposure, such as a hemoglobin adduct, is validated against a determination of a DNA adduct in a target organ (if they occur simultaneously). For markers of exposure, concurrent validity is satisfied by understanding the stoichiometric relationship between the exposure and the internal or biologically effective dose. For markers of effect, concurrent validity is satisfied by a strong correlation between the marker and the disease or dysfunction of interest. Concurrent validity is usually determined in cross-sectional studies.

Predictive validity refers to a marker's ability to predict the criterion (9). For example, a marker of altered structure and function, such as abnormal sputum cytology, could be validated against subsequent diagnostic confirmation of lung cancer. Predictive validation requires obtaining samples of subjects,
measuring some marker, waiting the necessary time for the effec- 
t (criterion) to occur, then assessing the observed correlation 
(10,14). Other factors to consider in predictive validation might 
include intervening or modifying characteristics that could in- 
fluence the occurrence of the end point and stochastic effects in 
the development of the outcome criterion. In general, the degree 
of predictive validity depends on the extent of the correlation be- 
tween the marker and the criterion. Predictive validity applies to 
makers of exposure, effect, or susceptibility.

Predictive validation is performed using a longitudinal (pro- 
spective) study design. Since one of the drawbacks in assessing 
predictive validity is the potentially long time course necessary 
for the development of the criterion, there are time-compressing 
study designs that are useful. One is the contemporaneous case- 
control design and another is the retrospective case-control 
design; both are limited in their ability to assess marker validity.

A contemporaneous case-control design involves obtaining 
samples of individuals with and without the criterion of interest 
(i.e., a disease) then assessing those individuals for the presence 
of a marker. The retrospective case-control design involves select- 
ing individuals with and without a disease (the criterion) and 
then attempting to identify marker status prior to the appearance 
of the disease or study end date. Clearly, these approaches are 
limited. A contemporaneous case-control study, using markers 
of exposure, will not provide an unambiguous answer concerning 
predictability if it is difficult to tell whether the marker predicts 
the criterion disease or is merely the result of it. The retrospective 
 case-control study is difficult to perform because it is not easy to 
find historic information on the presence of many markers.

It is possible to judge the criterion validity of a marker in terms 
of its sensitivity, specificity, and predictive value. Griffith et al. 
(15) have distinguished the terms sensitivity and specificity as 
they—refer to laboratory methods to detect a marker and as they 
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 detec-
tion system to respond in the presence of the marker. Population sen-
sitivity, 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 that are negative for both the marker 
and the event, to the number of subjects that are negative for the event 
(15).

Griffith et al. also identified two study designs that are useful 
for determining population sensitivity and specificity (15): The 
first is based on two independent samples of fixed size. In this 
design, the health status or exposure status of each subject is 
ascertained and observations are collected until the pre-set sam-
ple sizes are reached in each group. Neither the marker frequen-
cy nor the disease frequency play a role. The data might be col-
lected as subjects are identified or in a case-control study from 
medical records. Also, archived biological samples might be us-
ed. The second approach is to select a single sample of fixed size 
from the population of interest, and to distribute the subjects into 
a four-fold table according to the presence or absence of the 
marker, and the presence or absence of the exposure or disease. 
Sensitivity is then estimated as the ratio of the number of subjects 
positive for both the marker and the disease to the number of 
subjects with the disease. Specificity is estimated is the ratio of 
subjects negative for both the marker and the disease to the sub-
jects negative for the disease.

The best way of appraising criterion validity is to compare a 
marker with a criterion selected as the true characteristic or as 
the “gold standard” (12,16). This is exemplified by efforts to 
determine the validity of a new procedure for determining 
whether malignant or premalignant bladder cells can be detected 
by assessing DNA hyperploidy (17). If DNA hyperploidy is a 
valid marker of bladder cancer, hyperploidy should occur prior 
to visible morphological change, which is routinely evaluated by 
Papanicolaou cytology. Therefore, appropriate validation of the 
marker is not against the cytology, but against a positive bladder 
biopsy (the gold standard) some time in the future.

In epidemiologic studies, markers that are invalid measures of 
a phenomenon can result in misclassification of exposure, effect, 
or susceptibility. As Hogue and Brewster (3) observed, “An ex-
posure variable may be misclassified if the marker of exposure 
s has a sensitivity or specificity less than 1.0. That is, someone who 
is truly exposed is classified as being not exposed, or someone 
who is truly not exposed is classified as being exposed.” If, for 
example, a marker of biologically effective dose is the basis for 
exposure classification, misclassification will occur if that 
marker does not correspond to the actual amount of xenobiotic 
that interacts with critical macromolecules. This could occur 
with certain DNA adducts if the amount that persists is affected 
by the repair rates and if the repair rate varies among individuals.

In summary, the quality of risk assessments depends on the 
quality and validity of measurements. As Matanoski (18) observ-
ed, “If epidemiologists are to address problems faced by risk 
assessors, they must design studies, measure exposures and 
analyze results with a considered view of this specific use. This 
will require new perspectives on the measurement of exposures 
such as biomarkers and better methods for estimating 
exposures.” With regard to the design of studies, there is a need 
to use valid markers if the studies are to be of value in risk 
assessments.

The Office of Technology Assessment (19) also recognized this 
problem:

It is generally not possible to gather reliable information about a 
population and concurrently gather validating information about a 
marker used to measure outcome, unless another marker with known 
validity, and a known relationship to the new marker is also used in 
the study. Even though that is technically feasible, it is probably not 
an efficient way to gather validating data. (19)

Reliability

Marker validity is also dependant on reliability; that is, the 
degree to which a marker will be a valid representative or predi-
tor of an event is influenced by the reliability with which it can 
be measured. Reliability encompasses the unsystematic, random 
variation observed upon repeated measurements (9,22). In the 
measurement of continuous variables, such as with most 
biological markers, errors of various kinds are inevitable, and the 
absolutely correct measurement never can be determined (20). 
If a measure of a biological marker yields results that differ 
markedly from one occasion to another, it is of little value in 
research or quantitative risk assessment.

It is possible to use quantitative indices of the extent of random
variation of a biological marker. These indices can be used to
determine whether the reliability of a given measure is sufficient
for the purpose being considered. The two most common indices
are the standard error of the measurement and the reliability
coefficient (20). To assess random errors, multiple measure-
ments are needed to compensate for the fact that the random er-
or in the arithmetic mean of several measurements is likely to
be much less than the random error in an individual measure-
ment (20). In most epidemiologic research using biologic
markers, there are seldom large numbers of individual values.
Thus, only a small number of individuals can be used as a sam-
ple of the infinitely larger population to which the distribution
refers. The standard error indicates how the mean of that sam-
ple is distributed around the mean of the larger population.
Hence, the standard error of the mean reflects the reliability
of the sample mean as an indicator of the population mean (20).
This may not be as informative as the reliability coefficient for
evaluating markers to be used in risk assessments.

The reliability coefficient is technically known as the intraclass
coefficient of variability (21) and ranges from 0 to 1. If each
measurement is identical, then the intraclass coefficient is 1.0.
The greater the variation between measurements, the less the
reliability. Fleiss (21) has evaluated the impact of unsystematic
variation in measurement, described the untoward consequences
unreliability, and recommended how unreliability can be con-
trolled. The untoward consequences described by Fleiss include:
the need to increase sample sizes to overcome unreliability; the
systematic biased reduction of correlations between a health
measure and the measured extent of exposure to an environ-
mental risk factor; and high rates of misclassification in case-control
studies of the association between exposure and disease (20). All
of these pertain to studies using biologic markers of exposure or
effect. Fleiss (21) recommends that unreliability be controlled by
conducting pilot studies and replicating measurement pro-
cedures on each study subject. In some cases the measurement
of the amount of a marker is not an end in itself but is used to
calculate some other value, thereby propagating measurement er-
ers (20). Since correct values from measurements are generally
never known, calculations will, perforce, involve errors. Thus, it
is useful to know how errors in individual measurements affect
the results of subsequent calculations (20). For example, individual
errors in a sum or difference of measurements are added and
standard errors are combined with the root sum of squares (20).
Acknowledgment of these calculation errors should be includ-
ed in studies and subsequent risk assessments. When such errors
become significant, appropriate adjustments should be made.

Internal Study Validity

Another building block of quantitative risk assessment is the
study from which inferences about the association between ex-
posure and effect are drawn. Last (9) has defined the internal
validity of a study as the degree to which index and comparison
groups are selected and compared so that, apart from sampling
errors, the observed differences between the dependent variables
are attributed only to the hypothesized effect. This is validity in
the estimation of effect, and it is dependent on the ability to con-
trol bias. Internal study validity has been widely discussed in
epidemiological textbooks. Hence, in this section we will discuss
some issues of internal validity that pertain to the use of biological
markers. Some of this discussion is specific for markers, but there
are other general issues that also merit comment.

Bias is a distortion that may result when evaluating an associa-
tion and can occur when subject selection is unequal according
to disease or exposure status. In selecting subjects for studies in-
volving biologic markers, it is necessary to identify factors such
as background rates of markers and the range of normal variables
so that classification and subject selection are equal for the
groups being compared. These issues have been discussed else-
where (5,7). Bias can also result from misclassification of sub-
jects based on exposure or disease and failure to adjust for
other variables that are also predictive of the disease of interest.

Misclassification

Differential misclassification of exposure or disease can
reduce the validity of a study (3,7). Biologic markers that allow
for the reduction of misclassification enhance study validity.
Similarly, biologic markers can contribute to the reduction of
non differential misclassification. This type of misclassification,
which has been considered a lesser threat to validity, can result in
bias toward the null value (22).

The key to valid epidemiologic studies and, hence, valid quan-
titative risk assessment, 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 biological markers (23). However, as Rogan (23)
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 biological
markers can provide empirical data, which are preferential to the
use of deductively derived estimates (23).

For example, Lawrence and Taylor (24) demonstrated the
value of empirical exposure measurements when they were con-
fronted with the problem of assessing historical PCB exposures
of women who manufactured electrical capacitors. The purpose
of their investigation was to determine the effects of PCB ex-
posure on the women's reproductive outcomes during the period
1979 to 1983. Though the investigators did not have actual serum
PCB measurements for that period, they did have a complete
work history for each subject and industrial hygiene data that
allowed classification of each job in terms of a low, medium, or
high 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
a part of a general company survey. Using those data, the in-
vestigators 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 concentrations, derived from a sample
of subjects, was used as a biologic marker to construct a more ac-
curate estimate of the true exposure than was available using job
classification data.

Analytical Adjustment for Other Variables

When there are multiple variables to be considered in a study,
proper data analysis depends on the choice of the correct
mathematical model. The strongest models take into account a priori hypotheses specific to the topic under study. The incorporation of biologic markers in study designs and mathematical models also implies an understanding of the direction and mechanism of action. Additionally, by controlling measurement validity, it is also possible to partially control study validity, as measurement errors can produce biased estimates of regression coefficients used in models (25).

Longitudinal studies that employ biological markers will find increasing use in quantitative risk assessments. The validity of those study results will depend, in part, on the analytical approach selected. Such studies may involve repeated measures of a continuous random variable. Thus, there may be measurement errors that are considered random between persons, but which are autocorrelated within persons. The use of autoregressive modeling for the analysis of longitudinal data by epidemiologists is increasing and is likely to be used more frequently in studies involving biological markers. These models allow for the treatment of the time course of change of a variable (26). Other methods for analyzing repetitive measures that assume a Gaussian error structure have been reviewed by Louis (25), who concluded that this area needs continued statistical, numerical, and interpretive research and development.

External Validity

Risk assessment is an effort to address a condition of incomplete data (27). Hence, risk assessment involves the extrapolation (or generalization) from 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 extrapolating between populations or species; from high doses to low doses; and between different organs within a species. All of these efforts can be enhanced by using biologic markers common to each population or species. Allometric assessments 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 (28).

There are many uncertainties attendant to extrapolating to a large population from data derived from an epidemiologic study of a smaller group. The characteristics that make a study internally valid are often barriers to extrapolation. Extrapolation is, nevertheless, current practice in risk assessment. Using valid biological markers may allow some evaluation of whether a particular extrapolation is warranted; the variability is too extreme; or if differences in susceptibility have resulted in sensitive subgroups (27).

Extrapolation to low doses (or exposures) involves determining (or assuming) the shape of the dose-response curve. Establishing a dose-response relationship in a risk assessment might be considered a meta analytic procedure in some instances. That is, results from different studies might be combined to provide a larger sample size or a broader range of dose estimates. The validity of this effort can be enhanced if the same markers are used in different studies or if different markers have been shown to be correlated (i.e., have construct validity).

The contribution of macromolecular adducts to low-dose extrapolation has been the most heralded potential improvement to risk assessment. However, the use of biologic markers also can be a source of confusion in risk assessments. Most of the studies of adducts in humans have not yet demonstrated a clear dose response (1,29). This may be due to the wide variability in human response and the current inability to determine true individual exposures. Until the sources of variability can be identified and their impact evaluated, the absence or faulty characterization of a dose-response will limit the usefulness of this class of biologic markers in risk assessments (30,31). A potentially major source of differential susceptibility in dose response is the phenotypic variation of metabolic parameters (30). Rarely has this variation been considered in risk assessments.

The effect of the choice of a dose variable on risk estimates can be severe, especially when the pattern of exposure that the estimates are thought to reflect differs from the predominant pattern experienced by a study cohort (32). The use of a biological marker of exposure can help reduce the impact of using an ambiguous dose variable because it can more accurately reflect the true dose, even in studies where exposures are observed to have occurred over a wide range. For example, attempts have been made to compare biologically effective doses at high exposures where tumors are observed to low exposure concentrations to determine whether linearity of the carcinogenic effect is a valid assumption. Perera (1,29) has concluded that extensive data on DNA, RNA, and protein binding indicate that macromolecular effects, at the lowest administered doses, generally follow first-order kinetics (i.e., the rate of binding in target organs in vivo is directly proportional to 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, such as protein adducts, to dose have been shown to be constant over a dose range of 10⁻³ mole/kg to 10⁷ mole/kg (28,33). However, as Swenberg (34) 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 more than one order of magnitude (34).

Examples of Using Biologic Markers in Risk Assessment

The theoretical discussion of marker validity can be applied to risk assessments concerning the fumigant and fuel additive, ethylene dibromide (EDB) and the sterilant and chemical intermediate, ethylene oxide (EtO). Examination of the data concerning these two substances and their relationship to the disease process can provide some insight into the question of marker validity. This examination is summarized in Table 1. As will be seen from the following discussion, what appears to be a valid marker of EDB exposure and consequent disease risk turns out to be valid only at high exposures. The data concerning EtO, however, provides reason for optimism that selection of the appropriate biological marker can provide a more precise estimate of exposure-response at low doses and, therefore, risk.
| Table 1. Aspects of validity in cancer risk assessments of ethylene bromine (EDB) and ethylene oxide (EtO). |
|------------------------------------------|
| **Validity type** | **EDB** | **EtO** |
| EDB, bromine release | Not valid over wide range of exposures | Valid over wide range of exposures |
| EDB, hemoglobin alkylation | Association only with acute toxicity | Association with acute and chronic toxicity |
| **Criterion** | **Reliability** | **Internal study** | **External** | **Usefulness in quantitative risk assessment** |
| Not related to cancer | Measure is reproducible | Br release related to acute response | Poor surrogate of cancer biologically effective dose | Can overestimate exposure response leading to underestimate of true exposure-response relationship |
| Related to genotoxicity | Measure is reproducible | Associated with exposure | Good surrogate of cancer biologically effective dose | Better measure of airborne exposure and biologically effective dose. |

**Bromine Release in Ethylene Dibromide Toxicity**

In 1977, when the National Institute for Occupational Safety and Health recommended standards for occupational exposure to EDB, it was established that EDB caused mutations in fungi, plants, bacteria, insects, and mammalian cell systems, and that it induced cancer in several mammalian species (35). The data presented in that criteria document described several biochemical events that allowed investigators to estimate the internal dose of EDB.

First, as EDB was absorbed, glutathione production initially decreased, but then recovered. The decrease in the amount of glutathione was associated with the release of 2 moles of bromine for every mole of glutathione that disappeared. The production of free bromine could be correlated to the airborne exposure concentration, providing an indication of dose. Further evidence was provided to show that the production of S,S'-ethylenebis(glutathione) was saturable (35). More recent data indicates that when the first molecule of glutathione reacts with EDB, it can form a three-membered sulfur-containing ring that can alkylate DNA to form S-[2-(N7-guany1) ethyl] glutathione. This alkylation can occur prior to the detoxification reaction of EDB with the second molecule of glutathione (36).

These simple data offer some insight into the overall relationship between EDB exposure and cancer development. The fact that the detoxification pathway is only one of the metabolic pathways indicates that detoxification removes only a portion of the EDB from the system, the remainder being available for reaction with cellular macromolecules. Second, it is possible that EDB does not react with cellular macromolecules until the detoxification pathway has become saturated. If this latter scenario is adopted, then consideration must be given to the existence of a threshold of exposure. The first choice, on the other hand, provides support for the concept that there is no threshold.

Data from other species clearly show that EDB alkylates macromolecules and causes mutations, even at doses well below those that saturate the detoxification pathway, lending support to the theory that there is no threshold for the carcinogenic response. Finally, the production of tumors appears to be related to the cumulative dose (i.e., the exposure concentration multiplied by the duration of exposure).

If the quantitative relationships between exposure concentration, exposure duration, bromine production, adduct formation, gene mutation, and tumor expression were understood, then it would be feasible to use bromine production as a marker of increased cancer risk for measures of bromine prior to saturation of the pathway. Does the information about bromine production make sense in the context of EDB induced cancer? Certainly the information makes sense, at least qualitatively. EDB is used as a fuel additive because its bifunctionality is exploited to remove excess lead from engines (7). It is that same reactive property that allows EDB to act as a bifunctional alkylation of macromolecules. When the alkylation of DNA occurs, the cell attempts to repair the damage. If the rate of repair is less than the rate of alkylation, then the damage persists and can lead to a variety of untoward effects. The observation of enhanced DNA repair rates in mammalian systems supports this mechanism. However, it is important to note that the initial studies on bromine production were conducted at high, intragastric doses that saturated the metabolic detoxification mechanisms (35).

Other studies in which animals were exposed to EDB in air at lower concentrations indicated that the rate of metabolism was about 100 times greater than the rate of absorption, and thus exposure by inhalation may not pose the same threat as exposure by other routes such as feeding or gavage (35). Subsequent inhalation studies revealed that inhalation exposures at 10 ppm resulted in tumor development in mammals (35). Based on pharmacokinetic data, an exposure at 10 ppm would result in the absorption of as little as 0.4 μmole EDB/L of air, a concentration well below that shown to saturate detoxification mechanisms (35). These data indicate that EDB can exert its effects in two ways: by direct action on the tissues that it contacts; and systemically. The latter mechanism indicates that normal detoxification mechanisms do not adequately remove all the EDB, even at relatively low doses. Based on this information, it appears that bromine production, while qualitatively consistent with a possible carcinogenic mechanism, is not a good quantitative marker for EDB-induced carcinogenesis.

In order to obtain more precise information on the relationship between EDB exposure and cancer induction, a marker more sensitive to cellular activity than bromine release is needed. One such marker might be the formation EDB-DNA adducts, or as appears to be the case for EtO, the formation of hemoglobin adducts.

**Hemoglobin Alkylation by Ethylene Oxide**

Qualitatively, the data concerning the toxicity of EtO parallels that of EDB. Each of those chemicals is acutely toxic. EtO and EDB can cause mutations in a variety of plant, bacterial, insect, and mammalian species both in vitro and in vivo, and a number of investigators have clearly established the relationship between EtO exposure and alkylation of hemoglobin, DNA, and cancer development. For example, Burgnone et al. (37) have demonstrated that the extent of in vivo hemoglobin alkylation is proportional to the airborne concentration of EtO and the concentration of EtO in blood. Calleman et al. (38,39) and Osterman-Golkar (40) have shown that the amount of EtO in blood is proportional to the formation of DNA adducts. In a related study, Yager (41) has demonstrated that the frequency of
sister chromatid exchange in peripheral blood of EtO-exposed workers is proportional to cumulative dose (i.e., ppm × hr). Finally, Calleman et al. has shown that there is a relationship between the extent of hemoglobin alkylation by EtO and the number of rats with tumors following inhalation exposure to EtO (38,39). Calleman used thes data to estimate the risk of developing leukemia as a result of EtO exposure (38,39).

It is clear that the formation of alkylated hemoglobin by EtO satisfies the requirements of a valid biological marker. Though the formation of that particular marker appears to be an event that occurs independent of those related to EtO-induced cancer development, the formation of hemoglobin adducts by EtO appears to be a good surrogate for predicting risk. This conclusion is based on the assumption that other mammalian hemoglobin would respond similarly, however, the precise relationships do need to be elucidated. These relationships have been demonstrated in subsequent research (42,43).

Conclusion

The framework presented here and in a previous paper (6) may serve as a basis for evaluating the validity of biological markers for research and for quantitative risk assessments. At present, there are few valid biological markers that can be used to conduct quantitative risk assessments. Before a marker is useful in risk assessment, it should be found to have content, criterion, and construct validity, and it should be shown to be reliable. Pilot studies should be performed to establish background levels, the range of normal, confounding factors, and optimal collection and analytical techniques. Research studies using biological markers will need to be of appropriate sample size and pay attention to the proper selection of subjects and the use of appropriate statistical techniques (5,6).

If studies are to be useful in risk assessment, they must be generalizable but, more importantly, they must be internally valid. Hence, to satisfy the ultimate need for generalizability and still be internally valid, studies should involve heterogeneous population samples with homogenous subgrouping within the samples. If separate studies are conducted for use in risk assessments, efforts should be made to use similar markers and to pay attention to confounding factors.

Failure to consider the validity of components of a risk assessment can lead to erroneous conclusions. For example, in the case of EDB, if attempts were made to construct risk arguments based on bromine release data, it might have been concluded that there is a threshold of exposure that must be passed before the carcinogenic process can be initiated. The EtO data, on the other hand, clearly show relationships between airborne exposure concentrations, time, and events at the molecular level that are at least indicative of a genotoxic and carcinogenic mechanism that is consistent with generally accepted theories of carcinogenicity.

REFERENCES

1. Pera, F. P. Biological markers in risk assessment. Environ. Health Perspect. 76: 141–146 (1987).
2. Hattis, D. The value of molecular epidemiology in quantitative health risk assessment. In: Environmental Impacts on Human Health (S. Dragga, J. C. Cohrsenn, and R. E. Morrison, Eds.), Praeger, New York, 1987, pp. 89–115.
3. Rowland Hogue, C. J., and Brewster, M. A. Developmental risks: epidemiologic advances in health risk assessment. In: Epidemiology and Health Risk Assessment (L. Gordis, Ed.), Oxford University Press, New York, 1988, pp. 61–80.
4. Committee on Biological Markers, National Academy of Sciences. Biological markers in environmental health research. Environ. Health Perspect. 74: 3–9 (1987).
5. Schulte, P. A. Methodologic issues in the use of biological markers in epidemiologic research. Am. J. Epidemiol. 126: 1006–1016 (1987).
6. Schulte, P. A. A conceptual framework for the validation and use of biological markers. Environ. Res. 48: 129–144 (1989).
7. Hultka, B. S., and Wilcosky, T. Biological markers in epidemiologic research. Arch. Environ. Health 43: 83–89 (1988).
8. Alevanja, M., Aron, J., Brown, C., and Chandler, J. Cancer risk assessment models: anticipated contributions from biochemical epidemiology. J. Natl. Cancer Inst. 78: 633–643 (1987).
9. Last, J. M., Ed. A Dictionary of Epidemiology. Oxford University Press, New York, 1983.
10. Nunnaly, J. C. Psychometric Theory. McGraw-Hill, New York, 1967.
11. Zeller, R. A., and Carmines, E. G. Measurement in the Social Sciences. Cambridge University Press, Cambridge, 1980.
12. Abramson, J. H. Making Sense of Data. Oxford University Press, New York, 1988, pp. 94–95, 152–155.
13. Cronbach, L. J., and Meehl, P. E. Construct validity in psychological tests. Psychol. Bull. 52: 281–302 (1955).
14. Giarelli, E. E., Campbell, J. P., and Zedock, S. Measurement Theory for the Behavioral Sciences. W. H. Freeman and Company, San Francisco, CA, 1981, pp. 1–30.
15. Griffith, J., Duncan, R. C., and Hultka, B. S. Biochemical and biological markers: implications for epidemiologic studies. Arch. Environ. Health 44: 375–381 (1989).
16. International Programme on Chemical Safety. Guidelines on Studies in Environmental Epidemiology, Environmental Health Criteria 27, World Health Organization, Geneva, 1983, pp. 133–136.
17. Hemstreet, G. P., Schulte, P. A., Ringen, K., Stringer, W., and Altekruse, E. B. DNA hypereployd as a marker for biological response to bladder carcinogenic exposure. Int. J. Cancer 42: 817–820 (1985).
18. Matanoski, G. M. Issues in the measurement of exposure. In: Epidemiology and Health Risk Assessment (L. Gordis, Ed.), Oxford University Press, New York, 1988, pp. 107–119.
19. Office of Technology Assessment. Technologies for Detecting Heritable Mutations in Human Beings, OTA-H-298. U.S. Government Printing Office, Washington, DC, 1986, p. 104.
20. Massey, B. S. Measures in Science and Engineering. Ellis Horwood Ltd., Chichester, UK, 1986, pp. 66–101.
21. Fleiss, J. 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.), University of Pittsburgh, Pittsburgh, PA, 1986, pp. 9–16.
22. Rothman, K. J. Modern Epidemiology. Little Brown and Company, Boston, MA, 1986, pp. 86–89.
23. Rogan, W. J. Relation of surrogate measures to measures of exposure. In: Epidemiology and Health Risk Assessment (L. Gordis, Ed.), Oxford University Press, New York, 1988, pp. 148–158.
24. Lawrence, C. E., and Taylor, P. R. Empirical estimation of exposure in retrospective epidemiologic studies. In: Environmental Epidemiology (F. C. Kopfler and G. F. Craun, Eds.), Lewis Publishers, Inc., Chelsea, MI, 1985, pp. 239–246.
25. Louis, T. A. General methods for analyzing repeated measures. Stat. Med. 7: 39–45 (1988).
26. Rosner, B., Munoz, A., Tager, I., Speizer, R., and Weiss, S. The use of an autoregressive model for the analysis on longitudinal data in epidemiologic studies. Stat. Med. 4: 457–467 (1985).
27. Ederich, L. S. Combining animal and human data: resolving conflicts, summarizing the evidence. In: Epidemiology and Health Risk Assessment (L. Gordis, Ed.), Oxford University Press, New York, 1988, pp. 197–207.
28. Slaga, T. I. Interperspecies comparisons of tissue DNA damage, repair, fixation, and replication. Environ. Health Perspect. 77: 73–82 (1988).
29. Perera, F. P. The significance of DNA and protein adducts in human biomonitoring studies. Mutat. Res. 205: 255–269 (1988).
30. Mutuelsky, A. G. 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.), Plenum Press, New York, 1988, pp. 7–9.
VALIDATION OF BIOMARKERS FOR QUANTITATIVE RISK ASSESSMENT

31. Brown, S. L. 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.), Plenum Press, New York, 1988, pp. 255–269.

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

33. Farmer, P. B., Newmann, H. G., and Henchler, D. Estimation of exposure of man to substances reacting covalently with macromolecules. Arch. Toxicol. 60: 251–260 (1987).

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

35. Criteria for a Recommended Standard: Occupational Exposure to Ethylene Dibromide. National Institute for Occupational Safety and Health, Centers for Disease Control, U.S. Public Health Service, Department of Health and Human Services, DHEW (NIOSH) Pub. No. 77-221, 1977.

36. Inkeep, P. B., Koga, N., Cmarik, J. L., and Guengerich, F. P. Covalent binding of 1,2-dihaloalkanes to DNA and stability of the major DNA adduct, S-[2-N7-quanyl] ethyl glutathione. Cancer Res. 46(6): 2839–2844 (1986).

37. Burgone, F., Pertellini, L., Faccini, G. B., Pasini, F., Bartolucci, G. B., and DeRosa, E. Ethylene oxide exposure: biological monitoring by analysis of alveolar air and blood. Int. Arch. Occup. Environ. Health 58: 105–112 (1986).

38. Calleman, C. J. Hemoglobin as a dose monitor and its application to the risk estimation of ethylene oxide. Thesis, Department of Radiobiology, University of Stockholm, Stockholm, Sweden, 1984.

39. Calleman, C. J., Ehrenberg, L., Jansson, B., Osterman-Golkar, S., Segerback, D., Svensson, K., and Wachtmeister, C. A. Monitoring and risk assessment by means of alkyl groups in hemoglobin in persons occupationally exposed to ethylene oxide. J. Environ. Pathol. Technol. 2: 427–442 (1978).

40. Osterman-Golkar, S., Farmer, P. B., Segelbach, D., Bailey, E., Calleman, C. J., Svensson, K., and Ehrenberg, L. Dosimetry of ethylene oxide in the rat by quantitation of alkylated histidine in hemoglobin. Teratog. Carcinog. Mutagen. 3: 395–405 (1983).

41. Yager, J. W., and Benz, R. D. Sister chromatid exchanges induced in rabbit lymphocytes by ethylene oxide after inhalation exposure. Environ. Mutagen. 4: 121–134 (1982).

42. Osterman-Golkar, S., and Bergmark, E. Occupational exposure to ethylene oxide. Relation between in vivo dose and exposure dose. Scan. J. Work Environ. Health 14: 372–377 (1988).

43. Hogstedt, C., Aringer, L., and Gustavsson, A. Epidemiologic support for ethylene oxide as a cancer causing agent. J. Am. Med. Assoc. 355: 1575–1578 (1986).