ELSEVIER

Toxicology 101 (1995) 73–88

TOXICOLOGY

Validity criteria for the use of biological markers of exposure to chemical agents in environmental epidemiology

P.A. Schulte*[a], G. Talaska[b]

[a] *Screening and Notification Section, National Institute for Occupational Safety and Health, CDC, Cincinnati, OH 45226, USA*

[b] *Department of Environmental Health, University of Cincinnati, Cincinnati, OH, USA*

Abstract

Biomarkers may prove very useful in increasing the precision of exposure estimates during field epidemiological studies of environmental and occupational exposures. However, the determination of validity of exposure biomarkers is a laborious process. It is also a process that needs collaboration between laboratory and field scientists if biological markers of exposure are to be useful tools in environmental epidemiology.

Keywords: Validity criterion; Biological marker; Exposure; Chemical agent; Environmental epidemiology

1. Introduction

Exposure assessment has proven to be a major difficulty in environmental epidemiology. Much epidemiological research involves a retrospective perspective and exposure reconstruction is difficult due to a lack of adequate information regarding exposure histories of subjects. In addition, there is an enormous potential for exposure assessment to be influenced by apparent disease occurrence, and the often long latency period between exposure and disease (National Research Council, 1991a,b). This paper addresses the role biological markers (biomarkers) can play in exposure assessment and the criteria for determining when a marker of exposure can be used in environmental epidemiology. General questions of marker validity will be addressed followed by specific examples involving biomarkers of biologically effective dose of carcinogens. For purpose of discussion, exposure will be considered as "an event that occurs when there is contact between a human and the environmental contaminant" (National Research Council, 1991a). Effective methods to measure exposure in the ambient air have been widely utilized (National Research Council, 1991b). These methods can now be supplemented by approaches that evaluate biological markers of exposure. Biological

* Corresponding author.

Elsevier Science Ireland Ltd.
SSDI 0300-483X(95)03020-3

markers of exposure are indicators of internal dose, biologically effective dose, and early biological effects that indicate the presence or amount of xenobiotic that enters a person or biological changes correlated with exposure.

This discussion of biological markers can be circumscribed by the broader question of what kind of exposure data epidemiologists need (Gann, 1986). As Gann notes:

> The answer depends on the development of a well-defined research question in each case. This question (or questions) should usually be defined before exposure or effect data are selected and should be based on a biological model of the exposure-effect relation that is as explicit as possible.

In selecting exposure data, it also helps to specify the level or type of environment/disease association that is sought:

> Four levels or types of association can be examined in epidemiological studies each calling for a different degree of precision and validity in the exposure data. Every analytic epidemiological study generates an exposure-response relationship, even the simplest study which might only compare two points: exposure and no exposure. Fig. 1 illustrates a hypothetical exposure-response curve for an air pollutant; in this case the 'curve' is linear and intersects the exposure axis at exposure level C. In studies with individual data on exposure and outcome, each individual will contribute a point, or individual data will be collapsed into groups to form fewer points. In studies with aggregate data, obviously each aggregate or group will contribute one point. In Fig. 1, level A refers to a study that seeks 'any association' between environment and disease — thus permitting comparison of populations use of more crude data such as might be available in data bases, since a correct answer can be achieved even if actual exposures are considerably different from those estimated. On the other hand, studies that ask questions regarding the shape (e.g., slope and position) of the exposure-response curve require more finely tuned data on exposure at two or more convenient points (segment B in Fig. 1).

If the research question concerns determination of a threshold or 'no effect' level of exposure, the investigators must be able to identify groups or individuals with exposure on both sides of, and close to, point C. This demands even greater refinement in the exposure data, with less room for non-random error.

Finally, questions concerned with the portion of the total disease burden attributable to the environmental agent must either identify popula-

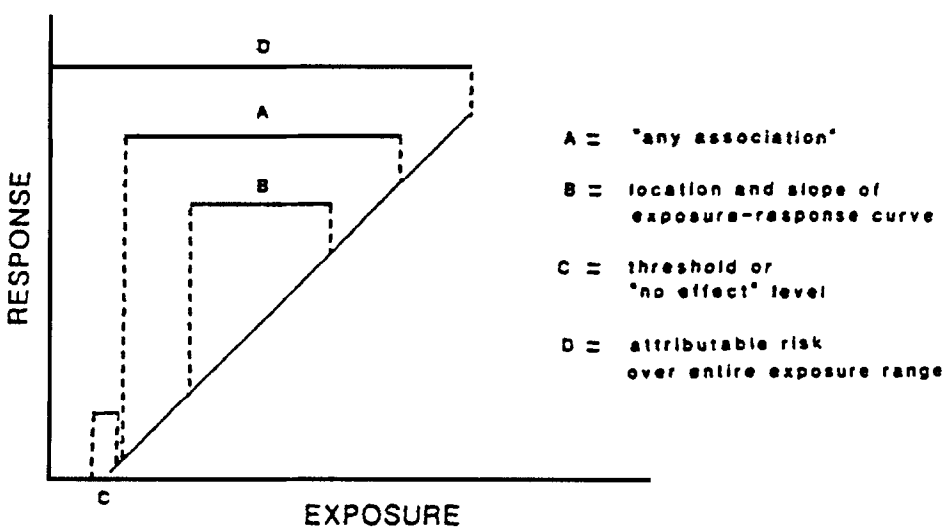

Fig. 1. Four distinct epidemiological questions and their exposure data requirements (Gann, 1986).

tion exposure across the entire range of exposure (segment D), or study a representative sample of cases, as in a case control study (Gann, 1986).The use of biomarkers for each of these levels will be the subject of the following discussion. At any level of association, valid biological markers of exposure will be those dose markers that have: (1) biological relevance; (2) defined pharmacokinetics; (3) temporal relevance; (4) defined background variability and identification of confounders.

2. Biological relevance

The biological relevance of markers is the extent to which it represents the underlying biological event. This has been referred to as 'content validity' (Schulte and Mazzuckelli, 1991). If an exposure marker has content validity it will represent part of a chain of events that are a subset of exposure opportunities and a pool from which outcome events are likely to occur. Hence, a DNA adduct formed from exposure to polycyclic aromatic hydrocarbons (PAH) will have content validity if it represents the interaction of PAH and DNA. The ultimate validity of a marker of exposure will depend on its relationship both to exposure and to outcome. This relationship is very difficult to prove in human studies due to the long latency between exposures, the appearance of the marker and the disease. But, until these studies are done, the importance of any marker to individual risk will be problematic. At the current level of understanding discussions of group risk, rather than individual risk, are warranted. Without demonstration of direct relationship to exposure and outcome, each biomarker study is actually a test of the biological relevance of the marker and adds to the web of association.

3. Understanding of pharmacokinetics and temporal relevance

The effective selection and use of biological markers of exposure depends on an understanding of the underlying pharmacokinetics and pharmacodynamics of suspected toxicants and the biomarkers (National Research Council, 1991b; WHO, 1986; Andersen, 1987; Smith, 1987; Droz, 1993; Lauwerys, 1983). Knowledge of pharmacokinetics is important in determining the frequency and timing of sampling and the tissues or fluids most appropriate for study. It also guides the interpretation of dose and effect data obtained in a target tissue or a surrogate.

Appropriate sampling of biological specimens requires that the temporal relationship of markers to external exposure or to disease be understood. Whether a marker reflects recent or cumulative exposures, peaks or averages, depends on the pharmacokinetics of the contaminant, and the persistence of the marker in the biological sample being assayed (which is, in large part, a function of the turnover rate or half-life of the sample) (National Research Council, 1991a).

Understanding of temporal relevance is essential for developing monitoring strategies and interpreting results. Most measures of internal dose, for example, reflect recent exposures. An exception would be a substance that is fat-soluble and is stored in adipose tissue (National Research Council, 1991a). Hemoglobin is a good integrating dosimeter over the 4-month life span of erythrocytes, which, unlike white blood cells, lack repair systems. In contrast, human serum albumin, another biological dosimeter has a half-life of 20–25 days. The elimination kinetics for a marker like carcinogen-DNA adducts is complicated by the fact that there are apparently two compartments for adducts in DNA. The majority of adducts appear to reside in a compartment where there is rapid loss ostensibly due to DNA repair, while other adducts reside in a compartment that is longer lived and consistent with cell turnover (Randerath et al., 1985).

In addition, it should be known in advance if an exposure is acute, chronic or intermittent, and whether the marker appears quickly after the exposure (as with many metabolites) or has a significant latent period, as when cells have to migrate to the surface of a tissue to be collected (Rosin, 1992). Chronic exposure generally results in steady-state marker levels if exposure duration is of sufficient intensity and duration so that marker formation is balanced by loss from such factors as repair, elimination and cell turnover.

Acute exposure can often be captured easily if the kinetics have been studied. The effects of intermittent exposure are often the most difficult to capture with biomarkers, particularly if there is a latent period in marker appearance (Schulte and Perera, 1993).

The manner and timing of sample collection can contribute to ambiguous or meaningless results. For example, in a study to determine exposure to environmental tobacco smoke with nicotine or its metabolites as markers, different results were obtained when different sampling protocols were used. Physiological pharmacokinetic modelling has demonstrated that the single-time point sampling protocol is of little use, unless a marker has a long half-life, which is rare for most exposures (Schwartz and Balter, 1988). Therefore, a good understanding of the pharmacokinetics is essential for proper timing of sample collection (National Research Council, 1991a).

The natural history of a biological specimen is also an important factor to consider:

> The period reflected by white blood cells or lymphocytes is considerably more complex. For example, adducts on DNA from white cells can reflect both past and current exposures, because a subset (T cells) is very long-lived. T cells make up approximately 60–90% of lymphocytes, which in turn constitute about 22–28% of peripheral white cells in blood. Thus, T cells constitute a maximum of 25% of white cells. The estimated half-life of T cells is 3 years. In contrast, B cells and monocytes constitute roughly 1–2 and 1–7%, respectively, of circulating white cells and have lifetimes ranging from days to weeks. Granulocytes make up the remaining 66–85% of white cells and are short-lived (hours to days) (National Research Council, 1991a).

Thus, it may be appropriate to fractionate lymphocytes and make marker measurements only on the subfraction actually exposed. Long-lived T cells are the cell type of choice in cases of past-discontinued or chronic exposure. In some cases fractionation of the target cell population may increase assay sensitivity and be better correlated with outcome. For example, when total cumulative exposure is more important to long-term health effect and for risk estimates, measuring marker levels in T-lymphocytes would give better information than measuring the marker exposure variability (National Research Council, 1991a). Savela and Hemminki (1990) saw no significant difference between groups of smokers and non-smokers when measuring carcinogen-DNA adducts in granulocytes or all white blood cells, while a significant difference was seen when measuring these lesions only in T-lymphocytes.

The natural history of the biological specimens will also have an impact on epidemiological study designs:

> In retrospective case-control studies, for example, one would want a permanent marker left decades earlier by an initiating carcinogen. In the case of discontinued exposures, even the longest-lived markers will be diluted by cell turnover, thereby leading to underestimation of past or cumulative dose. Only if exposure had been continuous and had not changed substantially over the decades (and only if the disease had not altered metabolism) would current measurements of the marker be directly representative of critical prior exposure. Many other exposure patterns are relevant to case-control and cohort studies (current but interrupted, continuous but of varying magnitude, etc.). The effect of each pattern on a marker like DNA adducts in white blood cells would involve different distributions of adducts among white-cell populations leading to a varied pattern of persistence (National Research Council, 1991a).

4. Understanding of 'background' variability and confounding variables

It is important to know the range of values of a given biological marker in a 'normal' population (National Research Council, 1991a). Normal can be considered to mean without occupational exposure or without observed or hypothetical environmental exposure. Pristine populations, however, are rare, so 'non-exposed' populations generally have some exposure. The extent of these exposures can vary greatly. Byrant et al. (1988) reported that 4-aminobiphenyl adducts could be detected in non-smokers. At the extreme, Savela and Hemminki (1990) reported that lymphocyte

DNA adduct levels were as high in a Silesian control group as in a Finnish occupationally exposed group. Care must be taken not to be deceived by the extensive variation in biochemical individuality; what is considered 'healthy' in some individuals might indicate a health risk in others (Schulte, 1987). The range of 'normal' can be large. For example, it is well known that cholinesterase activity in people not exposed to organophosphorus insecticides covers a wide range (WHO, 1985). This normal range needs to be known to allow for interpretation of abnormalities resulting from exposure:

> Interindividual variation and intraindividual variation are important contributors to 'noise' or background in monitoring or epidemiological studies and should be characterized before large-scale application of particular biological markers. Such data, however, can be generated only by large-scale surveys with repeated sampling and efforts to control for confounding variables. Thus, a background study is an important epidemiological exercise in itself. These types of studies are often difficult to get funded since they generally involve assessments in healthy people. Nevertheless, they are critical to developing and validating biological markers for use in environmental epidemiology. With respect to carcinogen-DNA and carcinogen-protein adducts, substantial interindividual variation and intraindividual variation have been observed with PAHs, 4-ABP, and nitrosamines (Bryant et al., 1988; Harris et al., 1985; Perera et al., 1987). SCEs also vary widely within and between subjects (National Research Council, 1991a; Carano and Natarajan, 1988).

Biological markers are subject to greater variability than conventional exposure-assessment techniques, because the body actively participates in the collection, distribution, and elimination of absorbed contaminants. Confounding variables that must be accounted for in studies that use biological markers include age, sex, race, cigarette smoking, alcohol consumption, diet, drugs or other environmental exposures, genetic factors, and pre-existing health impairment. for example, alcohol intake is the most common cause of reduced metabolism of industrial chemicals (Fiserova-Bererova, 1987). Thus, in a study of those chemicals in which the chemical or a metabolite is used as a biological marker of internal or biologically effective dose, one would have to account for alcohol consumption to understand and interpret the data properly.

Assay variability also tends to cloud potential associations. Under the best conditions this variability only contributes to random noise and is not biased. However, if there is some systematic difference in the order of sample analysis and there is a significant downward or upward trend in reported results, then this type of variability can contribute to false positive or negative conclusions. Run-to-run variability can be important in particular where assays require reagents (such as enzymes) which do not come from the same lot. This type of variability can be controlled through the use of internal or external standards, repeated measurements, or by randomization techniques which distribute this variability with equal probability between groups (Talaska et al., 1987). Fig. 2 shows the results of several measurements made of a 4-aminobiphenyl-DNA adduct standard made before and during a study of carcinogen-DNA adducts in human urinary bladders (Talaska et al., 1991). The initial datum is a determination made prior to the analysis of the human samples — clearly adduct recovery was low. A change in methodology resulted in a marked improvement in adduct recovery. Methods were then held constant for the remainder of the study, even though it was realized that the

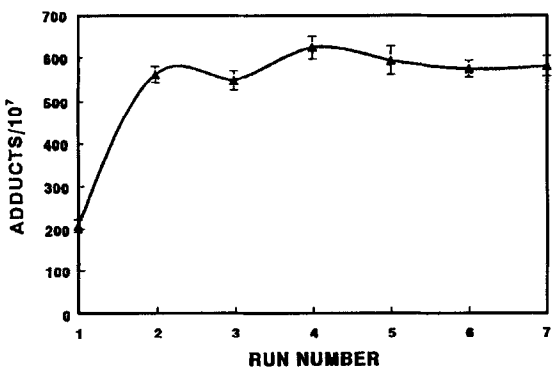

Fig. 2. Run-to-run variation in 4-ABP-DNA adduct standard by ^{32}P-postlabelling.

recovery of the adduct still was only about 30% percent with the improved method. The variation in the reported results obtained with the standard was viewed as acceptable and added less noise than further attempts to improve recovery.

5. Studies of the association between exposure and effect

When the research question is whether any association exists, an investigator has the option of comparing populations with maximum contrasts. If markers of exposure are the basis for these contrasts, the intragroup variability must be small enough so that the contrast can be observed, otherwise the populations will be misclassified as to exposure. Using exposure biomarkers as dependent or independent variables in environmental epidemiology requires that they have biological relevance, known pharmacokinetics, and that potential confounding factors are accounted for. Critical in this regard is attention to other sources of exposure such as occupation, diet, hobbies, etc., that might account for the marker (Schulte and Perera, 1993).

6. Studies of the location and slope of exposure-response curve

In this assessment, the exposure markers need to be precise enough to reflect two or more different exposures on the curve. Here the question of marker variability is important. If the variation within a person is greater than the variation between persons it will be difficult to determine what frequency a marker represents on the exposure axis.

7. Studies of the threshold or 'no effect level'

Markers used to indicate where the curve intersects the exposure level need to be able to be distinguished from background. At such low levels of exposure it is necessary to account for confounding factors that can influence the marker level.

This is generally the case in animals where one is dealing with inbred strains and confounding factors are controlled. Extensive data on DNA, RNA, and protein binding, in experimental animal systems indicate that these macromolecular effects at the lowest administered doses generally follow first-order kinetics; i.e., the degree of initial binding in target organs in vivo is usually directly proportional to presented dose. In some cases, that relationship also holds at very low doses similar to doses that might be encountered by humans as a result of environmental contamination (National Research Council, 1991a; Wogan and Gorelick, 1985).

Human data on adducts, however, do not demonstrate a strong proportional relationship between exposure and adduct frequency (response), most likely reflecting the great variability in the population. For example, frequencies of 4-aminobiphenyl (4-ABP) hemoglobin adducts were significantly higher in smokers than in nonsmokers in two early studies, but were not significantly correlated with amount smoked (Bryant et al., 1988; Perera et al., 1987). More recent hemoglobin and DNA adduct studies have shown significant correlations with dose for some markers, however, there is always wide variability in individual response (Philips et al., 1988; Bartsch, 1990; Vineis et al., 1990; Talaska et al., 1991; Schulte et al., 1992). This is undoubtedly due to factors involving nutrition (some people ingest anticarcinogens) and metabolism (Wattenberg, 1983). However, less interindividual variation would be anticipated with compounds like ethylene oxide and propylene oxide which, unlike PAH or 4-ABP, do not need to be metabolically attivated. Indeed, the frequency of hemoglobin adducts in humans were reasonably proportional to the estimated dose received. Exposure levels as low as 0.1 ppm were correlated with response in two studies (Mayer et al., 1991; Schulte et al., 1992), but another study did not show a correlation between exposure and hemoglobin-adduct frequency (Van Sittert et al., 1985), at airborne concentrations of ethylene oxide less than 0.05 ppm. For workers exposed to propylene oxide, good agreement was seen between the degree of hemoglobin alkylation and estimated propylene oxide exposure (National Research Council, 1991a; Osterman-Golkar et al., 1984).

Another example illustrates exposure-adduct

relationships over various exposure levels:

> Better defined exposure-response relationships were shown by the significant correlation between PAH-DNA adducts measured by immunoassay in peripheral white blood cells from Finnish foundry workers and their occupational exposure to PAH (Perera et al., 1988). Workers were classified as having high (more than 200 ng/m^3), medium (50–200 ng/m^3), or low (less than 50 ng/m^3) exposure BaP (as an indicator of PAH). The mean adduct concentrations (in fmol of adduct per μg of DNA) were 1.5 (high-exposure group), 0.62 (medium), 0.24 (low), and 0.066 (controls, unexposed, healthy workers seen at same clinic). These results were corroborated with the postlabelling method carried out on white-cell DNA from the same worker population (Hemminki, 1988; Philips et al., 1988). However, despite the correlation between DNA adducts and exposure at the group level, there was significant variation among individuals within the exposed groups. Adducts measured with immunoassay ranged from non-detectable to 2.8 fmol μg of DNA. The current PAH levels in the plant are approximately 30-fold lower than those which were previously shown to increase PAH-DNA binding (National Research Council, 1991a).

The broad variation may indicate that workers were receiving exposure by other routes such as ingestion, and that there are differences in processes such as activation, detoxification and DNA repair which are partially under genetic control (National Research Council, 1991a).

8. Attributable risk over entire exposure range

This level or type of association can be considered in two ways: a study where exposure is sampled at various points in the range and where exposure is not known. The exposure distribution curve will need to be developed by estimation or modelling. In these later instances the collection, calibration and analysis of biomarkers data must be compatible with pharmacokinetic models (Hattis and Silver, 1993).

Validity criteria: taking a marker from the laboratory to the field

Validity may have slightly different meanings in the laboratory and in the field. For environmental epidemiology to utilize valid exposure biomarkers both types of validity must be assured.

9. Natural history of the development of biomarkers

There is a natural history of the development of biomarkers; their content validity is evaluated on several levels or stages of experimentation. Those measurements which succeed and give insight into the disease process are candidates for further development and ultimately held as reliable for human studies. During the course of this evaluation more measurements, rather than fewer, must be made to ascertain whether the marker responds as appropriate to the various exposure situations.

Thus, it becomes important for the investigator to have a clear notion of the strengths and limitations of the biomarker. In many cases this understanding is best obtained by discussions with collaborators having different expertise. Laboratory scientists who work with the marker daily are likely to better appreciate the strengths and weaknesses of the measurements themselves, for example, their analytical sensitivity and specificity. Epidemiologists are often better prepared to discuss these same issues in terms of the assay as a whole. Thus, biomarkers must be validated both in terms of the marker itself, addressing such issues as content validity and temporal relevance, as well as the accuracy, precision, sensitivity and specificity of biomarker measurements in the field as they relate to exposure and disease.

There is also a certain probability that potential misclassification may be increased by the addition of an analytical measurement. Biomarkers are estimates of physiological events following an exposure to an agent. As estimators and not the parameters themselves, biomarkers are subject to the same types of variability as other estimates. A purpose of this section is to define components of measurement variability which may be introduced by the use of a biomarker, and to delineate several approaches to reduce the impact of analytical variability.

9.1. Developmental stage

The development or natural history of a cancer

biomarker in the laboratory often begins when a hypothetical association between an exposure or disease and some physiological change is tested. In many cases there is an incomplete understanding of the natural history of the disease process such that the events between 'cause' or exposure and the disease-effect are often unknown and are a 'black box'. Biomarkers promise to be useful in reducing uncertainty relating exposure to the disease process. However, when originally proposed they are often measurements of an unknown entity and must be tested against other established markers, or the occurrence of the disease itself, in studies which establish overall content validity. An untested marker can be considered in development until it can be demonstrated that the marker has sufficient content validity (Schulte and Perera, 1993), and that it is sufficiently sensitive and reproducibile to be used in the study of human populations.

The development of techniques for carcinogen-DNA adduct analysis is a useful model for the process of assay validation. The content validity of this measurement was suggested by Brooks and Lawley (1964), who saw that the ability of a chemical to form DNA, but not RNA or protein adducts, best predicted carcinogenicity. With the development of highly-sensitive techniques such as enzyme-linked immunoassays, ^{32}P-postlabelling, synchronous fluorescence and, most recently, mass spectrometry, the content validity of carcinogen-DNA adduct analysis has been more rigorously defined. Specific carcinogen-DNA adducts (Van de Poll et al., 1989) have been shown to induce to both point and gene mutations in a dose-related manner (Van Zeeland et al., 1985; Arce et al., 1987; Talaska et al., 1987). Other work has shown that adducts for certain carcinogens were found in a wide variety of target and non-target tissues, with DNA adducts in the target tissues usually occurring at higher levels than in the non-target organs although there were reported exceptions (Stowers and Anderson, 1984; Schurdak and Randerath, 1985; Schurdak et al., 1987). Another advantage of the current analytical methods for carcinogen-DNA adduct determinations is a decreased tissue requirement which allows work to be done on the basis of a single animal (Randerath et al., 1985). Animal studies using these techniques could detect differences between exposure groups using sample sizes of from five to seven animals (Randerath et al., 1985; Talaska et al., 1987). Reliable data were obtained on the basis of groups of animals receiving the same dose and/or on the basis of the individual animal in some cases. The markers were required to have a high degree of reliability. This was often demonstrated by the reproducibility of standards, the development of standard curves and by repeated measure. However, several authors have pointed out the need to analyze samples using designs (e.g., stratified block analysis) to minimize the effects of run-to-run variability which can be caused by different reagent lots and use of enzymes, the activity of which can decline with time. These studies demonstrated the sensitivity limits of many of the techniques proposed for use as human carcinogen biomarkers. Recognition that human exposure would likely result in lower marker levels than in these experimental situations spurred on the development of more sensitive variants of analytical techniques (Gupta, 1985; Reddy and Randerath, 1986; Gupta and Earley, 1988; Roberts et al., 1988).

These data gave investigators confidence that measurement of DNA adducts was important in carcinogenesis and had the potential for giving worthwhile information to human exposure and effect. Carcinogen-DNA adducts are now recognized as pre-mutagenic lesions and biomarkers of carcinogen effective dose. Although carcinogen-DNA adducts are neither necessary nor sufficient for neoplasia, their measurement has significantly increased knowledge of early steps in chemical carcinogenesis. On the other hand, certain issues still must be resolved through animal studies. For example, the influence of DNA repair and cell proliferation on fixing of the damage caused by carcinogen-DNA adducts remains to be determined. And the question of whether peak DNA adduct levels, persistent adducts, or total integrated adduct levels are better predictors of genetic effect has yet to be answered satisfactorily. But while these issues are being studied, these developmental animal models demonstrate the content validity of carcinogen-DNA adduct mea-

surements, encouraging their use in preliminary human studies.

9.2. Field validation

Once validated in the laboratory it is necessary for a marker to be validated in groups of people so that the investigator knows the extent to which it varies by demographic, behavioral, geographic or other relevant descriptors. Does it change with age? Is it the same in all races, sexes, in smokers and non-smokers? Does diet affect it? It is not necessary that every one of these factors is known, but an investigator should know the answer with regard to many of the variables that describe a potential study population. Key in validating an exposure marker for environmental epidemiology is the identification of sources of exposure that might account for measurable levels of markers in putatively unexposed populations. All of these field validation efforts should be viewed as epidemiological studies. Hence, they require attention to principles of epidemiological research.

Bull (1989) developed a decision model for biomarkers of exposure for the U.S. Environmental Protection Agency (USEPA). The objective was to determine levels of exposure in populations so that risk assessments could be conducted. Since the universe of biomarkers is potentially large, a conceptual model is needed for evaluating the completeness of information concerning a specific exposure biomarker. The output from the model would be used to determine whether a specific biomarker should be considered for further development using USEPA resources. The model is shown in Fig. 3 and Table 1. This model may serve as a checklist in evaluating potential biomarkers of exposure.

9.3. Trial human applications

Trial studies are done to determine whether the marker can measure defined differences in exposure in humans. The broadest comparisons are made: for example, groups with no exposure are compared with those having heavy exposure. These studies provide positive reinforcement for the analytic methodology. Exposure to compounds with a documented effect is important. Exposure to tobacco smoke has often been used in studies of this type because of its known contribution to lung and bladder cancer (Vineis et al., 1990). Since the major reason for doing these studies is to determine whether the biomarker responds as expected, measuring the effect in the target organ is critically important and tissues may be obtained by methods which are too invasive for routine monitoring purposes. Phillips et al. (1988), examined the levels of carcinogen-DNA adducts in human surgical lung samples. They found about a five-fold increase in total carcinogen-DNA adducts in smokers, the increase being correlated with the number of cigarettes smoked per day, suggesting a dose-response. Van Schooten et al. (1990) also looked at surgical samples of lungs of smokers and saw that adducts were higher in the normal but not the tumor tissue of tobacco smokers, consistent with that expected from the biology, the rate of cell turnover and metabolism in these tissues. In addition, they saw good agreement between two techniques of carcinogen-DNA adduct analysis (correlation = 0.97).

The urinary bladder is another target for carcinogens in tobacco smoke, with smokers estimated to be from 2–9 times more likely to develop urinary bladder cancer. Talaska et al. (1991) saw that a variety of specific carcinogen-DNA adducts were elevated in tobacco smokers. However, a dose-response could only be demonstrated for a single specific adduct, suggesting that there was significant interindividual variation in response to a given dose.

Significant differences in carcinogen-DNA adduct levels were detected in smokers and the magnitude of the differences approximately corresponded to the excess risk noted for tobacco smokers in epidemiological studies. Dose-response effects were noted in several studies, but the interindividual variability seen in others suggested the importance of pharmacokinetic factors affecting absorption (such as depth of smoke inhalation) or metabolism. The issues of data analysis important to studies at this level of development include the consistency of methods, serial repeats of standards and the reproducibility of

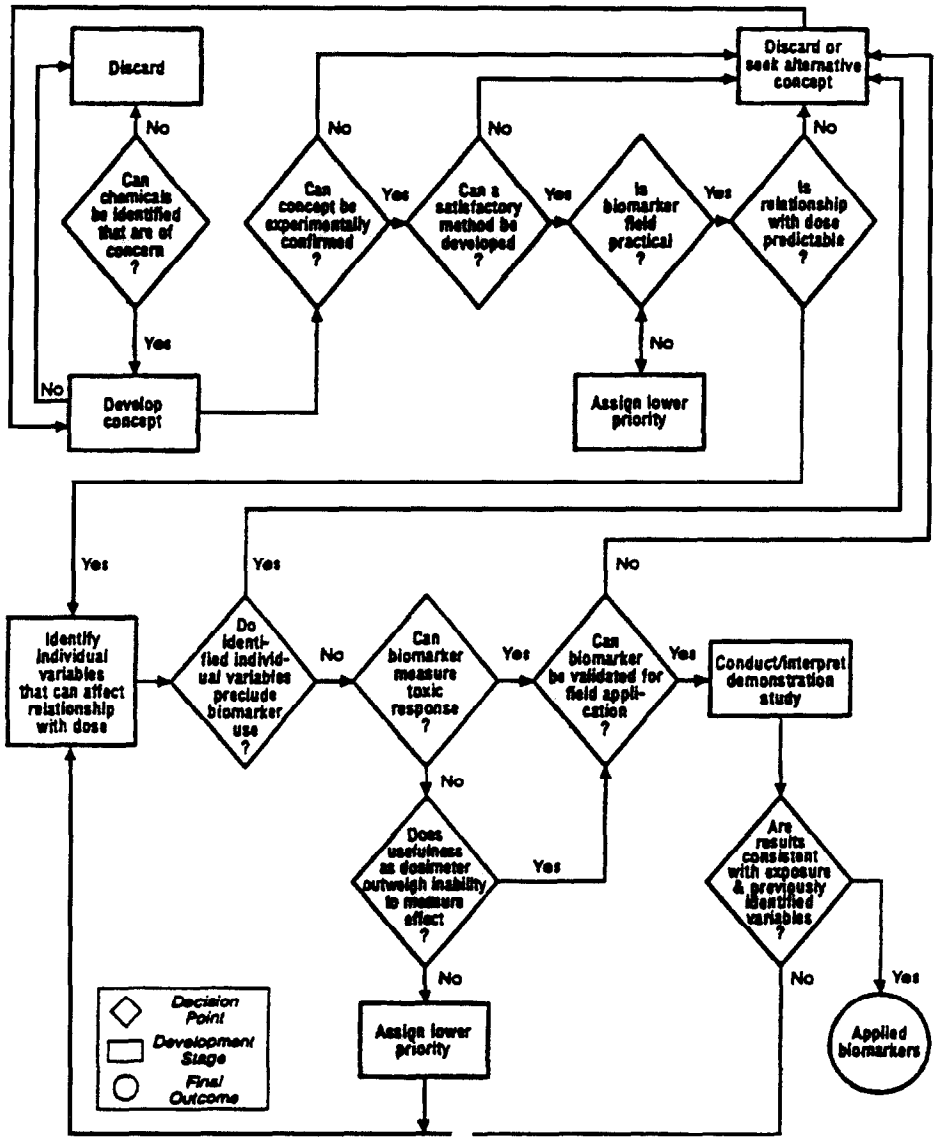

Fig. 3. Decision model for development of biomarkers of exposure (Bull, 1989).

repeated measurements. Repeat measurements are often possible at this stage because fairly large tissue samples are often available. The possibility for run-to-run variation in results is high because many of the assays used for carcinogen-DNA adduct analysis require enzymes as critical reagents. Fig. 4 shows the results of repeated measure of a series of human urinary bladder samples. In some cases there is good agreement between the two sets of the repeated measurements, while in other cases the correspondence was less good. A rule was established at the onset of the study to include all the data for each sample unless there was a technical reason for exclusion, so the means of all analyses were reported as results for an individual.

To be valid, a biomarker must also be practical. For carcinogen-DNA adduct analysis, obtaining

Table 1
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 humans 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).

target organ samples is an issue, since surgical samples will not be available during population-based studies. The alternatives are to seek out useful surrogates for target organ analysis, or to develop techniques to obtain target organ samples non-invasively. Selection of an appropriate surrogate tissue may not be simple owing to the fact that the target organ is a target because of some peculiarity of pharmacokinetics. For example, it appears that the urinary bladder is the target for aromatic amine carcinogens because this organ receives the carcinogen in an active form and in relatively high concentrations and then holds the compound until micturition, in some cases a period sufficient to allow interaction with DNA (Kadlubar et al., 1991). In the lung, carcinogens may be activated in situ, which may not permit the reactive forms contact with the central compartment. Therefore, if the proposed marker is to indicate the effective dose to the target organ and not just exposure, the correspondence between the marker levels in the surrogate and target tissue must be demonstrated. For example, Reddy and Randerath (1990) studied the relationship between carcinogen-DNA adduct levels in lymphocytes and a variety of target organs in an animal model. They saw that with the exception of benzo[*a*]pyrene there was no corre-

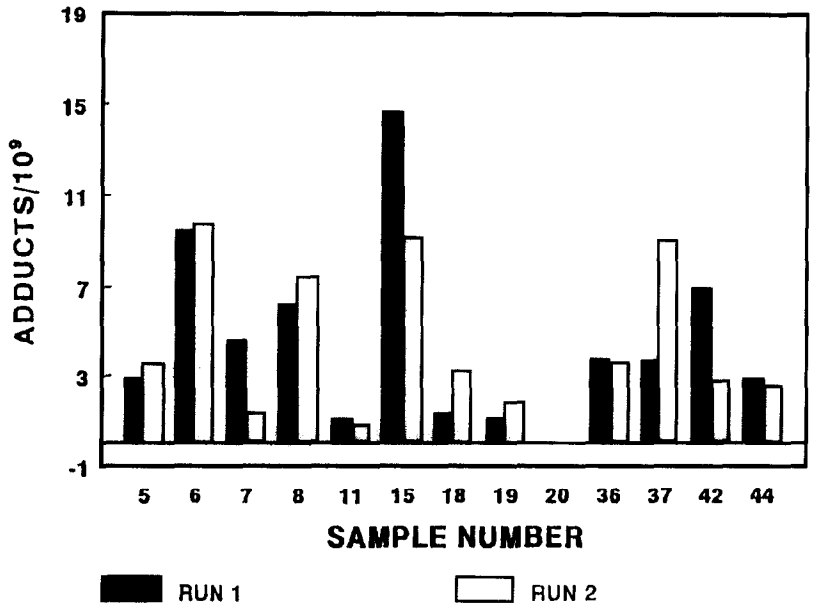

Fig. 4. Repeat variability in human urinary bladder biopsy samples. Run 1, left-hand columns; run 2, right-hand columns.

spondence between lymphocyte and target organ levels. This would indicate that using lymphocytes in human studies may be limited to indicating internal dose (and not effective dose) of carcinogens. Similarly, studies are currently underway to determine the correspondence between carcinogen-DNA adduct levels in exfoliated urothelial cells and the urinary bladder in humans. Additionally, results reported in one dog with chronic exposure to 4-aminobiphenyl have been encouraging (Talaska, 1990, 1992).

Studies at this stage can also be useful to point out the limitations of analytical approaches so that efforts can be made to improve the techniques. for example, Talaska et al. (1991) investigated the relationship between tobacco smoking and carcinogen-DNA adducts in exfoliated urothelial cells. While adduct levels were generally higher in smokers than non-smokers, and a dose-response was demonstrated within the group of smokers, the differences between smokers and non-smokers were not statistically significant. It appears that what may have actually been background material in some non-smokers samples were misinterpreted as carcinogen-DNA adducts reducing the assay signal-to-noise ratio and sensitivity. Clearly, changes in this technique are needed to improve sample cleanup and increase DNA yield. Synchronous fluorescence techniques may also have sensitivity problems for another reason, in that detection of carcinogen-DNA adducts in human samples requires as much 1 mg of DNA, which translates into about 1 g of tissue sample for each analysis (Manchester et al., 1988). Improvements to reduce this tissue requirement will make this method extremely valuable.

9.4. Secondary human studies

In this phase of the natural history of a biomarker the object is to determine whether the predicted relationship between the biomarker and exposure and effect modifiers has the expected content validity. In this case, the marker will be required to generate two (or more) discrete dose-response curves for levels of the effect modifier. Obviously this will require measurement of exposure, the marker of interest and the suspected effect modifier. For cancer markers this

stage is critical to more general use in population studies.

A series of papers which have reported on the interaction between 4-aminobiphenyl-hemoglobin adducts and *N*-acetyltransferase phenotype in a large group of tobacco smokers (Vineis et al., 1990; Talaska, 1991). In initial studies, the 4-aminobiphenyl-hemoglobin adduct biomarker was seen to respond to two types of tobacco associated with different urinary bladder risk (Bryant et al., 1988). The ability of this biomarker to respond to smoking cessation was also described (Skipper and Tannenbaum, 1990). Acetylation phenotype was suspected to be an effect modifier in epidemiological studies of human urinary bladder cancer (Cartwright and Glashan, 1984). The purpose of this work was to determine if the 4-aminobiphenyl-hemoglobin adduct marker would also respond in a similar fashion to this phenotypic difference. The levels of the 4-aminobiphenyl-hemoglobin adducts were increased in slow acetylators in each of the groups tested, non-smokers, and black and blond tobacco smokers. It is interesting that the effect of acetylator phenotype seemed strongest in the groups with the lowest tobacco intake (Talaska, 1988). Therefore, with these data two dose-response curves could be constructed to control for the effect modifier acetylator phenotype. Despite these preliminary data, ultimately, the role of the 4-aminobiphenyl-DNA in the causal pathway of carcinogenesis still needs to be confirmed.

9.5. Prospective human studies

The ultimate test of a cancer biomarker is whether or not it can be used to predict disease. Recently, Ross et al. (1992) conducted a prospective study of 18000 persons by measuring the levels of urinary aflatoxin metabolites and aflatoxin B_1-N^7-deoxyguanosine adducts, then following the individuals until tumor incidence. They found 22 cases of liver cancer. Although this adduct is not thought to be directly responsible for carcinogenesis, those persons who excreted this marker were almost five-times more likely to have the disease than those who did not. In addition, persons positive for excreted DNA adducts and hepatitis B antigens were at almost 50-fold greater risk. (An estimated 90% of hepatic tumors in China are thought to be associated with hepatitis B. This is the first prospective study indicating that a DNA adduct marker is associated with increased human cancer.) This type of approach to validation is costly and labor intensive. As an alternative, researchers should look to banked specimens and other creative approaches to validate markers.

10. Future cancer biomarker studies: exposure intervention studies

An exposure biomarker should respond to changes in the exposure. There have been several interesting studies conducted which showed that certain markers respond to exposure changes. As mentioned above, 4-aminobiphenyl-hemoglobin adducts were monitored in a group of people involved in a smoking cessation effort (Maclure et al., 1990; Skipper and Tannenbaum, 1990). The loss of these adducts in the blood approximately followed an expected curve based on the biological half-life of hemoglobin. Studies of carcinogen-DNA adduct in ex-cigarette smokers indicated that this marker responds to exposure interventions, although whether the kinetics of adduct loss follow cell-turnover kinetics in the urinary bladder or lung cells has not been determined as accurately as has been the case for the 4-aminobiphenyl-hemoglobin adducts (Philips et al., 1988; Talaska, 1988).

Intervention studies during the course of field trials will be very useful, both as an evaluation of the marker, as well as an evaluation of the intervention once the marker has been accepted as valid. Intervention studies involving evaluation of the effective dose will require estimation of external exposure as well as internal and effective doses.

References

Andersen, M.E. (1987) Tissue dosimetry in risk assessment, or what's the problem here anyway? In: Pharmacokinetics in

Risk Assessment, National Academy Press, Washington, DC, pp. 8–23.

Arce, G.T., Cline, D.T. Jr. and Mead, J.E. (1987) The ^{32}P-postlabelling method of quantitative DNA adduct dosimetry of 2-acetylaminofluorene-induced mutagenicity in Chinese hamster ovary cells and *Salmonella typhimurium* TA1538. Carcinogenesis 8, 515–520.

Bartsch, H., Caporaso, N., Coda, M., Kadlubar, F.F., Malaveille, C., Skipper, P., Talaska, G., Tannenbaum, S.R. and Vineis, P. (1990) Carcinogen-hemoglobin adducts, urine mutagenicity and metabolic phenotype in active and passive cigarette smokers. J. Natl. Cancer Inst, 82, 1826–1830.

Brooks, P. and Lawley, P.D. (1964) Evidence for binding of polynuclear aromatic hydrocarbons to nucleic acids of mouse skin: relationship between carcinogenic power of hydrocarbons and their binding to deoxynucleic acids. Nature 202, 781–785.

Bryant, M.S., Vineis, P., Skipper, P. and Tannenbaum, S.R. (1988) Hemoglobin adducts of aromatic amines: Associations with smoking status and type of tobacco. Proc. Natl. Acad. Sci. USA 85, 9788–9791.

Bull, R.J. (1989) Decision model for the development of biomarkers of exposure. US Environmental Protection Agency, EPA 600/x-89/163, US Government Printing Office, Washington, DC.

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

Cartwright, R.A. and Glashan, R.W. (1984) The epidemiology and management of occupational bladder cancer. In: L.T. Smith and D. Prout, (Eds), Bladder Cancer, Butterworths, London, pp. 125–150.

Droz, P.O. (1993) Biological monitoring and pharmacokinetic modeling for the assessment of exposure. In: P.A. Schulte and F.P. Perera, (Eds), Molecular Epidemiology, Academic Press, San Diego, CA, pp. 137–157.

Fiserova-Bererova, V. (1987) Development of biological exposure indices (BEIs) and their implementation. Appl. Indust. Hyg. 2, 87–92.

Gann, P. (1986) Use and misuses of existing data bases in environmental epidemiology: the case of air pollution. In: F.C. Kopfer and G.F. Craun, (Eds), Environmental Epidemiology, Lewis Publishers Inc., Chelsea, MI, pp. 109–122.

Gupta, R.C. (1985) Enhanced sensitivity of ^{32}P-postlabeling analysis of aromatic carcinogen: DNA adducts. Cancer Res. 45, 5656–62.

Gupta, R.C. and Earley, K. (1988) ^{32}P-Adduct assay: comparative recoveries of structurally diverse DNA adducts in the various enrichment procedures. Carcinogenesis 9, 1687–93.

Harris, C.C., Vahakangas, K., Newman, M.J., Trivers, G.E., Shamssuddin, W., Sinopoli, N., Mann, D.L. and Wright, W.E. (1985) Detection of benzo(*a*)pyrenediol epoxide-DNA adducts in peripheral blood lymphocytes and antibodies to the adducts in serum from coke oven workers. Proc. Natl. Acad. Sci. USA 82, 6672–6676.

Hattis, D. and Silver, K. (1993) Use of biomarkers in risk assessment. In: P.A. Schulte and F.P. Perera, (Eds), Moleculer Epidemiology, Academic Press, San Diego, CA, pp. 251–273.

Hemminki, K., Perera, F.P., Phillips, D.H., Randerath, K., Reddy M.V. and Santella, R.M. (1988) Aromatic DNA adducts in white blood cells of foundry workers. In: H. Bartsch, J. Hemminki and I.K. 0'Neill (Eds), Methods for Detecting DNA Damaging Agents in Mans Applications in Cancer Epidemiology and Prevention, IARC Scientific Pub No. 89, Lyon IARC.

Kadlubar, F.F., Dooley, K.L., Teitel, C.H., Roberts, D.W., Benson, R.W., Butler, M.A., Bailey, J.R., Young, J.F., Skipper, P.W. and Tannenbaum, S.R. (1991) Effects of voiding interval on metabolism, pharmacokinetics, blood hemoglobin adduct formation, and liver and urinary bladder DNA adduct levels in Beagle dogs administered the carcinogen, 4-aminobiphenyl. Cancer Res. 51, 4371–4377.

Lauwerys, R.R. (1983) Industrial Chemical Exposure: Guidelines for Biological Monitoring, Biomedical Publications, Davis, CA.

Maclure, M., Bryant, M.S., Skipper, P.L. and Tannenbaum, S.R. (1990) Decline of the hemoglobin adduct of 4-aminobiphenyl during withdrawl from smoking. Cancer Res. 50, 181–184.

Manchester, D.K., Weston, A., Choi, J.-S., Trivers, G., Fennessey, P.V., Quintana, E., Farmer, P.B., Mann, D.L. and Harris, C.C. (1988) Detection of benzo[*a*]pyrene diolepoxide-DNA adducts in human placenta. Proc. Natl. Acad. Sci. USA 85, 9243–9247.

Mayer, J., Warbuton, D., Jeffrey, A.M., Pero, R., Walles, S., Andrews, L., Toor, M., Latriano, L., Wazneh, L., Tang, D., Tsai, W.Y., Kuroda, M. and Perera, F.P. (1991) Biologic markers in ethylene oxide-exposed workers and controls. Mutat. Res. 248, 163–176.

National Research Council, Committee on Advances in Assessing Human Exposure to Airborne Pollutants. (1991a) Human Exposure Assessment for Airborne Pollutants. Advances and Opportunities, National Academy Press, Washington, DC.

National Research Council, Committee on Environmental Epidemiology. (1991b) Environmental Epidemiology, National Academy Press, Washington, DC.

Osterman-Golkar, S, Bailey, E., Farmer, P.B., Gort, S.M. and Lamb, J.H. (1984) Monitoring exposure to propylene oxide through the determination of hemoglobin alkylation. Scand. J. Work Environ. Health 10, 99–102.

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

Perera, F.P., Santella, R.M., Brenner, D., Poirier, M.C., Munshi, A.A., Fischman, H.K. and Van Ryzkin, J. (1987) DNA adducts, protein adducts and sister chromatid exchange in cigarette smokers and nonsmokers. J. Natl. Cancer Inst. 79, 449–456.

Phillips, D.H., Hewer, A., Martin, C.N., Garner, R.C. and King, M.M. (1988) Correlation of DNA adduct levels in human lung with cigarette smoking. Nature 336, 790–792.

Randerath, E., Agrawal, H.P., Weaver, J.A., Bordelon, C.B. and Randerath, K. (1985) ^{32}P-Postlabeling analysis of DNA adducts persisting for up to 42 weeks in the skin, epidermis and dermis of mice treated topically with 7,12-dimethyl benz[*a*]anthracene. Carcinogenesis 6, 1117–1126.

Reddy, M.V. and Randerath, K. (1990) A comparison of DNA adduct formation in white blood cells and internal organs of mice exposed to benzo[*a*]pyrene, dibenzo[*c*,*g*]carbazole, safrole and cigarette smoke condensate. Mutat. Res. 241, 37–48.

Reddy, M.V. and Randerath, K. (1986) Nuclease PI-mediated enhancement of sensitivity of ^{32}P-postlabeling test for structurally diverse DNA adducts. Carcinogenesis 7, 1543–1551.

Roberts, D.W., Benson, R.W., Groopman, J.D., Flammang, T.J., Nagle, W.A., Moss, A.J. and Kadlubar, F.F. (1988) Immunochemical quantification of DNA adducts derived from the human bladder carcinogen 4-aminobiphenyl. Cancer Res. 48, 6336–6342.

Rosin, M.P. (1992) The use of the micronucleus test on exfoliated cells to identify anticlastogenic action in humans. Mutat. Res. 267, 265–276.

Ross, R.K., Yuan, J.M., Yu, M.C., Wogan, G.N., Qian, G.S., Tu, J.T., Groopman, J.D., Gao, Y.T. and Henderson, B.E. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. Lancet 339, 943–46.

Savela, K. and Hemminki, K. (1990) DNA adducts in leukocytes and granulocytes of smokers and nonsmokers detected by the ^{32}P-postlabelling assay. Carcinogenesis 12, 503–8.

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

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

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

Schulte, P.A. and Perera, F.P. (1993) Validation. In: P.A. Schulte and F.P. Perera, (Eds), Molecular Epidemiology: Principles and Practices, Academic Press, San Diego, CA, pp. 79–107.

Schurdak, M.E., Stong, D.B., Warshawsky, D. and Randerath, K. (1987) N-Methylation reduces the DNA-binding of ^{7}H-dibenzo(*c*,*g*)carbazole ~300-fold in mouse liver but only ~2-fold in skin: possible correlation with carcinogenic activity. Carcinogenesis 8, 1405–1410.

Schwartz, S.L. and Balter, N.J. (1988) Occupational exposure and health effects assessment criteria using physiological pharmacokinetic modeling. 4th International Symposium: Epidemiology in Occupational Health, August 1988, Stockholm, Sweden.

Skipper, P.L. and Tannenbaum, S.R. (1990) Protein adducts in the molecular dosimetry of chemical carcinogens. Carcinogenesis 11, 507–518.

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

Stowers, S.J. and Anderson, M.W. (1984) Ubiquitious binding of benzo(*a*)pyrene metabolites to DNA and protein in tissues of mouse and rabbit. Chem. Biol. Interact. 51, 151–166.

Talaska, G., Al-Juburi, A.Z.S.S. and Kadlubar, F.F. (1991) Smoking-related carcinogen-DNA adducts in biopsy samples of human urinary bladder: Identification of *N*-deoxyguanosin-8-yl-4-aminobiphenyl as a major adduct. Proc. Natl. Acad. Sci. USA 88, 5350–5354.

Talaska, G., Au, W.W., Ward, J.B. Jr., Randerath, K. and Legator, M.S. (1987) The correlation between DNA adducts and chromosomal aberrations in the target organ of benzidine exposed, partially-hepatectomized mice. Carcinogenesis 8, 1899–1905.

Talaska, G., Dooley, K.B. and Kadlubar, F.F. (1990) detection and characterization of carcinogen-DNA adducts in exfoliated urothelial cells from 4-aminobiphenyl-treated dogs by [32P]postlabelling and subsequent thin layer- and high pressure liquid chromatography. Carcinogenesis 11, 639–646.

Talaska, G., Schamer, M., Skipper, P., Tannenbaum, S., Caporaso, N., Kadlubar, F., Bartsch, H., Malaveille, C. and Vineis, P. (1992) Carcinogen-DNA adducts in exfoliated urothelial cells: techniques for non-invasive human monitoring. Environ. Health Perspect. 99, 289–91.

Umbenhauer, D., Wild, C.P., Montesano, R., Suffhill, R., Broyle, J.M., Huh, N., Kirsten, Y., Thomale, S., Rajewsky, M.F. and Lu, S.H. (1985) OG methyldeoxyguanosine in oesophageal DNA among individuals at high risk of oesophageal cancer. Int. J. Cancer 36, 661–665.

van de Poll, M.L.M., van der Hulst, D.A.M., Tates, A.D., Mulder, G.J. and Meerman, J.H.N. (1989) The role of specific DNA adducts in the induction of mucronuclei by *N*-hydroxy-2-acetylaminofluorene in rat liver in vivo. Carcinogenesis 10, 717–722

van Schooten, F.J., Hillebrand, M.J., van Leeuwen, F.E., Lutgerink, J.T., van Zandwijk, N., Jansen, H.M. and Kriek, E. (1990) Polycyclic aromatic hydrocarbon — DNA adducts in lung tissue from lung cancer patients. Carcinogenesis 11, 1677–81.

Van Sittert, N.J., de Jong, G., Clare, M.G., Davies, R., Dean, B.J., Wren, L.J. and Wright, A.S. (1985) Cytogenetic immunological and hematological effects in workers in an ethylene oxide manufacturing plant. Br. J. Indust. Med. 42, 19–26.

Van Zeeland, A.A., Mohn, G.R., Neuhauser-Klaus, A. and Ehling, U.H. (1985) Quantitative comparison of genetic effects of ethylating agents on the basis of DNA adduct formation. Use of *O*-ethylguanine as a molecular dosimeter

for extrapolation from cells in culture to the mouse. Environ. Health Perspect. 62, 163–169.

Vineis, P., Caporaso, N., Tannenbaum, S., Skipper, P., Glogowski, J., Bartsch, H., Coda, M., Talaska, G. and Kadlubar, F. (1990) Acetylation phenotype, carcinogen-hemoglobin adducts, and cigarette-smoking. Cancer Res. 50, 3002–3004.

Wattenberg, L.W. (1983) Inhibition of neoplasia by minor dietary constituents. Cancer Res. 43, 2448–2453.

WHO. (1986) Principles of Toxicokinetic studies. Environmental Health Criteria 57, World Health Organizaton, Geneva.

WHO. (1975) Early detection of health impairments in occupational exposure to health hazards. WHO technical report 571, World Health Organization, Geneva.

Wogan, G.N. and Gorelick, N.J. (1985) Chemical and biochemical dosimetry of exposure to genotoxic chemicals. Environ. Health Perspect. 62, 5–18.