A Method for Biomarker Validation and Biomarker-Based Dose Response: A Case Study with a Bayesian Network Model for Benzene

Haber LT, Maier MA, Hack CE, Shulte P, Fowler B, Lotz WG, Savage RE Jr.

Running Header: Biomarker Validation – Benzene Bayesian Network

ABSTRACT

To facilitate the use of biomarkers in human health risk assessments, a biomarker decision support system is presented that systematically identifies, documents, validates, and incorporates biomarkers into occupational risk assessments. A biomarker database structure was developed and decision rules were identified to organize the diverse types of data to support an occupational risk assessment for benzene. A suite of biomarker validation approaches was applied to evaluate potential biomarkers of exposure and effects of exposure for benzene-induced acute myeloid leukemia (AML). Traditional biomarker evaluation approaches based on the Hill criteria and regression analysis were coupled with a Bayesian network approach to validate (or discount) individual biomarkers and ultimately link the validated biomarkers along the exposure-disease continuum. Dose-response analyses using validated biomarkers were conducted to contrast various biomarker-based dose-response approaches. Although multiple analysis and validation approaches are described in this benzene case study, it is envisioned that the system will be most useful as a set of options allowing the user to choose the analytical approach(es) appropriate to the data set and needs of the analysis. Ultimately, this work aims to identify appropriate tools for critically evaluating biomarkers, as well as to provide a quantitative approach for linking changes in biomarkers of effect both to exposure information and to changes in disease response. Such linkage can provide a scientifically valid point of departure that incorporates precursor dose-response information without being dependent on the difficult issue of a definition of adversity for precursors.

Key Words: Biomarker, Bayesian, Benzene, Validation, OELs

INTRODUCTION

While it is difficult to trace the origin of the term "biomarker," it is clear that biomarkers have found widespread use in contemporary biomedical sciences. In 1989, the National Research Council on Environmental Studies and Toxicology defined a biomarker as any cellular or molecular indicator of toxic exposure, adverse health outcome or susceptibility (NAS/NRC, 1989). That definition paved the way for a further delineation of biomarkers into three categories – biomarkers of exposure, effect and susceptibility. While various disciplines foster unique applications of biomarkers to their special needs, a shared aspect is that the transformation from a measured biological endpoint to a validated biomarker with a specific public health application is a difficult and tedious process. Albertini (2001) noted that the mere availability of biomarkers does not mean that they will be useful for human studies directed at public health issues, and that although many biomarkers have been identified, few have been validated to the point of known usefulness (Albertini, 1998). Perera (2000) pointed out that although biomarkers can play a key role in cancer epidemiological studies, many studies failed to use validated biomarkers or employed study designs that did not adequately consider the biology of the endpoints. Established guidelines for biomarker validation exist (Barker, 2003; Schulte, 2005) and Schulte (2005) emphasized the importance of validation prior to use of biomarkers. However, lack of validation has continued to be a key barrier to increased biomarker application in human health risk assessment (Maier et al., 2004). The appropriate use of biomarker data is an ongoing issue in risk assessment, as noted by several authors (e.g., Paustenbach and Gailbraith, 2006).

To attempt to address this barrier, we describe a biomarker decision support system (Figure 1) to facilitate the systematic identification, documentation, validation, and incorporation of biomarkers into occupational risk assessment. The steps in the decision support system are a)

categorization of inputs; b) data evaluation; c) analysis and validation of biomarkers and dose-response; d) development of the risk assessment output. For the case study demonstrating the application of the decision support system, we extracted biomarker data from public databases or the appropriate literature,[1] integrated the data into a user-friendly biomarker database of our design, analyzed the biomarker data using a suite of validation tools, and demonstrated the contribution of biomarker data in the dose-response assessment for occupational exposure limit (OEL) development. We focus on application of the concept to occupational risk assessment, since biomarkers of exposure have a rich history of use in this area and because there is significant potential for biomarkers to enhance the scientific underpinnings in the development of OELs.

To test and demonstrate the decision support system for occupational safety and health applications, benzene was chosen as the case study chemical. Benzene was chosen based on (1) the potential for occupational exposure, (2) severity of toxic response, (3) availability and continuity of biomarker data (i.e., linkage to effects), (4) adequate knowledge of mode of action, and (5) the perceived need for an updated OEL, based on the availability of new data.

Ideally, exposure and effect biomarkers can be linked to predict dose-response behavior using biologically-based dose-response models that are structured based on detailed quantitative knowledge of the mode of action. In most cases, however, such a quantitative level of understanding is not available, and alternative approaches are needed. Thus, we demonstrate a suite of tools to validate and link biomarkers across the exposure-dose-disease continuum, ranging from qualitative analyses to a relatively novel Bayesian network analysis approach. For illustrative purposes, multiple analysis and validation approaches are described for the case study in this paper. However, it is envisioned that the system is most useful as a set of options, with

the user choosing the analytical approach(es) appropriate to the data set and needs of the analysis. In the case study, the Bayesian network is also used to quantitatively integrate data on biomarkers and effects in a dose-effect biomarker-disease response analysis to determine a point of departure based on data spanning the continuum, as opposed to the conventional approach of using either effect or biomarker data exclusively in the dose-response assessment. A goal of this work is to identify a range of tools for critically evaluating biomarkers, as well as to provide a quantitative approach for linking changes in biomarkers of effect (also called precursors to the adverse effect) both to exposure information and to changes in disease response. Such linkage can provide a scientifically valid point of departure that incorporates precursor dose-response information without being dependent on the difficult issue of a definition of adversity for precursors.

METHODS AND RESULTS

Biomarker Data Search

The biomarker data were extracted from a review of the data on biomarkers, toxicology, and information on the underlying biology of the case study chemical, benzene. Relevant studies were identified by searching the literature using the National Library of Medicine TOXNET cluster of databases and the Medline database, from recent comprehensive toxicology reviews, and by contacting researchers familiar with benzene toxicology.

Biomarker Database

The systematic documentation and integration of the diverse types of biomarker data obtained from multiple studies required developing a biomarker database. We developed a customized biomarker database in Microsoft Access® to house data related to benzene biomarkers and document study selection. The relational database included four levels: 1) general study information; 2) species or cell type investigated; 3) specific information on biomarkers reported; and 4) dose-response data.[2]

Categorization of Inputs

Studies were categorized as providing information on one or more biomarker types: exposure, internal dose, effective dose, early effects, mild effects, severe effects, or susceptibility. These categories have been well described elsewhere and we applied definitions consistent with those provided in risk assessment guidance documents (e.g., EPA, 1994). This initial categorization step provided documentation of the range of studies identified and served as an initial screening point to allow for additional targeted literature searches that focused on biomarkers of most interest.

The disease endpoints considered in this analysis were anemia, leukopenia, and acute myeloid leukemia (AML). Central nervous system (CNS) effects were not included in the analysis, since they are most relevant following acute high-level exposures and likely result from a different mechanism than that for the blood effects. Although the mechanism of action of benzene's effects on hematological outcomes is not fully known, there are several potential biomarkers of benzene exposure or of effects of exposure, and several have been linked to clinical disease. The identified biomarkers were classified as biomarkers of exposure (Table 1A) or effect (Table 1B).

The remaining steps in the decision support system were applied to a subset of the database consisting of those studies that met the following criteria: 1) measured endpoints had accompanying exposure estimates; 2) the endpoints evaluated were those most relevant to the diseases (adverse effects) under consideration, based on prior knowledge about possible modes of action; and 3) study quality was adequate to yield reliable estimates. In addition, the preference was to use studies with multiple endpoints spanning the exposure-disease continuum in the same population, but few studies that evaluated multiple endpoints across the continuum were available, and so this criterion was not used to exclude studies. A total of 27 studies were retained after applying these screening criteria for data adequacy to the larger pool of available studies identified in the published literature. Of these studies, 21 included exposure biomarkers and 12 included effects biomarkers, including the range of biomarkers in Tables 1A and 1B. (See Supplemental Material for additional study details.) The effect biomarkers were evaluated against the endpoints listed in Table 1C.

Data Evaluation

While there were multiple human studies providing data for most of the biomarkers of exposure (except for the albumin adducts and urinary benzene triol), for most of the biomarkers of effect, there was only one useful human study that tied the biomarker data to exposure measurements. In addition, for two of the effect categories, all or most of the available biomarker data came from only one study. Possible exposure biomarkers considered included benzene in the blood, urine, and exhaled air; urinary levels of the benzene metabolites trans,trans-muconic acid (ttMA), S-phenylmercapturic acid (SPMA), catechol, benzene triol, and phenol; and protein adducts formed by benzene metabolites.

A significant challenge in developing quantitative evaluation approaches that combine diverse data types is the need to format data in a manner that maintains the integrity of the data, but allows for combining datasets. For studies meeting the study adequacy criteria described above, biomarker data were normalized to facilitate inter-study data comparisons and the combining of studies for the validation analysis. This was done by converting common endpoints reported in different studies to the same response units, and then normalizing all data relative to some nominal response level.

Data were collected from several studies of different groups with different exposure scenarios and experimental designs. To enable quantitative analysis, the data were grouped into exposure categories determined by examining the range of exposure (as reported by the study authors) associated with the various candidate biomarker and disease observations. The data were ordered by increasing exposure concentration, and cut-points for the exposure categories were chosen with the goal of spreading the observations as evenly as possible over the exposure categories. For example, the exposure ranges were chosen so that observations of AML were in different exposure groups, rather than being lumped together. Using the ranges of exposures also facilitated the evaluation of the linkage across the exposure-disease continuum, since various biomarkers and effects were grouped together in a common exposure category. The eight resulting exposure groups were 0 up to 0.01 ppm, 0.01 up to 0.5 ppm, 0.5 up to 1.5 ppm, 1.5 up to 3 ppm, 3 up to 10 ppm, 10 up to 25 ppm, 25 up to 50 ppm, and 50 ppm and greater. While the selection of cut-points was an ad hoc exercise for this demonstration project, it is a typical meta-analysis approach for analysis of epidemiological data.

Biomarker Analysis and Validation

The generic decision support system was designed to accommodate the use of a suite of analysis validation tools. The particular tools used for a given chemical will vary depending on the nature of the dataset available. For the benzene case study, the analytical tools were qualitative analysis using the Hill criteria (EPA, 2005), qualitative graphical analysis, linear regression, and Bayesian network modeling. While use of the entire range of tools may not have been needed to identify appropriate biomarkers for this case study, the results illustrate the strengths and weaknesses of the different tools. In addition, sufficient data were available to use the power of the Bayesian network to quantitatively incorporate the biomarker data into the dose-response analysis. Other chemical data sets will require only simple regression tools to identify the appropriate biomarkers, or the potential mechanisms of action may not be sufficiently known to implement the more powerful quantitative approaches.

The various biomarkers of exposure and effect were ordered into pathways hypothesized based on knowledge of mechanisms of action to lead to the relevant endpoints. The initial identification of mechanistic pathways allowed for several alternative pathways from ambient exposure to benzene-induced anemia, leukopenia, and AML (Figure 2). Two pathways to anemia were considered. One pathway involved disruption of heme synthesis, and several biomarkers in the heme synthesis pathway were considered. The other possible pathway for benzene-induced anemia begins with decreases in blood progenitor cells, which may be caused directly by cytotoxicity mechanisms or secondary to genotoxicity. Decreases in progenitor cells could also lead to leukopenia. Three possible pathways to AML were considered, with multiple biomarkers considered for each pathway. The pathways considered involved (1) genotoxicity, (2) oxidative stress (which could lead indirectly to DNA damage), and (3) changes in gene expression related to cell proliferation. The potential pathways identified were based on current

information on mode of action hypotheses with putative biomarkers for which data amenable to analysis were available.[3]

The first approach for validation and analysis of the proposed pathways involved qualitative application of the Hill criteria of (1) consistency, (2) strength of association, (3) specificity, (4) temporal relationship, (5) biological plausibility/coherence, and (6) dose-response relationship (EPA, 2005). Biological plausibility/coherence was evaluated against current hypotheses in benzene mechanisms as articulated in several recent comprehensive toxicity reviews (Bird et al., 2005; Eastmond et al., 2001; Smith, 1996). All of the biomarkers under consideration passed this criterion. To evaluate the Hill criterion of temporality, one would ideally compare the duration of exposure required for changes in biomarker levels compared with the duration required for effects on disease incidence. This was not possible, however, given the nature of the datasets evaluated (human epidemiology studies with undefined or only longer-term exposures). In lieu of this information on temporal patterns of exposure and response, a qualitative approach was taken, using a general mode of action understanding to judge whether each biomarker would be expected to be affected prior to the clinical outcome of concern.

Specificity is the weakest of the Hill criteria, and the least likely to be fulfilled. As is commonly seen with effect endpoints and biomarkers, none of the effect biomarkers are specific to benzene. The specificity of several of the exposure biomarkers was reviewed by ACGIH (2001) in its consideration of biological exposure indices (BEIs). The measures of benzene internal dose, benzene in blood, urine, and breath, are specific to benzene. ACGIH (2001) considered the urinary levels of the benzene metabolites ttMA and SPMA to be sufficiently specific that both were adopted as recommended BEIs. Neither of these compounds are

themselves toxic agents, although ttMA is produced by the oxidation of the toxic intermediate trans,trans-muconaldehyde. In contrast to SPMA and ttMA, urinary levels of the benzene metabolites hydroquinone, catechol, 1,2,4-benzene triol, and phenol are not specific to benzene, because, in addition to reflecting benzene exposure, these levels may also reflect direct exposure to these chemicals or exposure to other chemicals that are metabolized to the same compounds. Albumin adducts of benzene oxide (BO) and 1,4-benzoquinone (BQ) are fairly specific, and reflect internal dose of reactive metabolites that have been proposed to cause the hematotoxic and leukemogenic effects of benzene. Because specificity is a weak criterion, this criterion alone was not used to exclude potential biomarkers.

Regression analyses were used to evaluate the consistency of the association between environmental exposure to benzene (in ppm) and biomarkers of exposure (Table 2). This was done by estimating the coefficient of determination[4] (R^2). The selection of potential biomarkers based on coefficients of determination is somewhat arbitrary, but, when coupled with the other Hill criteria, this approach can be informative. The regression analysis was conducted both for the entire concentration-response range observed, and for exposures only at lower concentrations in the range of regulatory or toxicological significance (i.e., up to approximately 1 ppm). Regression analysis was not performed for albumin adducts with 1,4-benzoquinone or benzene oxide, or for urinary benzenetriol, because there were only three observations and no observations below 1 ppm (except for the control values) for these metabolites.

The benzene metabolites were generally much better markers of exposure than were benzene in urine, breath, or blood (Table 2). Both urinary ttMA and SPMA correlated well with benzene exposure, consistent with the ACGIH (2001) recommendation of these metabolites as BEIs for benzene exposure. Correlation with benzene exposure at low concentrations was better

for urinary ttMA than for urinary SPMA. Correlation at low benzene exposures was poor for other metabolites, consistent with the ACGIH (2001) conclusion that those metabolites were not specific and sensitive enough in the low exposure range. Overall, only two biomarkers of exposure had acceptable correlation coefficients both overall and in the low concentration range, urinary ttMA and benzene in blood (BB). ACGIH (2001) considered benzene in blood to be a specific and sensitive biomarker, but noted that it is markedly affected by current or recent exposure, rather than average or cumulative exposure. Based on these considerations, urinary ttMA and BB were included in the initial Bayesian network modeling described below; only ttMA was chosen for the final network.

To evaluate the Hill criteria of consistency and dose-response for biomarkers of effect, the values of such markers from multiple studies were plotted together and graphical comparisons were used to evaluate the various candidate precursors to AML (Figure 3), as well as for anemia and leukopenia (data not shown). As noted earlier, data were normalized to facilitate presenting multiple different endpoints on the same graph. For many of the biomarkers of effect, only one appropriate human study was available, limiting the ability to evaluate consistency. Moreover, in considering the dose-response, a key decision was choosing an appropriate dose metric; a common dose metric was needed in order to analyze the relationships among the various endpoints. AML incidence is generally analyzed as a function of cumulative exposure (e.g., ppm-years of exposure). On the other hand, measured values of exposure biomarkers are more likely associated with shorter term exposure than with cumulative exposure, due to clearance of metabolites from the body, the turnover rate of blood cells where the markers where observed, or other recovery mechanisms during periods with no exposure. Assuming that average exposures are relatively constant over periods of months to years - despite potential

acute exposure fluctuations - using chronic time-weighted average (TWA) exposure will be a good surrogate for both cumulative and short-term exposures of interest. This assumption is appropriate for risk assessments for the types of effects of interest for benzene that are likely to be due to toxic mechanisms resulting from longer-term cumulative exposures. Therefore, the analyses shown in Figure 3 were based on TWA, using either average exposure concentration data reported by the authors or using average exposure duration to convert from ppm-years to TWA. Several of the biomarkers considered had dose-response curves that were not consistent across studies or across dose groups within a study; these were excluded from further validation and, for clarity, are generally not shown in Figure 3.

The graphical analysis of the biomarkers of effect was a fairly crude qualitative evaluation, based on the shape and position of the biomarker response curve relative to the disease response curve. We considered this qualitative information more informative than regression analyses for the effects biomarkers, in light of the limited data for each biomarker and the inconsistencies in the leukemia dose-response (even though all of the dose-response data for leukemia were from a single study[5]). A key consideration is that the biomarker curve should rise before, that is to the left of, the disease response curve, but in order to be predictive of disease, it must not plateau or level off before an increase in the disease response is apparent. Despite the limitations of this approach, some broad conclusions are possible, based on comparison of the biomarker curves with that for leukemia.

Based on inspection of Figure 3, the most consistent precursors to AML are the genotoxicity markers of hyperdiploidy in chromosome 7, micronuclei, and the oxidative stress marker 8-OHdG in lymphocytes. Each of the curves for these three markers exhibited smooth dose-response curves that appear to rise above background levels slightly before an increase in

AML incidence is observed. The micronuclei and 8-OHdG curves also have some data at higher doses where these markers appear to plateau or level off, which improves the dose-response characterization, since steep dose-response curves at high doses may make predictions at high doses more uncertain. The hyperdiploidy and aneusomy data are somewhat limited, with only three data points, missing both a plateau and data in the range where AML incidence begins to increase. This lack of key data introduces uncertainty and degrades the dose-response characterization. Furthermore, aneusomy appears to begin increasing at doses higher than those that cause AML, and the dose-response curves for aneusomy and AML clearly cross, which would not be expected if aneusomy causes AML. The single strand breaks (SSB) and breakage at chromosome 9 had dose-response curves that were not inconsistent with AML, but these curves increased very sharply (and before any apparent increase in AML incidence in the case of SSB) and were not observed to peak or level off, limiting their utility for dose-response modeling. Based on this application of the Hill criteria, micronuclei, hyperdiploidy, and 8-OHdG were identified as the best AML precursors from the available datasets for use as candidates for the next phase of the validation scheme (Bayesian network analysis, see below).

Red blood cell count was the marker chosen to represent the anemia endpoint, since the red blood cell abnormalities did not have a consistent dose-response. Based on consideration of the biology of erythropoiesis, which was generally consistent with the graphical analysis for the anemia pathway, the Bayesian network model included effects on one marker from the heme synthesis pathway (ALA in erythrocytes), hemoglobin, and the progenitor cells.

There was considerable inter-study variability in the observations of white blood cell counts (WBC). As with the analysis for anemia, the blood progenitor cells behaved in a manner consistent with the biology of hematopoiesis, and both precursor cell types for which there was

adequate data, the “colony-forming unit -- granulocyte, erythrocyte, monocyte, megakaryocyte” cells (CFU-GEMM), and “colony-forming units – granulocyte, macrophage” (CFU-GM), were included in the Bayesian network. The one remaining Hill criterion, strength of association, was evaluated by examining parameters estimated for a Bayesian network, as described in the next section.

Bayesian network

Since data were available on endpoints spanning the exposure-response continuum, with multiple plausible pathways, we used a validation and dose response approach based on Bayesian network modeling.

The Bayesian network approach is described in detail in Appendix B in the supplemental material and was constructed using MCSim (Bois et al., 2005) with supporting statistical analyses in Microsoft Excel®; the model code is presented in Appendix C of the supplemental material. The network (Figure 2) was based on the hypothesized mode of action of hematotoxicity, the availability of adequate data from the screened pool of potential studies, and the results of the Hill criteria-based biomarker evaluation discussed above. Benzene concentration and exposure biomarkers (i.e., urinary ttMA and benzene in blood) are combined into a single node in the schematic shown in Figure 2 for clearer presentation of the network, although separate arrows from benzene concentration and the exposure markers to the next level of early effect markers are present in the actual network. In addition, the coded network allows for benzene concentration to act directly upon the effect markers so that it is possible for the validation analysis to reject poor biomarkers of exposure.

As described further in Appendix B, the Bayesian network is a series of power (for exposure markers only) and logistic regression equations describing the relationships between the nodes in the network (i.e., between the "steps" of the pathway, see Figure 2). Briefly, each marker or disease endpoint (i.e., each node) shown in the network (except for benzene concentration in air) is represented by an equation with a constant term, and one slope term for each of the node's parents (i.e., the nodes from which the arrows originate). The network was calibrated, or fit to the data, via Markov Chain Monte Carlo (MCMC) simulation to find distributions for the parameters of the regression equations. This calibration was followed by several analyses ("diagnostics") performed to choose among competing biomarkers (e.g., different markers of genotoxicity) and to identify elements of the network of potential biomarkers that are most predictive of disease.

The one remaining Hill criterion not addressed above, the benzene – exposure marker – effect marker – disease strength of association, was quantitatively evaluated by examining the estimates of the slope parameters in the Bayesian network. Generally speaking, larger means and smaller variances for estimates of the slope parameters inter-relating the nodes of the network indicate stronger associations between the nodes. (See Tables B1 and B2 of Appendix B in the supplemental material.) Considerations in evaluating the strength of the association (and the confidence in that association) between benzene exposure, the various markers, and the disease outcomes included the sign and magnitude of the slope, the probability that the slope is positive, the standard error or uncertainty for the slope, and the correlation coefficients (not shown) between the disease and the various biomarkers. All of these analyses were conducted based on the posterior distributions determined from the Bayesian analysis. However, inclusion of nodes in the final network also took into account other considerations in addition the

magnitude of the slope parameter. In some cases, the link between two markers was very strong, but there were very weak associations both up and down the network from the two closely associated markers. In other cases, markers were included in the network based on the biological plausibility that they belonged in the mechanistic pathway to disease (e.g., BFU-E), even though the associations with its parents and daughters in the network were not the strongest.

Figures 4 and 5 illustrate the evaluation of the probability that the slope from a precursor to disease is greater than zero when the distribution for that parameter is assumed to be normal. For the markers shown in figures 4 and 5, the slope is expected to be positive, and a higher probability indicates greater support for that assumption, while a smaller probability (e.g., 50% or less) reflects weaker association between the precursor and the disease. In this case, the test indicates that 8-OHdG (Figure 5) is a better biomarker than micronuclei (Figure 4), because the slope parameter from 8-OHdG to AML is more likely to be positive (i.e., probability of 85% vs. 61%). The standard deviation of the slope parameters relative to their means were also compared (when a lognormal distribution was assumed) to help choose among competing biomarkers. The analyses using the normal and lognormal assumptions were generally consistent regarding the biomarkers that were most appropriate to include in the network. Although the slope from 8-OHdG to AML did not have the greatest magnitude, it did have the smallest standard deviation relative to the mean, and was again indicated to be the best early effect biomarker for AML. Thus, based on the database and decision support system used, 8-OHdG was determined to be the best precursor of AML amongst the biomarkers examined.

The diagnostics were less conclusive for validating biomarkers along the pathways to anemia and leukopenia. This is likely due to sparse data along the pathway to anemia (e.g., for the heme synthesis pathway) and inconsistent dose-response data for the white blood cell counts.

However, based on the graphical comparison and knowledge of the mode of benzene toxicity, the pathways through the progenitor cells were also included in the final network of biomarkers. 8-OHdG was also linked to CFU-GEMM in the final biomarker network based on the biological argument that oxidative stress or genotoxicity (for which 8-OHdG may be an indicator) could lead to reduced blood progenitor cell colony formation. Since the network modeling of anemia and leukopenia was not very successful, these pathways are not discussed further in this paper.

The resulting network of biomarkers with the strongest linkages to disease, using the diagnostics described above and the biological considerations just presented, is shown as the shaded pathway in Figure 2.

Dose-Response Analyses

Dose-response analyses for AML incorporating the validated biomarkers of exposure and effects were conducted using four approaches (Table 3). The first three approaches were basically benchmark concentration modeling, using different dependent and independent variables, and corresponding definitions of the benchmark response (BMR). First, a logistic model was used to directly evaluate the relationship between the concentration of benzene in air and AML. This dose-response was conducted to provide a frame of reference for evaluating the impact of incorporating the biomarkers. Second, a logistic model was used to evaluate the relationship between exposure biomarkers and AML, in order to derive the effective exposure biomarker value corresponding to the theoretical risk often used as the basis for development of OELs (i.e., 1/1000 incidence). This effective exposure biomarker was then converted to an equivalent benzene exposure level using a power regression model relating the exposure biomarker to benzene concentration in air. The third approach was analogous to the first

approach, except that the dependent variable was the validated precursor immediately preceding AML (8-OHdG). The degree of precursor response used as the BMR was one control standard deviation from the control mean response (EPA, 2000). The power and logistic models for the second and third approaches were fitted using maximum likelihood techniques in EXCEL. In the fourth approach, a Monte Carlo analysis of the Bayesian network model was conducted using MCSim (Bois et al., 2005) to calculate the benzene concentration that, when propagated through the biomarker network, produced an AML response rate of 1/1000. The results of these alternative biomarker-based dose-response analyses for the AML endpoint are shown in Table 3.

As shown in Table 3, the median effective concentration of benzene in air (ECs) derived using the biomarker of exposure (i.e., urinary ttMA) and the precursor to disease (8-OHdG) are within a factor of two of the reference EC from the model of benzene in air versus disease. However, the network analysis results in an EC that is significantly lower than the reference EC from modeling benzene in air versus disease. The network dose-response modeling approach was also attempted using hyperdiploidy as the key precursor to AML. The resulting dose-response (data not shown) was much more uncertain than that obtained using 8-OHdG. This is further support of the conclusion that 8-OHdG is the preferred biomarker based on these data, although a definitive conclusion is not possible, due to the limitations of the data on biomarkers of genotoxicity used in the model.

DISCUSSION

The previous section demonstrated a systematic approach for comparing and choosing among exposure or effect biomarkers along the continuum from exposure to disease and for incorporating the chosen biomarkers in an exposure-biomarker-disease response analysis to

compute biomarker-based points of departure, ECs, for OEL derivation. We have presented a case study to illustrate the utility of the different tools. Additional work would be needed to fully develop a reliable OEL.

In the case study of benzene, for each endpoint (AML, anemia, and leucopenia), a reference EC was estimated using the conventional approach of regressing disease response on benzene concentration in air. Analyses using biomarker-based approaches were then conducted and compared to the conventional approach. The values reported here are not intended to correspond directly to current OELs for several reasons: the calculated ECs are best estimates rather than lower confidence bounds, as would often be used in a full risk assessment; uncertainty factors were not considered; alternative biological models were not considered; and the ECs arise directly from data analysis and do not consider risk policy considerations often embedded in OEL determinations.

The precursor-based EC, which was based on directly modeling the air concentration vs. the precursor data, is smaller than the reference EC. This is the expected result for an evaluation of the concentration vs. precursor, since the precursor event occurs at a lower dose than the disease (or perhaps the same dose), and cannot have a higher threshold. Issues surrounding the identification of relevant precursor events for direct use in risk assessment, and the definition of adversity have been a hindrance in developing consensus on risk values. Recent risk assessments, such as for perchlorate (Ginsberg and Rice, 2005; Johnston et al., 2005; Strawson et al., 2005), provide an example of the debate surrounding this issue.

The network-based approach presented here does not remove the need for clear articulation of causal proximity of a validated biomarker to the adverse effect of interest or for consideration of the degree of change in the ultimate endpoint of interest that is adverse.

Nevertheless, at the very least, it provides a systematic approach to determine whether a specific precursor event can be considered validated and thus become a candidate for this type of articulation. Moreover, the use of a Bayesian network provides a quantitative approach for linking changes in precursors both to exposure information and to changes in disease response, thus providing a scientifically valid point of departure that incorporates precursor dose-response information without being dependent on the difficult issue of a definition of adversity for precursors when they are suggested for use directly as a point of departure. The Bayesian network also provides a means to improve the dose-response estimate for the final endpoint by taking into account the dose-response information for earlier steps in the pathway leading to that endpoint.

In our case study, for AML, the network dose-response model yielded a median estimate of the EC that was approximately an order of magnitude smaller than the reference EC based on the traditional dose-response of benzene in air versus AML. This was apparently due to the presence of more information in the low-dose region where changes in 8-OHdG are detectable but effects on AML mortality are not. The network incorporates all of the information from the exposure-response continuum, and the result suggests that incorporation of the more sensitive early effect markers (i.e., 8-OHdG) may help inform the AML response at lower exposures. In this case, this results in a smaller estimate of the EC, though this will not be true in all cases, since it could result in higher estimates for other compounds.

The use of 8-OHdG as a biomarker of oxidative DNA damage has met with some controversy due to uncertainties regarding human variability in background levels of oxidative DNA damage, the potential confounding of analyses due to *ex vivo* DNA oxidation, variability in the sensitivity of alternate detection methods, and the absence of a clear dose-response in some

epidemiology studies (Collins et al., 2004; Pilger and Rudiger, 2006; Angerer et al., 2007). Despite these complications, 8-OHdG remains a commonly used biomarker in many current studies. Moreover, the benzene exposure study that was used for the quantitative analysis (Liu et al., 1996) in our study showed that a significant dose-response pattern for 8-OHdG was achieved, and reported that inter-laboratory variability does not preclude the importance of trends observed within a specific study. Finally, 8-OHdG was one of several potential genotoxicity endpoints that were found to be useful for developing the dose-response analysis, and this consistency in the findings provides added confidence in our results.

Bayesian networks have received much attention recently in the field of human health toxicology for application in the identification and analysis of gene regulatory networks (Dojer et al., 2006; Imoto et al., 2003; Toyoshiba et al., 2004; Toyoshiba et al., 2006). However, to our knowledge, our application for identifying relevant linkages of biomarkers along the exposure-disease continuum to support dose-response assessments has not been adequately explored.

Although we have demonstrated the utility of this approach in the benzene case study, there were several limitations in that assessment. The human data were too limited to evaluate the impact of several other biologically-plausible precursors to AML in the overall model. This observation is useful as a validation tool as it identifies key biomarkers as candidates for additional data collection. Even for 8-OHdG, only one set of adequate data in humans that linked biomarker and exposure data was available, limiting the degree to which consistency could be evaluated. Thus, there are a variety of extensions of this proof of concept examination that might be needed for broader applicability of this method. For example, we limited the analysis to in vivo studies, studies with human data, and exposure and effect biomarkers (but not susceptibility biomarkers). It is possible that endpoints not analyzed in this case study due to

inadequate human exposure-response data for validation are actually important events in the exposure-response continuum, and therefore could be good biomarkers of exposure or effects for benzene. Significant data on benzene biomarkers has been collected in animals and in mechanistic in vitro studies (e.g., Au et al., 2002; Bird et al., 2005; Faiola et al., 2004). There is opportunity for an enhancement to the current method related to combining data from these diverse sources.

PBPK and biologically-based dose-response models are likely to be important for such enhancements. The approach using Bayesian network models presented above should allow inclusion of biomarkers in the dose response for chemicals lacking complex biological models, but refinements of the linkages in the network can be supplied as such models are defined. For example, PBPK models may define the linkages between exposure and biomarkers of exposure, and allow the integration of data across multiple species. Such models, including the recently published human PBPK model for benzene (Yokley et al., 2006), could be used to refine the analysis presented here.

The methods demonstrated here showcase the potential impact of using biomarker information to estimate points of departure in support of the OEL development process. The Bayesian network technique, in particular, appears to be a novel approach for validation and use of biomarker data; it provides a scientifically and mathematically sound rationale for selection and analysis of biomarkers for occupational safety and health applications. Extending the approaches we developed here to conduct a more complete analysis, whether for benzene or for other compounds, including human, experimental animal, and in vitro data, will be a focus of future work.

REFERENCES

ACGIH (American Conference of Governmental Industrial Hygienists). (2001). Documentation of the biological exposure indices. 7th Edition. ACGIH Worldwide.

Albertini, R. J. (1998). The use and interpretation of biomarkers of environmental genotoxicity in humans. *Biotherapy* **11,** 155-167.

Albertini, R. J. (2001). Developing sustainable studies on environmental health. *Mutat. Res.* **480-481,** 317-331.

Angerer, J., Ewers, U., Wilhelm, M. (2007). Human biomonitoring: State of the art. *Int. J. Hyg. Environ. Health*. **210**, 201-228.

Au, W. W., Oberheitmann, B., Harms, C. (2002). Assessing DNA damage and health risk using biomarkers. *Mutat. Res*. **509**, 153-163.

Barker, P. E. (2003). Cancer biomarker validation: standards and process: roles for the National Institute of Standards and Technology (NIST). *Ann. N. Y. Acad. Sci.* **983**,142-150.

Bird, M.G., Grerim, H., Snyder, R., Rice, J.M. (2005). International symposium: Recent advances in benzene toxicity. *Chem. Biol. Interact*. **153-154**, 1-5.

Bogadi-Sare, A., Brumen, V., Turk, R., Karacic, V., Zavalic, M. (1997). Genotoxic effects in workers exposed to benzene: with special reference to exposure biomarkers and confounding factors. *Ind. Health*. **35**, 367-373.

Bois, F., Maszle, D., Revzan, K., Tillier, S., and Yuan, Z. (2005). MCSIM: logiciel de simulation [MCSim Simulation Package]. Experimental Toxicology Unit, INERIS, Verneuil-en-Halatte, France. http://toxi.ineris.fr/en/activites/toxicologie_quantitative/mcsim/mcsim.php

Boogaard, P. J., and Van Sittert, N. J. (1995). Biological monitoring of exposure to benzene: a comparison between S-phenylmercapturic acid, trans,trans-muconic acid, and phenol. *Occup Environ Med*. **52**(9), 611-620.

Collins, A.R., Cadet, J., Moller, L., Poulsen, H.E., Vina, J. (2004). Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? *Arch. Biochem. Biophys.* **423**, 57-65.

Crump, K. and Allen, B. (1985). Methods for quantitative risk assessment using occupational studies. *Amer. Stat.* **39**, 442-450.

Dojer, N., Gambin, A., Mizera, A., Wilczynski, B., Tiuryn, J. (2006). Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics* **8**, 7:249.

Eastmond D. A., Schuler, M., Frantz, C., Chen, H., Parks, R., Wang, L., Hasegawa, L. (2001). Characterization and mechanisms of chromosomal alterations induced by benzene in mice and humans. HEI Research Report.

Egeghy, P. P., Tornero-Velez, R., and Rappaport, S. M. (2000). Environmental and biological monitoring of benzene during self-service automobile refueling. *Environ. Health Perspect.* **108** (12), 1195-1202.

EPA (United States Environmental Protection Agency). (1994). Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry. EPA/600/8-90/066F.

EPA (United States Environmental Protection Agency). (2000). Benchmark dose technical guidance document [external review draft]. Risk Assessment Forum, Washington, DC; EPA/630/R-00/001. Available from: <http://www.epa.gov/iris/backgr-d.htm>.

EPA (United States Environmental Protection Agency). (2005). Guidelines for carcinogen risk assessment. Risk Assessment Forum, Washington, DC; EPA/630/P-03/001B.

Faiola, B., Fuller, E., Wong, V.A., Pluta, L., Abernethy, D.J., Rose, J., Recio, L. (2004). Exposure of hematopoietic stem cells to benzene or 1,4-benzoquinone induces gender-specific gene expression. *Stem Cells*. **22**, 750-758.

Forrest, M. S., Lan, Q., Hubbard, A. E., Zhang, L., Vermeulen, R., Zhao, X., Li, G., Wu, Y., Shen, M., Yin, S., Chanock, S. J., Rothman, N., and Smith, M. T. (2005). Discovery of Novel Biomarkers by Microarray Analysis of Peripheral Blood Mononuclear Cell Gene Expression in Benzene-Exposed Workers. *Environ. Health Perspect*. **113**(6), 801-807.

Fustinoni, S. Buratti, M., Campo, L., Colombi, A. Consonni, D. Pesatori, A. C., Bonzini, M., Farmer, P., Garte, S., Valerio, F., Merlo, D. F., Bertazzi, P. A. (2005). Urinary t,t-muconic acid, S-phenylmercapturic acid and benzene as biomarkers of low benzene exposure, *Chem. Biol. Interact.* **153-154**, 253-256.

Garte, S., Popov, T., Georgieva, T., Bolognesi, C., Taioli, E., Bertazzi, P., Farmer, P., and Merlo, D. (2005). Biomarkers of exposure and effect in Bulgarian petrochemical workers exposed to benzene. *Chem. Biol. Interact*. **153-154**, 247-251.

Ginsberg, G., and Rice, D. (2005). The NAS perchlorate review: questions remain about the perchlorate RfD. *Environ Health Perspect*. 2005 Sep;**113**(9), 1117-1119. Erratum in: *Environ Health Perspect*. 2005 Nov;**113**(11), A732.

Hotz, P., Carbonnelle, P., Haufroid, V., Tschopp, A., Buchet, J.P., Lauwerys, R. (1997). Biological monitoring of vehicle mechanics and other workers exposed to low concentrations of benzene. *Int. Arch. Occup. Environ. Health*. **70,** 29-40.

Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., Miyano, S. (2003). Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *Proc. IEEE Comput. Soc. Bioinform. Conf.* **2**, 104-13.

Javelaud, B., Vian, L., Molle, R., Allain, P., Allemand, B., Andre, B., Barbier, F., Churet, AM., Dupuis, J., Galand, M., Millet, F., Talmon, F., Touron, C., Vaissiere, M., Vechambre, D., Vieules, M., and Viver, D. (1998). Benzene exposure in car mechanics and road tanker drivers. *Int. Arch. Occup. Environ. Health.* **71**, 277-283.

Johnston, R. B. Jr, Corley, R., Cowan, L., Utiger, R. D. (2005). The NAS perchlorate review: Adverse effects? *Environ. Health. Perspect.* **113**(11), A728-9.

Lan, Q., Zhang, L., Li, G., Vermeulen, R., Weinberg, R.S., Dosemeci, M., Rappaport, S.M., Shen, M., Alter, B.P., Wu, Y., Kopp, W., Waidyanatha, S., Rabkin, C., Guo, W., Chanock, S., Hayes, R.B., Linet, M., Kim, S., Yin, S., Rothman, N. and Smith, M.T. (2004). Hematotoxicity in workers exposed to low levels of benzene. *Science* Dec 3; **306**(5702):1774-1776.

Liu, L., Zhang, Q., Feng, J., Deng, L., Zeng, N., Yang, A. and Zhang, W. (1996). The study of DNA oxidative damage in benzene-exposed workers. *Mutat. Res.* **370**(3-4): 145-150.

Maier, A., Savage, R. E. Jr., Haber L. T.. (2004). Assessing biomarker use in risk assessment - a survey of practitioners. *J. Toxicol. Env.Health* Part A, **67**, 687-695.

Muzyka, V., Bogovski, S., Viitak, A., and Veidebaum, T. (2002). Alterations of heme metabolism in lymphocytes and metal content in blood plasma as markers of diesel fuel effects on human organism. *Sci. Total Environ.* **286**(1-3), 73-81.

NAS/NRC. 1989. Report of the oversight committee. In: Biologic markers in reproductive toxicology. Washington, DC: National Academy of Sciences, National Research Council, National Academy Press.

Ong, C. N., Kok, P. W., Lee, B. L., Shi, C. Y., Ong, H. Y., Chia, K. S., Lee, C. S. and Luo, X. W. (1995). Evaluation of biomarkers for occupational exposure to benzene. *Occup. Environ. Med.* **52**(8), 528-533.

Ong, C. N., Kok, P. W., Ong, H. Y., Shi, C. Y., Lee, B. L., Phoon, W. H., Tan, K. T. (1996). Biomarkers of exposure to low concentrations of benzene: a field assessment. *Occup Environ Med.* **53**, 328-333.

Paustenbach, D., Galbraith, D. Biomonitoring and biomarkers: Exposure assessment will never be the same. *Environmental Health*, **114**(8), 1143-1149.

Perera, F. P. (2000). Molecular epidemiology: on the path to prevention? *J. Natl. Cancer Inst.* **92**(8), 602-612.

Pilger, A, and Rudiger, H.W. (2006). 8-Hydroxy-2'-deoxyguanosine as a marker of oxidative DNA damage related to occupational and environmental exposures. *Int. Arch. Occup. Environ. Health*. **80**, 1-15.

Popp, W., Rauscher, D., Muller, G., Angerer, J., and Norpoth, K. (1994). Concentrations of benzene in blood and S-phenylmercapturic and t,t-muconic acid in urine in car mechanics. *Int. Arch. Occup. Environ. Health.* **66**(1), 1-6.

Qu, Q., Melikian, A., Li, G., Shore, R., Chen, L., Cohen, B., Yin, S., Kagan, M., Li, H., Meng, M., Jin, X., Winnik, W., Li, Y., Mu, R., and Li, K.. (2000). Validation of biomarkers in humans exposed to benzene: Urine metabolites. *Am J Ind Med*. **37**, 522-531.

Rappaport, S., Waidyanatha, S., Yeowell-O'Connell, K., Rothman, N., Smith, M., Zhang, L., Qu, Q., Shore, R., Li, G., and Yin, S. (2005). Protein adducts as biomarkers of human benzene metabolism. *Chem. Biol. Interact.* **153-154**, 103-109.

Rothman, N., Haas, R., Hayes, R., Li, G., Wiemels, J., Campleman, S., Quintana, P., Xi, L., Dosemeci, M., Titenko-Holland, N., Meyer, K., Lu, W., Zhang, L., Bechtold, W., Wang, Y., Kolachana, P., Yin, S., Blot, W., and Smith, M.. (1995). Benzene induces gene-duplicating but not gene-inactivating mutations at the glycophorin A locus in exposed humans. *Proc. Natl. Acad. Sci.* **92**, 4069-4073.

Rothman, N., Li, G.L., Dosemeci, M., Bechtold, W.E., Marti, G.E., Wang, Y.Z., Linet, M., Xi, L.Q., Lu, W., Smith, M.T., Titenko-Holland, N., Zhang, L.P., Blot, W., Yin, S.N. and Hayes, R.B. (1996). Hematotoxicity among Chinese workers heavily exposed to benzene. *Am. J. Ind. Med.* **29**(3), 236-246.

Sarto, F., Cominato, I., Pinton, A. M., Brovedani, E. M., Peruzzi, M., Bianchi, V., Levis, AG. (1984). A cytogenic study on workers exposed to low concentrations of benzene. *Carcinogenesis* **5**(6), 827-832.

Schulte, P. A. (2005). The use of biomarkers in surveillance, medical screening and intervention. *Mutat. Res.* **592**(1-2), 155-163.

Sherwood, R. J., and Sinclair, G. C. (1999). New PBPK model applied to old occupational exposure to benzene. *AIHA J*. **60**, 259-265.

Smith, M. T. (1996). The mechanism of benzene-induced leukemia: a hypothesis and speculations on the causes of leukemia. *Environ. Health Perspect.* **104** Suppl 6,1219-25.

Stommel, P., Muller, G., Stucker, W., Verkoyen, C., Schobel, S. and Norpoth, K. (1989). Determination of S-phenylmercapturic acid in the urine- an improvement in the biological monitoring of benzene exposure. *Carcinogenisis*, **10**(2), 279-282.

Strawson, J., Dourson, M.L., Zhao, Q.J. (2005). The NAS perchlorate review: is the RfD acceptable? Environ. Health Perspect. **113**(9), 1117-1119.

Toyoshiba, H., Yamanaka, T., Sone, H., Parham, FM., Walker, NJ., Martinez, J., Portier, CJ. (2004). Gene interaction network suggests dioxin induces a significant linkage between aryl hydrocarbon receptor and retinoic acid receptor beta. *Environ. Health. Perspect.* **112**(12), 1217-1224.

Toyoshiba, H., Sone, H., Yamanaka, T., Parham, F. M., Irwin, R. D., Boorman, G. A., Portier, C. J. (2006). Gene interaction network analysis suggests differences between high and low doses of acetaminophen. *Toxicol. Appl. Pharmacol*. **215**(3), 306-316.

Verdina, A., Galati, R., Falasca, G., Ghittori, S., Imbriani, M., Tomei, F., Marcellini, L., Zijno, A., Del Vecchio, V., Crebelli, R. (2001). Metabolic polymorphisms and urinary biomarkers in subjects with low benzene exposure. *J. Toxicol. Environ. Health Part A*. **64**, 607-618.

Waidyanatha, S., Rothman, N., Fustinoni, S., Smith, M.T., Hayes, R.B., Bechtold, W., Dosemeci, M., Guilan, L., Yin, S. and Rappaport, S.M. (2001). Urinary benzene as a biomarker of exposure among occupationally exposed and unexposed subjects. *Carcinogenesis.* **22**(2), 279-286.

Wiwanitkit, V., Suwansaksri, J., Nasuan, P. (2001). Urine trans,trans-muconic acid as a biomarker for benzene exposure in gas station attendants in Bangkom, Thailand. *Ann. Clin. Lab. Sci.* **31**(4), 399-401.

Yokley, K., Tran, H.T., Pekari, K., Rappaport, S., Riihimaki, V., Rothman, N., Waidyanatha, S., Schlosser, P.M. (2006). Physiologically-based pharmacokinetic modeling of benzene in humans: A Bayesian approach. *Risk Analysis*, **26**(4), 925-943.

Zhang, L., Eastmond, D. A., Smith, M. T. (2002). The nature of chromosomal aberrations detected in humans exposed to benzene. *Crit Rev Toxicol.* **32**(1),1-42.

Footnotes

[1] We initially sought to evaluate a diversity of biomarkers for benzene, including animal studies and *in vitro* data, in light of the pilot nature of the study. As a result of the wealth of data for benzene, we decided to exclusively study biomarkers derived from human studies as a simplified approach for this case study.

[2] Copies of the database are available from the authors upon request. Although only human data were used for this demonstration project, the biomarker database was developed to accommodate data from animal and in vitro studies as well.

[3] The choices of potential biomarkers were limited by the decision, noted earlier, to use only in vivo human data. However, this approach does not preclude the involvement of other relevant biological events, which were not included in the analysis due to the lack of adequate dose-response data.

[4] The coefficient of determination is a goodness of fit statistic from ordinary least squares regression (range 0 to 1) interpreted as the proportion of variation observed in the dependent variable (e.g., urinary metabolite level) that can be explained by the regression model. In other words, it estimates the proportion of the "error" that is explained by the model.

[5] Data were from the pliofilm cohort (Crump and Allen, 1985). Individual data for this cohort were generously provided by Dr. Crump and Dr. Van Landingham at ENVIRON.

Acknowledgements
The authors express their appreciation to Bruce Allen for his critical comments on the manuscript, and to Oliver Kroner, Melissa Kohrman, and Alison Willis for technical assistance. This work was funded by GSA Contract No. GS-10F-0369N from NIOSH.

Figure Legends

Figure 1. Conceptual representation of the biomarker decision support system. The sorting process is an iterative one, involving both judgments made prior to the analysis, and quantitative analysis to affect judgments.

Figure 2. Network of candidate biomarkers of benzene exposure and effect. The structure of the network was based on the literature on benzene biology and hypothesized modes of action. As described in the text, benzene concentration and exposure biomarkers (i.e., urinary ttMA and benzene in blood) are combined into a single node in the schematic shown in Figure 2 for clearer presentation of the network, although separate arrows from benzene concentration and the exposure markers to the next level of early effect markers are present in the actual network. The biomarker of exposure included in the final network (ttMA) is bolded. In addition, the coded network allows for benzene concentration to act directly upon the effect markers so that it is possible for the validation analysis to reject poor biomarkers of exposure. Gray highlights show the final network of biomarkers of benzene exposure and effect with the strongest linkages to disease, based on analysis of the full network of candidate biomarkers. ALA = delta-aminolevulinic acid; 8-OHdG = 8-hydroxyguanosine; CFU-GEMM = colony-forming unit -- granulocyte, erythrocyte, monocyte, megakaryocyte (a precursor cell type): BFU-E = burst-forming unit – erythroid; CFU-GM = colony-forming units – granulocyte, macrophage; RBC = red blood cell count; ttMA = trans,trans-muconic acid; WBC = white blood cell count.

Figure 3. Graphical analysis of AML precursors. Selected biomarkers of effect potentially related to AML, as well as the AML incidence, were plotted versus the concentration of benzene in air. The responses have been normalized to put them on roughly the same scale to facilitate comparisons, so the vertical axis is arbitrary. For clarity, only selected biomarkers are shown. Data are shown for individual biomarkers (e.g., aneusomy and hyperdiploidy) that are combined into one category (e.g., chromosome aberrations) in Table 1B.

Figure 4. Histogram of the uncertainty distribution for the slope from % micronuclei to AML in the full Bayesian network of all candidate biomarkers. Pr (Slope > 0) = 61%, based on the posterior frequency distribution for this parameter from the MCMC calibration of the Bayesian network of potential biomarkers. The histogram is based on 60,000 iterations of the Markov chain, that is, 20,000 from each of three independent chains. The bins, or ranges of parameter values defined for the histogram, are shown on the horizontal axis, and the number of times the parameter value falls within each bin is counted for the vertical axis values.

Figure 5. Histogram of the uncertainty distribution for the slope from 8-OHdG to AML in the full Bayesian network of all candidate biomarkers. Pr (Slope > 0) = 85%, based on the posterior frequency distribution for this parameter from the MCMC calibration of the Bayesian network of potential biomarkers. The histogram is based on 60,000 iterations of the Markov chain, that is, 20,000 from each of three independent chains. The bins, or ranges of parameter values defined for the histogram, are shown on the horizontal axis, and the number of times the parameter value falls within each bin is counted for the vertical axis values.

Table 1A. Candidate Biomarkers of Exposure Included in the Validation Analysis

Marker[1]	Sources	Qualitative Hill Criteria	Regression	Bayesian Network	Reason Excluded
Benzene internal dose					
Blood benzene	Ong et al., 1995; Eastmond et al., 2001; Bogadi-Sare et al., 1997; Ong et al., 1996; Hotz et al., 1997; Popp et al., 1994; Verdina et al., 2001	+[2]	+	–[2]	Outperformed by ttMA in Bayesian network, and does not represent metabolite in toxic pathway.
Breath benzene	Hotz et al., 1997; Egeghy et al., 2000; Sherwood and Sinclair, 1999	+	–		Poor correlation at low exposure (probably due to outlier).
Urinary benzene	Lan et al., 2004; Ong et al., 1995; Waidyanatha et al., 2001; Ong et al., 1996; Qu et al., 2000; Fustinoni et al., 2005	+	–		Per L of urine (high interindividual variability), poor overall correlation.
Metabolite internal dose					
Urinary ttMA	Ong et al., 1995; Rothman et al., 1996; Liu et al., 1996; Ong et al., 1996; Hotz et al., 1997; Popp et al., 1994; Qu et al., 2000; Boogaard and Sittert, 1995; Javelaud et al., 1998; Fustinoni et al., 2005; Verdina et al. 2001; Wiwanitkit et al., 2001	+	+	+	Validated, best correlation with high and low exposure, best in Bayesian network.
Urinary SPMA	Eastmond et al., 2001; Hotz et al., 1997; Popp et al., 1994; Qu et al., 2000; Boogaard and Sittert, 1995; Rappaport et al.,	+	–		Poor correlation at low exposure.

	2005; Stommel et al., 1989; Verdina et al., 2001				
BQ-Albumin adducts	Rappaport et al., 2005	–			Only 3 data points from a single study.
BO-Albumin adducts	Rappaport et al., 2005	–			Only 3 data points from a single study.
Urinary hydroquinone	Rothman et al., 1996; Ong et al., 1996; Hotz et al., 1997; Qu et al., 2000	+	–		Poor correlation at low exposure.
Urinary catechol	Rothman et al., 1996; Ong et al., 1996; Hotz et al., 1997; Qu et al., 2000	+	–		Poor correlation at low exposure.
Urinary benzene triol	Qu et al., 2000	–	+		Negative dose-response, only 3 data points from single study.
Urinary phenol	Rothman et al., 1996; Rothman et al., 1995; Ong et al., 1996; Hotz et al., 1997; Popp et al., 1994; Qu et al., 2000; Stommel et al., 1989	+	–		Poor correlation and not specific to benzene exposure at low exposure.

[1] Abbreviations: BO = benzene oxide; BQ = 1,4-benzoquinone; SPMA = S-phenylmercapturic acid; ttMA = trans,trans-muconic acid (ttMA)
[2] The "+" indicates that the candidate marker was considered appropriate, and proceeded to the next stage of the validation process. The "–" indicates the candidate marker was excluded in that stage of the validation process.

Table 1B. Candidate Biomarkers of Effect Included in the Validation Analysis

Marker[1]	Sources	Qualitative Hill Criteria	Graphical Comparison	Bayesian Network	Reason Excluded
Oxidative Stress					
Urinary 8-OHdG	Rothman et al., 1995	+	–[2]		Inconsistent dose-response
Lymphocyte 8-OHdG	Liu et al., 1996	+	+	+	Validated
Toxicogenomics					
CXCL16	Forrest et al., 2005	–			JUN expression chosen
ZNF331	Forrest et al., 2005	–			JUN expression chosen
JUN	Forrest et al., 2005	+	+	–	Outperformed by 8-OHdG
PF4	Forrest et al., 2005	–			JUN expression chosen
Genotoxicity					
N0 mutation	Rothman et al., 1995	–			only 2 observations
NN mutation	Rothman et al., 1995	–			only 2 observations
Chromosomal aberrations	Eastmond et al., 2001 ; Bogadi-Sare et al., 1997 ; Sarto et al., 1984 ; Zhang et al., 2002	+	+	–	(hyperdiploidy) outperformed by 8-OHdG
Lymphocyte micronuclei	Liu et al., 1996	+	+	–	Outperformed by 8-OHdG
Single DNA strand breaks	Rothman et al., 1995 ; Garte et al., 2005	+	–		Response increased sharply too early
Stem cell cytotoxicity					
Colony forming units (CFU) (GEMM)[2]	Lan et al., 2004	+	+	+	Validated
Colony forming units (GM)	Lan et al., 2004	+	+	+	Validated
Burst forming units (E)	Lan et al., 2004	+	+	+	Validated
LAPA (leukocyte alkaline phosphatase activity)	Javelaud et al., 1998	–			Only measured at <= 0.2 ppm
Anemia precursors					
Hemoglobin	Lan et al., 2004 ; Fustinoni et al., 2005	+	+	–	Outperformed by stem cell pathway
Protoporphyrin	Muzyka et al., 2002	+	–		ALA in erythrocytes chosen
Protoporphyrin-DNA adducts	Muzyka et al., 2002	+	–		ALA in erythrocytes chosen

Heme synthetase	Muzyka et al., 2002	+	–		ALA in erythrocytes chosen
ALA synthetase	Muzyka et al., 2002	+	–		ALA in erythrocytes chosen
ALA in erythrocytes[3]	Muzyka et al., 2002	+	+	–	Outperformed by stem cell pathway
ALA in lymphocytes	Muzyka et al., 2002	+	–		ALA in erythrocytes chosen

[1] Abbreviations: ALA = delta- aminolevulinic acid; CFU-GEMM = colony-forming unit -- granulocyte, erythrocyte, monocyte, megakaryocyte (a precursor cell type); CFU-GM = colony-forming units – granulocyte, macrophage; GSH = reduced glutathione; NBT = Nitroblue Tetrazolium . To simplify the presentation, some of the markers shown reflect composites of multiple related markers. For example, multiple markers of chromosomal aberrations were considered, including hyperdiploidy, aneusomy, and aberrations in several individual chromosomes.
[2] The "+" indicates that the candidate marker was considered appropriate, and proceeded to the next stage of the validation process. The "–" indicates the candidate marker was excluded in that stage of the validation process.
[3] Included in the Bayesian network based on biological plausibility even though failed in graphical analysis due to limited data.

Table 1C. Endpoints Used for Evaluation of Biomarkers of Effect

Endpoint	**Source**
Red blood cell counts	Rothman et al., 1996
Red blood cell abnormalities	
White blood cell counts	
White blood cell abnormalities	
Platelet counts	
Platelet abnormalities	
Acute myeloid leukemia	Crump and Allen, 1984

Table 2. Coefficients of Determination from Linear Regression of Potential Biomarkers of Exposure vs. Exposure Level (ppm)

Endpoint	R^2 (all)	R^2 (<~1 ppm)
Blood benzene	0.37	0.25
Breath benzene	0.77	0.02
Urinary benzene	0.07	0.93
Urinary ttMA	0.94	0.12
Urinary SPMA	0.95	0.05
Urinary catechol	0.93	<0.01
Urinary hydroquinone	0.98	<0.01
Urinary phenol	0.99	0.02

Table 3. Comparison of Biomarker-based Dose-response Approaches for Acute Myeloid Leukemia (AML)

Approach	BMR[1]	EC[2] (ppm)
Benzene in Air (ppm) vs. AML[3]	1/1000	7
Urinary ttMA (mg/g Cre) vs. AML[4]	1/1000	4
Benzene in Air (ppm) vs. 8-OHdG[5]	1 sd	5
Network Analysis of AML[6]	1/1000	0.7

[1] Benchmark response, 1/1000 extra risk above background for AML or 1 control standard deviation from the control mean for 8-OHdG
[2] Effective concentration of benzene in air leading to the specified BMR
[3] Using a logistic model to directly evaluate the relationship between the concentration of benzene in air and AML
[4] Using a logistic model to identify a biomarker level corresponding to a 1/1000 incidence of AML, followed by back-calculation of the corresponding concentration of benzene in air
[5] Using a logistic model to evaluate the relationship between the concentration of benzene in air and increases in 8-OHdG
[6] Using MCMC analysis of the Bayesian network

Figure 1. Conceptual Representation of the Biomarker Decision Support System

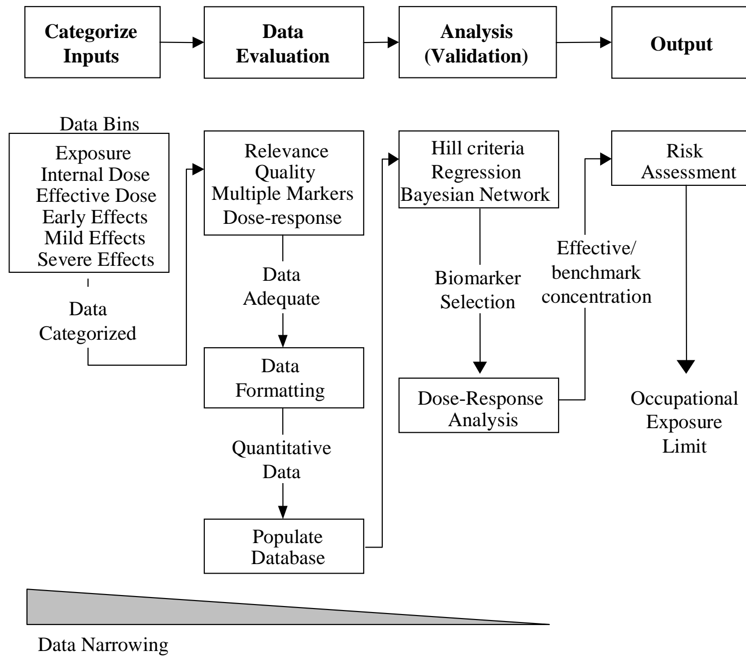

Figure 2. Network of candidate biomarkers of benzene exposure and effect.

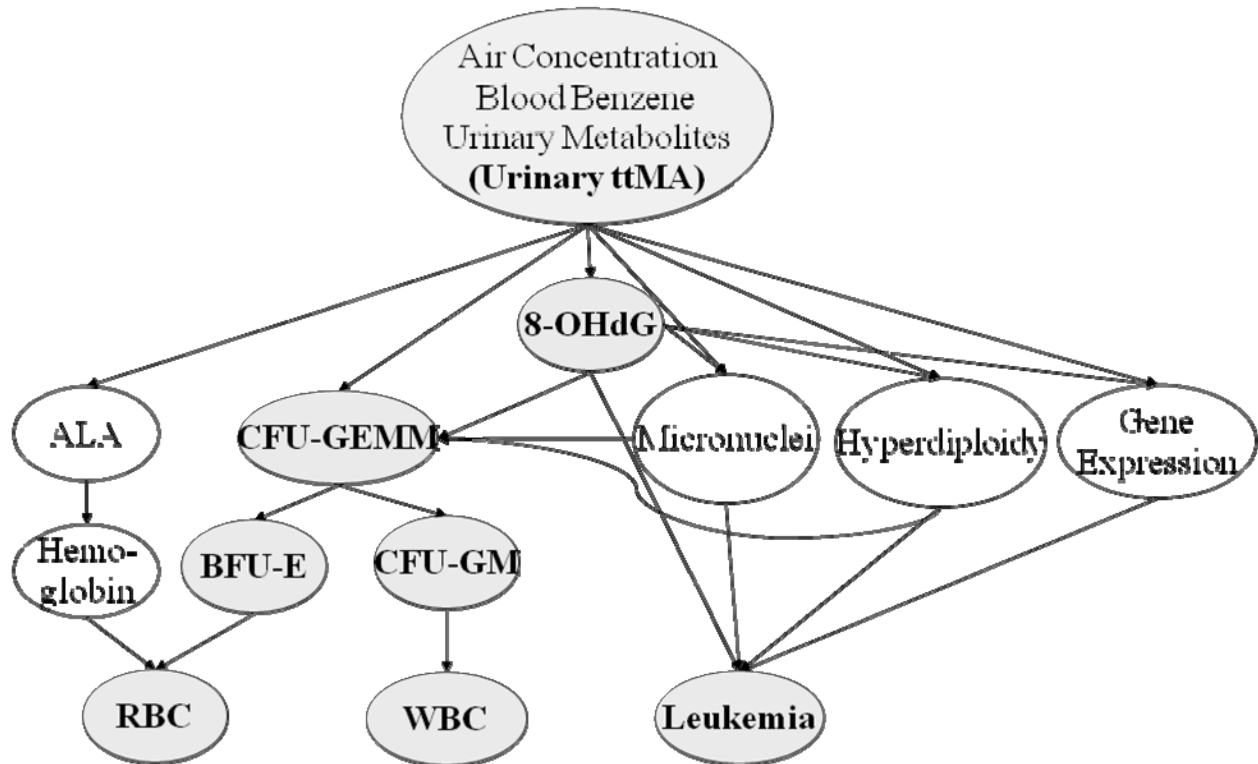

Figure 3. Graphical Analysis of Precursors of Leukemia Incidence

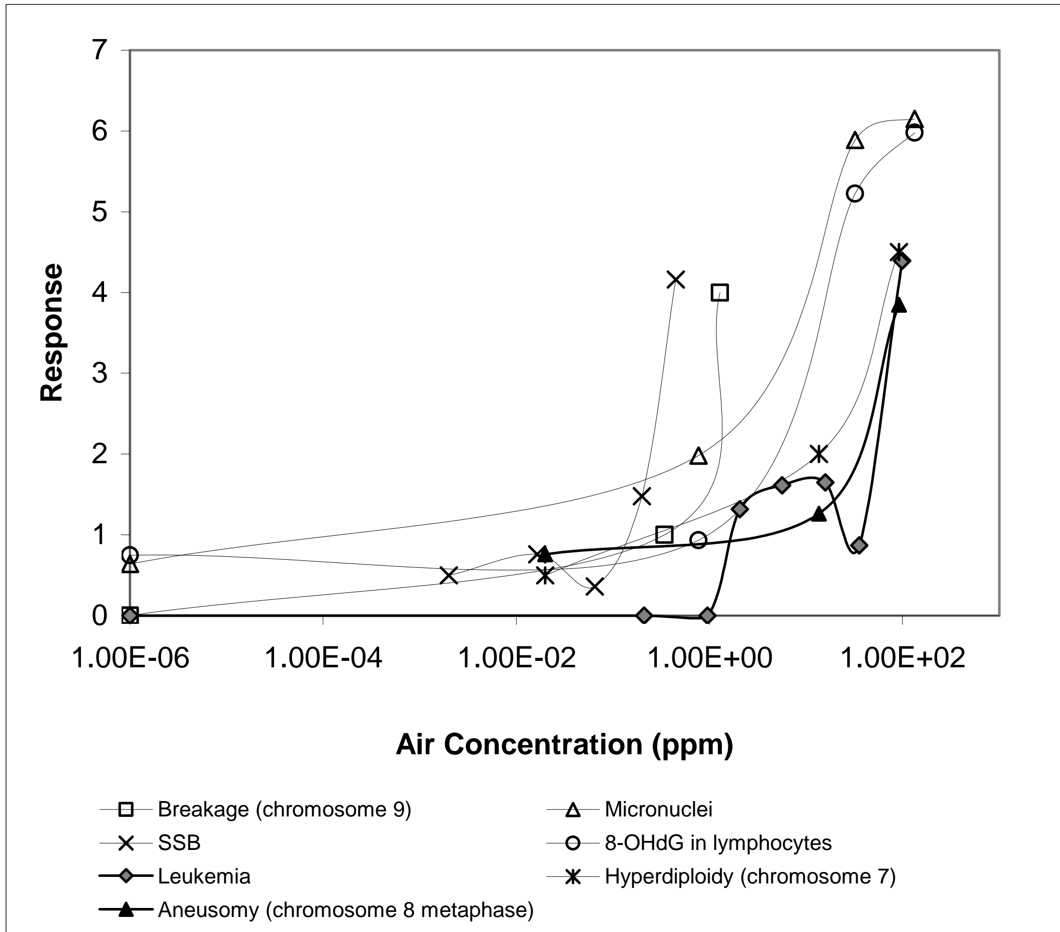

Figure 4. Histogram of the uncertainty distribution for the slope from % micronuclei to AML in the full Bayesian network of all candidate biomarkers.

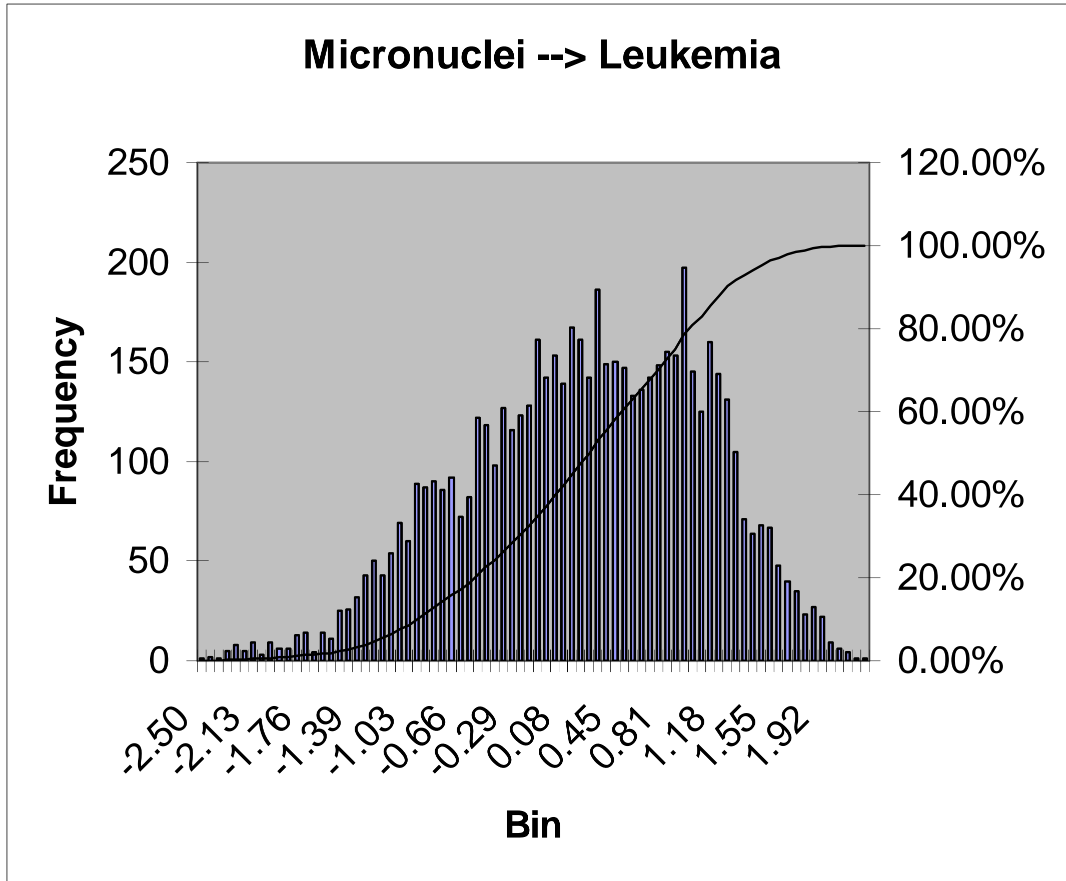

Figure 5. Histogram of the uncertainty distribution for the slope from 8-OHdG to AML in the full Bayesian network of all candidate biomarkers.