Carcinogenesis vol.34 no.1 pp.2–9, 2013
doi:10.1093/carcin/bgs382
Advance Access publication December 7, 2012

REVIEW

Low-dose metabolism of benzene in humans: science and obfuscation

Stephen M.Rappaport[1,*], Sungkyoon Kim[2], Reuben Thomas[1], Brent A.Johnson[3], Frederic Y.Bois[4] and Lawrence L.Kupper[5]

[1]Superfund Research Program and Center for Exposure Biology, School of Public Health, University of California, Berkeley, CA 94720, [2]Department of Environmental Health, Graduate School of Public Health, Seoul National University, Seoul 151-742, Republic of Korea, [3]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, [4]Université de Technologie de Compiègne, 60200 Compiègne, France and INERIS, DRC/VIVA/METO unit, 60550 Verneuil en Halatte, France and [5]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599-7420

*To whom correspondence should be addressed.
Email: srappaport@berkeley.edu Tel.: 510 642-4355 Fax: 510 642-5815

Benzene is a ubiquitous air pollutant that causes human leukemia and hematotoxic effects. Although the mechanism by which benzene causes toxicity is unclear, metabolism is required. A series of articles by Kim *et al.* used air and biomonitoring data from workers in Tianjin, China, to investigate the dose-specific metabolism (DSM) of benzene over a wide range of air concentrations (0.03–88.9 p.p.m.). Kim *et al.* concluded that DSM of benzene is greatest at air concentrations <1 p.p.m. This provocative finding motivated the American Petroleum Institute to fund a study by Price *et al.* to reanalyze the original data. Although their formal 'reanalysis' reproduced Kim's finding of enhanced DSM at sub-p.p.m. benzene concentrations, Price *et al.* argued that Kim's methods were inappropriate for assigning benzene exposures to low exposed subjects (based on measurements of urinary benzene) and for adjusting background levels of metabolites (based on median values from the 60 lowest exposed subjects). Price *et al.* then performed uncertainty analyses under alternative approaches, which led them to conclude that '… the Tianjin data appear to be too uncertain to support any conclusions …' regarding the DSM of benzene. They also argued that the apparent low-dose metabolism of benzene could be explained by 'lung clearance.' In addressing these criticisms, we show that the methods and arguments presented by Price *et al.* are scientifically unsound and that their results are unreliable.

Human metabolism of benzene

Benzene is an important industrial chemical that is also present in petroleum products and combustion effluents. Given its great volatility, this constellation of emission sources has made benzene a truly ubiquitous air contaminant (1). Although occupational exposures to high doses of benzene cause acute myeloid and acute non-lymphocytic leukemias (1), evidence of hematotoxic effects (2) and lymphohematopoietic cancers (3) in workers exposed to benzene <1 p.p.m. raises concerns about exposures to low concentrations as well. Urban populations throughout the world and cigarette smokers are routinely exposed to air concentrations of benzene in the range of 1–20 p.p.b. (4).

Although the mechanism by which benzene causes toxicity is not completely understood, metabolism appears to be required (5–7). Benzene is metabolized to a myriad of reactive species (benzene oxide, the benzoquinones, the muconaldehydes and benzene diolepoxide) (8) and more stable molecules that are excreted in urine (mainly phenol, hydroquinone, catechol and muconic acid with small amounts of benzenetriol and *S*-phenylmercapturic acid) (9). Significant concentrations of the phenolic compounds (phenol, catechol and hydroquinone) are observed in human urine even in the absence of prominent exposures to benzene and point to background sources, including diet, cigarette smoking and the microbiome (10–13).

Much of our current knowledge about human benzene metabolism has been gleaned from biomonitoring studies of Chinese workers (9,14–20). Given their importance in elucidating low-dose metabolism of benzene in humans, the publications by Kim *et al.* (18–20) deserve special attention. These articles described 620 paired air and urine measurements (unmetabolized benzene, phenol, hydroquinone, catechol, muconic acid and *S*-phenylmercapturic acid) from the largest of the Chinese studies, which included 389 workers in Tianjin, China (250 from factories using benzene and 139 from factories not using benzene). Since workers from benzene-using factories displayed hematotoxicity at air concentrations <1 p.p.m. (2), low-dose benzene metabolism was of particular interest. Because the personal air monitors used to measure airborne benzene could not detect air concentrations below about 0.2 p.p.m., Kim *et al.* (18) used a calibration model to predict air concentrations for the low-exposed subjects based on measurements of urinary benzene. Then, to adjust for background levels of each metabolite, Kim *et al.* (18,19) subtracted median metabolite concentrations observed in the 60 lowest-exposed subjects (range: <1–3 p.p.b.). Kim *et al.* (18,19) investigated the dose-specific metabolism (DSM) of benzene by dividing the background-adjusted concentration of each metabolite and their sum ('total metabolites') by the corresponding air concentration (μM per p.p.m. benzene).

Kim *et al.* initially aggregated subjects by exposure level (30 per group) to investigate the empirical relationship between DSM and benzene concentrations (18). As shown in Figure 1, DSM declined 14-fold as median benzene exposures increased from 0.027 p.p.m. to 15.4 p.p.m., with most of the reduction occurring <1 p.p.m. (18). The error bars shown in Figure 1 represent 5th and 95th percentiles of bootstrap distributions that account for sampling uncertainties and use of the calibration model to estimate low exposures. The scale of uncertainties relative to the overall change in DSM indicates that the mean trend of decreasing DSM with increasing benzene exposure was unlikely to be the result of chance.

Having used this combination of robust statistics to establish the empirical DSM relationship for benzene metabolites, Kim *et al.* then used natural spline (NS) and linear models to investigate metabolite levels as functions of benzene exposure plus covariates, including age, gender, body mass index (BMI), smoking and single-nucleotide polymorphisms of important metabolism genes (19,20). As shown by the dashed line in Figure 1, NS models smoothed and extended the empirical relationships with an overall 9-fold reduction in DSM between 0.03 and 88.9 p.p.m. The open circles and error bars represent 50th, 10th and 90th percentiles of bootstrap distributions that account for sampling uncertainties and NS modeling. Follow-up analyses of covariates showed that benzene metabolism was greater in females, declined with age (19) and was influenced by polymorphic metabolism genes (*CYP2E1*, *NQO1*, *EPHX1*, *GSTM1* and *GSTT1*) (20).

Additional evidence of enhanced benzene metabolism <1 p.p.m.

The results from Kim *et al.* indicate that DSM of benzene was greatest at the lowest investigated air concentration of 0.03 p.p.m. and declined with increasing air concentrations up to 90 p.p.m. These findings are bolstered by measurements of protein adducts of reactive benzene metabolites in Chinese workers that also pointed to DSM reductions at or <1 p.p.m. (21–26). Interestingly, toxicokinetic models of benzene metabolism indicated that DSM should

Abbreviations: BMI, body mass index; DSM, dose-specific metabolism; GM, geometric mean; NS, natural spline; p.p.b., parts per billion; p.p.m., parts per million, TMP, total metabolite production.

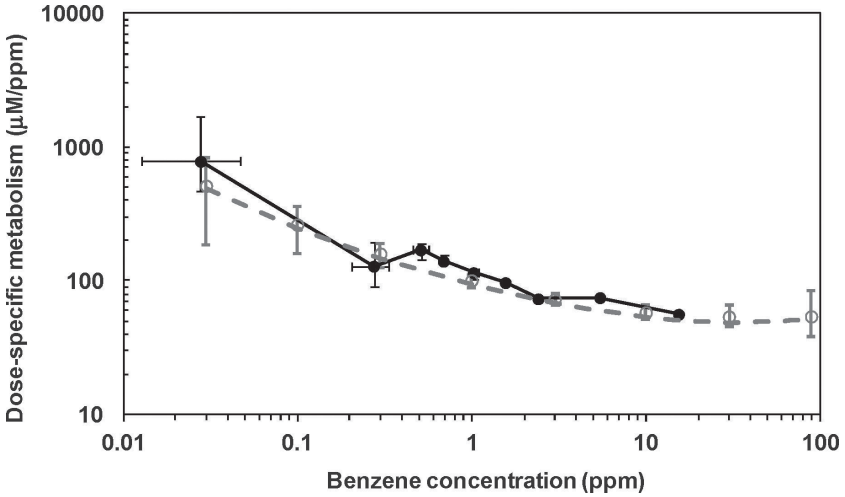

Fig. 1. Dose-specific metabolism of benzene as indicated by measurements of benzene metabolites in urine from Tianjin subjects after background adjustment. Closed circles show data aggregated by estimated benzene exposures (30 subjects per group) with error bars representing 5th and 95th percentiles of bootstrap distributions (18). The dashed curve represents the natural spline model of individual subjects' (geometric mean) benzene exposures between 0.03 and 88.9 p.p.m. (19). Open circles and error bars are 50th, 10th and 90th percentiles of bootstrap distributions (Supplementary Material, Table S.1, available at *Carcinogenesis* Online).

not diminish until air concentrations reached 10–100 p.p.m., when liver concentrations of benzene would begin to saturate metabolism by CYP2E1 (27–30). However, these models were based on experimental human and animal exposures to benzene >1 p.p.m. (or equivalent). In fact, the only experimental investigation of human metabolism <1 p.p.m. was conducted by Weisel *et al.* (31) who reported that four subjects inhaling 40 p.p.b. of ^{13}C-benzene for 2 h metabolized benzene more rapidly than had been observed in workers exposed to p.p.m. levels.

Rappaport *et al.* (32,33) fit Michaelis–Menten-like models, representing one and two saturable pathways, to benzene metabolite data combined from the Tianjin study and an earlier investigation of 44 Shanghai workers (median air concentration = 31 p.p.m.) (9). The weight of statistical evidence strongly favored two pathways rather than one pathway for metabolism of benzene to phenol and muconic acid (33) as well as total metabolites (32). This model predicted that almost three-fourths of benzene metabolism <0.1 p.p.m. resulted from the putative high-affinity (low-dose) pathway (33) and was supported by calculations based on independent data.

Criticism of Kim *et al.*

The conclusion of Kim *et al.* (18,19) that benzene metabolism is enhanced at sub-p.p.m. concentrations motivated the American Petroleum Institute (API) to fund a project by Price *et al.*(34), which reanalyzed the Tianjin data. After obtaining the air and metabolite data under the Freedom of Information Act, Price *et al.* (34) focused on the NS modeling results of Kim *et al.* (19) and the corresponding DSM calculations. Surprisingly, Price *et al.* ignored the robust empirical analyses from Kim *et al.*'s earlier *Carcinogenesis* article (18) that displayed the same overall DSM behavior (Figure 1), and did not discuss corroborating evidence, cited above, favoring enhanced benzene metabolism at or <1 p.p.m.

After reproducing the published results of Kim *et al.*, Price *et al.* discounted the finding of enhanced DSM for benzene <1 p.p.m. for the following reasons:

1. Kim *et al.* should have used subjects' mean values for statistical analyses rather than estimated geometric mean (GM) and median values.
2. Kim *et al.* should not have used a calibration model to predict benzene air concentrations for low exposed subjects because most of these workers were in factories where benzene was not used.
3. Alternative samples of subjects should have been considered for adjusting metabolite concentrations for background values.
4. Kim *et al.*'s uncertainty analyses did not include random errors from the calibration model.
5. Kim *et al.*'s conclusion that benzene is metabolized more efficiently <1 p.p.m. is inconsistent with knowledge about 'lung clearance.'

Price *et al.* then performed their own uncertainty analyses with NS models of the Tianjin data and concluded that the '… data appear to be too uncertain to support any conclusions of a change in the efficiency of benzene metabolism with variations in exposure' (abstract, last line).

Response to Price *et al.*

We will address each of the above criticisms of Price *et al.* considering the original work of Kim *et al.*, NS modeling results reported by Price *et al.*, follow-up uncertainty analyses with the NS models presented in our Supplementary Material (Sections 1 and 2), available at *Carcinogenesis* Online and a review of independent data regarding the 'lung clearance' of benzene (Supplementary Material, Section 3, available at *Carcinogenesis* Online). We will show that the methods and arguments presented by Price *et al.* are scientifically unsound and that their results are unreliable.

1. *Use of mean values rather than geometric mean or median values.* Because the 389 Tianjin subjects had up to fourpaired air and urine measurements (median of two per person), Kim *et al.* used estimated subject-specific GM values of air and metabolite concentrations for their analyses. They used median values from the 60 lowest-exposed subjects to estimate background concentrations of urinary metabolites and used median air and metabolite concentrations of groups of 30 subjects (aggregated by benzene exposure) in their empirical analyses (18). In fact, one can use any measure of location (e.g., mean, GM or median) to investigate paired phenomena such as air and metabolite levels and to adjust for background levels. However, when the range of observation is extremely large, as with the Tianjin dataset where subjects' GM air concentrations covered four orders of magnitude, it is common to explore relationships in the logarithmic scale (or simply the 'log scale') and to assume that the variates in question are log-normally distributed (35,36). Because the antilog of the mean of a set of logged observations is the sample GM (an estimator of the population median for a lognormal distribution), it is convenient to employ GM or median values of variates when performing log-scale analyses. This strategy has been widely used in science, engineering

and economics (36) as well as for characterizing databases of air and biological measurements (37,38) and for investigating the population toxicokinetics of benzene (28).

Price *et al.* contend that natural-scale mean (hereafter, simply 'mean') values rather than GM values should have been used to investigate exposure–metabolite relationships. However, they offer neither a scientific rationale nor any supporting references for the conjecture that mean rather than median values 'must' be used to adjust metabolite concentrations for background values, particularly in light of non-linear relationships between benzene exposures and metabolite levels. In fact, both median (and GM) or mean concentrations can be interpreted meaningfully in the natural scale for investigations of exposure–metabolite relationships. Recognizing that subjects are exposed to varying air concentrations of benzene from day to day (about 15-fold for Chinese benzene exposures) (22), median (and GM) air and metabolite values reflect 'typical' concentrations and days, whereas the mean values reflect 'average' concentrations over all days. Likewise, using metabolite levels of very low exposed subjects for background adjustment can employ either medians to represent typical background levels or means to represent average background levels. Thus, one could choose to model log-scale relationships between exposure and metabolite levels in terms of the logged median, GM or mean values. However, the combination of great within-subject variability of air concentrations on different days plus rapid metabolism (hours) complicates use of mean estimates for investigating the DSM of benzene (26,39).

A more serious issue concerns the lack of detail provided by Price *et al.* regarding their modeling of relationships between mean benzene exposures and mean metabolite levels. Price *et al.* indicate that they used the 'arithmetic mean' values for subjects with repeated measurements (p. 2096, under 'Modifications to data set'). We assume by this that they used the first moments of the natural scale observations of air and urinary analytes for obtaining subject-specific data. Apparently, Price *et al.* logged these estimated means and used the logged values to construct NS models of metabolite levels as functions of the air concentrations of benzene. However, they provide no information about the NS modeling other than to say (p. 2096, last sentence) that NS models were revised to include a 'bias correction factor', based loosely on Miller (40) that is embodied in Equation (7). [Note that Equation (7) contains an error and should be given as

$$\text{Urinary metabolite concentration} = \left(e^{[f(\ln(\text{Air benzene})]}\right)\left(e^{(\text{MSE}/2)}\right)] .$$

Because Price *et al.* did not report either their final NS models or even the numbers and values of knots (representing logged air concentrations) that they used, we could not replicate their results. In Appendix B of their Supplementary Material, which describes replication of Kim's NS models (but not new models of mean values), Price *et al.* contend (on pp. 3–4) that they '… were not able to independently determine the value of knots …' and therefore used the knot locations from Kim *et al.* This is curious because Kim *et al.* (19) (at the bottom of p. 2247 of their article) stated that six knots were '… assigned using equally spaced quantiles of the observations …' (as is common practice) and referred to Harrell's book for details (41). But in any case, use of Kim's knots would have been inappropriate under Price's Approaches B and C (described later) because sample sizes and ranges of observations differed markedly from those of the original models. The absence of basic details regarding Price *et al.*'s NS modeling, under alternative approaches for background adjustment, renders unreliable all their results save those used to replicate results by Kim *et al.* (19).

2. *Appropriateness of the calibration model.* The calibration model used by Kim *et al.* (18) to predict low benzene exposures from measurements of unmetabolized benzene in urine was motivated by Italian investigators who reported highly correlated benzene levels in air and urine in the p.p.b. to low p.p.m. range (42–45). Indeed, Kim *et al.* (18) showed that the distribution of measured air concentrations in the Tianjin study [0.2–88.9 p.p.m. (n = 228)] overlapped closely with data reported by Ghittori *et al.* (42) for benzene exposures of non-smokers [0.01–3.7 p.p.m. (n = 63)].

Of the 389 Tianjin subjects, 161 (41%) had air exposures predicted from urinary benzene measurements, i.e. 22 from factories with benzene and 139 from factories without benzene. Price *et al.* (p. 2095, right column, par. 1) contend that it was inappropriate to predict exposures of workers in factories without benzene because their urine measurements

> would be driven by non-occupational sources such as smoking, refueling vehicles, time spent in traffic, and dietary sources of benzene ... Because of the differences in the sources and timing of benzene exposures as compared to the occupationally-exposed workers, the relationship between the non-occupationally-exposed workers' benzene exposures and the levels in their spot urine samples cannot be assumed to follow the relationship that occurs in the occupationally-exposed subjects.

By erecting an artificial barrier between subjects in factories who used and did not use benzene, Price *et al.* ignore the fact that all Tianjin subjects were exposed to benzene from petroleum and combustion processes (vehicle exhausts, smoking, etc.), including the 22 benzene factory workers whose air levels were predicted from the calibration model. One cannot exclude such benzene sources by simply claiming that the subjects are 'non-occupationally exposed.' Furthermore, Kim *et al.* reported that the predicted low benzene concentrations were very reasonable when compared with independent measurements of benzene exposures in urban environments and among smokers (18). And finally, the timing of urine specimens within a workday was the same for all workers in the Tianjin study, regardless of whether the particular factory used benzene, and thus should not have biased results.

Given extensive validation of urinary benzene as a biomarker of short-term exposure, Price's criticism of the calibration model is poorly justified and, as we will show, uncertainties introduced by the calibration model were trivial. Because the scientific goal is to investigate DSM over the full range of benzene exposures, including those derived from ambient sources and smoking, it would be unscientific to ignore quantitative estimates of exposure across 41% of study subjects. Indeed, by classifying subjects with relatively high levels of urinary benzene as part of the background sample, Price *et al.* introduce substantial misclassification error into the analyses (discussed under Approach B).

3. *Alternative approaches for adjusting background levels of metabolites*. Because chemicals produced by benzene metabolism also arise from dietary and endogenous sources, Kim *et al.* adjusted subject-specific metabolite levels by subtracting median metabolite concentrations from the 60 lowest exposed subjects. Price *et al.* argue that this adjustment was inappropriate and introduced three alternative approaches, designated as 'A', 'B' and 'C.' Approach A maintained the 60 lowest exposed subjects for background correction but subtracted mean rather than median values. Approach B subtracted the estimated mean from 136 subjects from factories that did not use benzene and Approach C subtracted the estimated mean from 133 subjects exposed to air concentrations <0.03 p.p.m. Price *et al.* justified Approach C with the following statement (p. 2096, left column, par. 5): 'The third approach (C) is based upon the comment in Kim *et al*. (3) that the NS model predictions were "not reliable" below air benzene concentrations of 0.03 ppm.' Well, Kim *et al.* (18) never used the quoted words 'not reliable' and employed all data for constructing models, save those from the 60 lowest exposed subjects (background sample). Although Kim *et al.* limited their NS model 'predictions' of DSM to benzene exposures at or <0.03 p.p.m., this would have not been possible if data <0.03 p.p.m. had been removed from the models (discussed with uncertainty analysis).

Figure 2 shows distributions of exposure concentrations under the different approaches for defining background samples (shown along the bottom of the figure). Air concentrations are presented at left for subjects comprising background samples and at right for the remaining subjects available for modeling exposure–metabolite relationships. The 60 lowest exposed subjects, used for background samples by Kim *et al.* and Approach A, represent a 21-fold range of benzene concentrations (<0.001–0.003 p.p.m.), whereas the 136 subjects for Approach B represent a 3660-fold range (<0.001–0.533 p.p.m.) and the 133 subjects for Approach C represent a 206-fold range

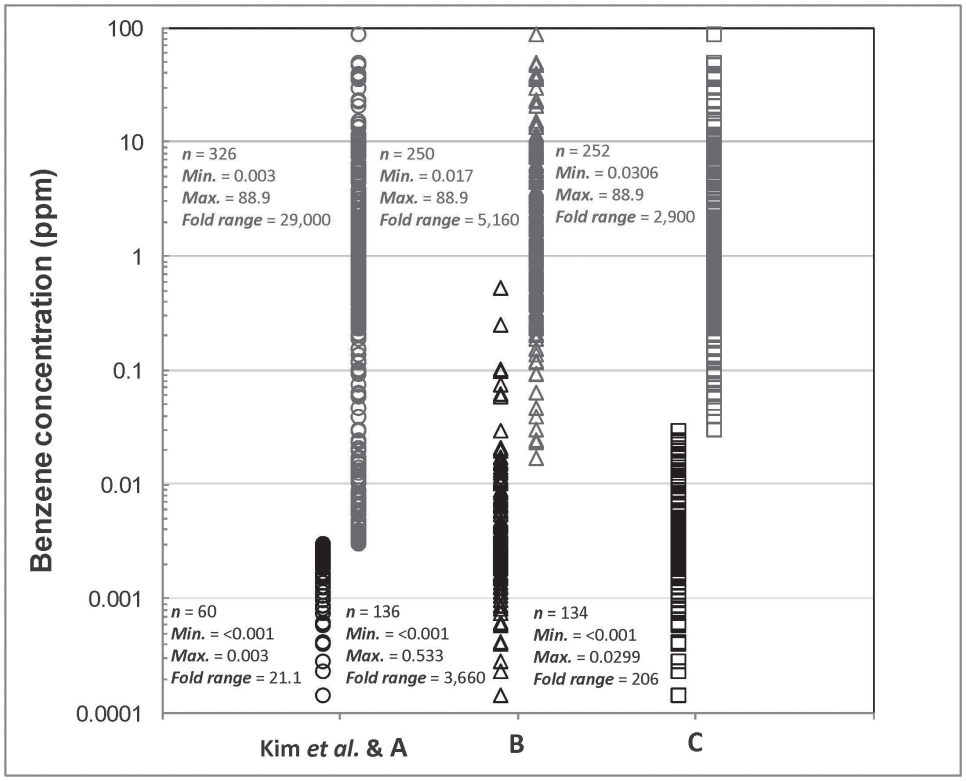

Fig. 2. Air concentrations of benzene observed across 386 subjects in the Tianjin dataset under different approaches for defining background samples. For each approach (shown at the bottom), exposure data are presented at left for subjects comprising background samples and at right for the remaining subjects available for modeling exposure–metabolite relationships.

(<0.001–0.0299 p.p.m.). Thus, by increasing the numbers of subjects in background samples for Approaches B and C, Price *et al.* greatly increase the corresponding ranges of benzene concentrations and the attendant misclassification of exposure. Price *et al.* also make fewer data available for NS models under Approaches B and C and greatly reduce the ranges of modeled air concentrations. Whereas Kim *et al.* (and Approach A) employed 326 subjects covering a 29000-fold range of air concentrations, Approach B includes 250 subjects covering a 5160-fold range and Approach C includes 252 subjects covering a 2900-fold range (Figure 2). By effectively removing much of the modeled data, under Approaches B and C, Price *et al.* widened confidence intervals for estimated parameters. Under Approach B, Price *et al.* created background and modeled samples that were highly overlapping in benzene concentrations and thereby introduced misclassification errors into the analysis and, under Approach C, Price *et al.* reduced the modeled data so as to diminish power to detect low exposure effects on metabolism. There should, therefore, be no surprise that estimates of DSM under Approaches B and C would differ substantially from those of Approach A and Kim *et al.*

We recognize that methods for background adjustment other than that employed by Kim *et al.* could be used to investigate the DSM of benzene. For example, a concurrent estimation of background and exposure effects for the Tianjin data (same model, all data together) could have advantages (32,33). However, there appears to be no scientific justification for arbitrarily expanding the range of benzene exposures in background samples by orders of magnitude while also reducing the numbers and ranges of modeled data (Figure 2).

4. *Uncertainty analyses.* Kim *et al.* performed bootstrapping to estimate uncertainties for both the empirical analyses (18) and NS modeling (19) (Figure 1). Although bootstrap distributions for the empirical analyses accounted for sampling uncertainties as well as for use of the calibration model, those for the NS models only considered sampling uncertainties. In their reanalysis of the Tianjin data, Price *et al.* focused exclusively on the NS models even though the robust empirical analyses showed essentially the same trend of DSM (Figure 1). This is apparently because Kim *et al.* did not include the calibration model in uncertainty analyses for the NS models, but did so for the empirical analyses. In any case, Price *et al.* refer repeatedly to the calibration model and (on p. 2096, left column, par. 2) imply that uncertainties in Kim's NS models were substantially greater than those reported. To test this conjecture, we repeated the bootstrap analyses for Kim's NS models with and without calibration uncertainty. The results are given in Supplementary Material (Section 1, Tables S.1 and S.2), available at *Carcinogenesis* Online, and are summarized in Figure 3, which shows 50th, 10th and 90th percentiles of bootstrap distributions obtained either with calibration uncertainty (solid and dashed curves) or without calibration uncertainty (circles and error bars). Clearly, the calibration model added trivial uncertainty to trends of DSM reported by Kim *et al.*, as would be expected from the earlier empirical results (Figure 1) and the fact that each calibration employed a rather large sample of subjects having both air and urinary measurements ($n = 228$).

Price *et al.* did not report parameters for their NS models of metabolite concentrations. Rather, results were presented as ratios of DSM values at the extremes of the range of modeled benzene concentrations between 0.03 and 88.9 p.p.m. (Note that Price *et al.* use 'total metabolite production' abbreviated 'TMP' instead of DSM.) Although we could not reproduce their findings, we discovered anomalous results in Price *et al.*'s TMP ratios that point to unsound methods. Their uncertainty analyses—summarized by box-and-whisker plots in Price *et al.*'s Figure 2—are inconsistent with point estimates derived from their observed data distributions (given on p. 2096 in the first two paragraphs under 'Results'). This is illustrated in our Figure 4, which juxtaposes the point estimates of Price's TMP ratios with the corresponding bootstrap distributions represented in Price *et al.*'s Figure 2. All point estimates of TMP ratios from Approaches A, B and C are biased upward relative to the confidence intervals estimated via bootstrapping. This suggests that the procedure used to obtain parameter estimates from bootstrap samples was different from that used to

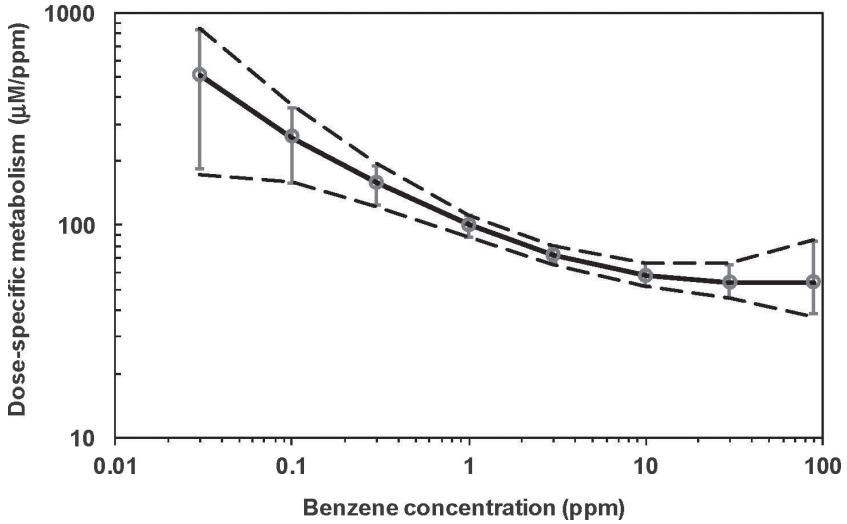

Fig. 3. Comparing uncertainty analyses of Kim *et al.* (19) with and without consideration of uncertainty from the calibration model (from Supplementary Material, Tables S.1 and S.2, available at *Carcinogenesis* Online). The open circles and error bars represent the 50th, 10th and 90th percentiles of bootstrap distributions (Supplementary Table S.1, available at *Carcinogenesis* Online), reproducing the original analyses of Kim *et al.*, without considering uncertainty from the calibration model. The solid and dashed curves in Supplementary Figure S.1, available at *Carcinogenesis* Online, represent the corresponding 50th, 10th and 90th percentiles of new bootstrap distributions, which include uncertainty from the calibration model (Supplementary Table S.2, available at *Carcinogenesis* Online). Comparing the two sets of results, it is apparent that uncertainties from the calibration model contributed only trivially to the bootstrap distributions.

obtain the point estimates. Because bootstrap samples are generated from the observed data distributions, one would expect that the point estimates would lie within the significant mass of bootstrap distributions. For example, Figure 1 shows that 50th percentiles of the bootstrap distributions from Kim *et al.* (19) (open circles) match almost perfectly NS modeling of the data distribution (dashed curve).

Although Price *et al*. provided insufficient details for us to determine the source(s) of these discrepancies, possible problems involve the error in their Equation (7) noted earlier and also Price *et al.*'s adjustment for 'model uncertainty' (p. 2096, right column, par. 2) to generate bootstrap distributions, but apparently not for modeling the data distributions. Unfortunately, Price *et al.* did not define 'model uncertainty' and provided no references for justification. If 'model uncertainty' is used in the context of say (46), where multiple models are shown to equally fit the data, then one could consider reconciling the predictions from the different models. Unfortunately, no efforts in this direction were made by Price *et al.* A consequence of adding this unjustified noise to the predictions of the NS models would be to increase the sizes of confidence intervals for the estimated parameters.

Under their Approach A, which used the same modeled and background samples as Kim *et al.*, Price *et al.* reported a point estimate of the TMP ratio of 9.4 (p. 2096, right column, par. 5), which is quite

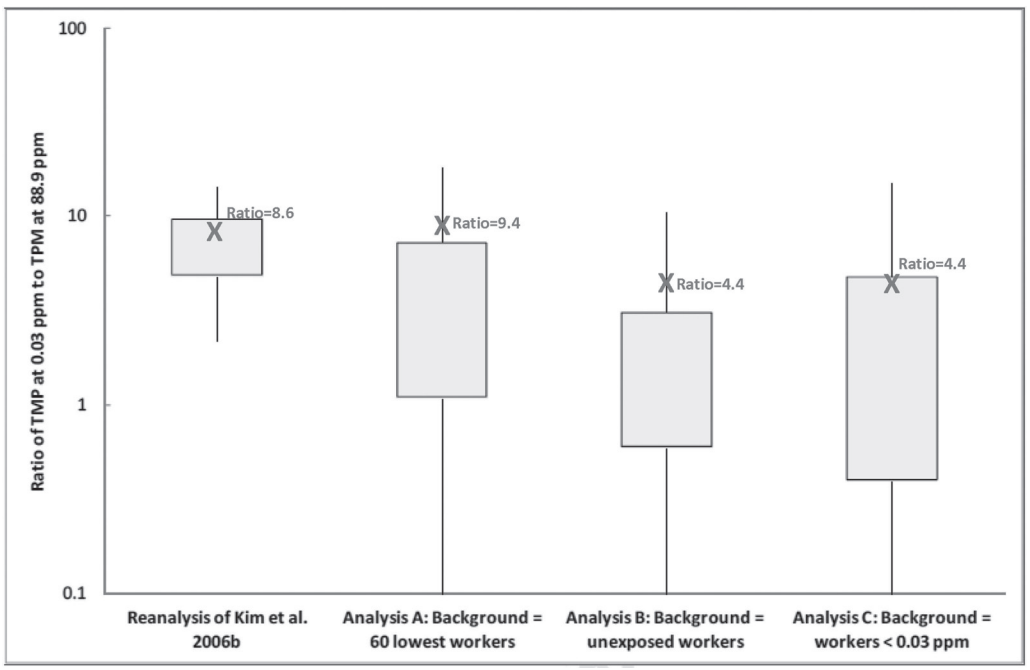

Fig. 4. Reproduction of Figure 2 from Price *et al.* showing box-and-whisker plots of bootstrap distributions for ratios of total metabolite production (TMP) at benzene concentrations of 0.03 and 88.9 p.p.m. (Note that Price *et al.*'s TMP is equivalent to 'DSM' in this article). Values of TMP ratios designated with Xs were added by the authors of this article to identify point estimates for data distributions given by Price *et al.* (p. 2096 in the first two paragraphs under 'Results').

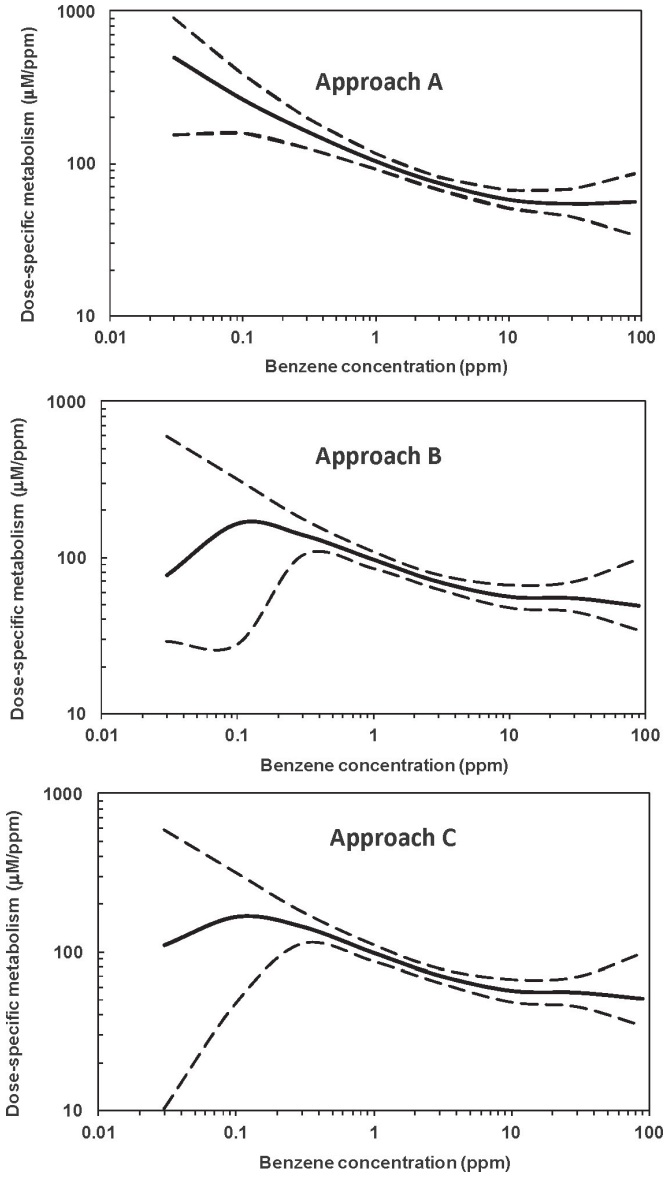

Fig. 5. Comparing uncertainty analyses for natural spline models of total metabolite concentrations as functions of subject-specific geometric mean air concentrations under Price *et al.*'s Approaches A, B and C (from Supplementary Material, Tables S.2, S.3 and S.4, available at *Carcinogenesis* Online). The solid curves represent 50th percentiles of bootstrap distributions and the dashed curves represent 10th and 90th percentiles of bootstrap distributions.

similar to the value of 9.2 obtained from the NS models of Kim *et al.* Thus, even after substituting estimated means for the GM and median values used by Kim *et al.*, Price recapitulated the finding of a 9-fold reduction in DSM of benzene between 0.03 and 88.9 p.p.m. Price *et al.* then redefined the background and modeled groups for Approaches B and C in a manner that would very likely obscure any effects of enhanced metabolism at low benzene exposures. Because NS model fits are both continuous and differentiable (47), model predictions of metabolite levels at 0.03 p.p.m., i.e. the lower bound used by Price *et al.* to define TMP ratios, are influenced by subjects exposed in a neighborhood around this air concentration. Whereas 90 subjects were available between 0.03 and 0.2 p.p.m. for Approach A, only 16 and 17 subjects were available under Approaches B and C, respectively. With very few data in the neighborhood around 0.03 p.p.m. under Approaches B and C, NS models of metabolite levels become unstable and have large variances at low air concentrations. Model instability would also be accentuated by inappropriate assignment of NS knots from Kim *et al.* for Approaches B and C, which have different ranges and sample sizes.

To gain insight into Price's alternative approaches, we generated bootstrap samples for NS models of subject-specific GMs (rather than mean values) under Approaches B and C, including all sources of uncertainty, at staged air concentrations between 0.03 and 88.9 p.p.m. (Supplementary Material, Section 2, Tables S.3 and S.4, available at *Carcinogenesis* Online). (Note that bootstrap distributions under Approach A were reported in Supplementary Table S.2, available at *Carcinogenesis* Online.) As shown in Figure 5, 10th and 50th percentile values of DSM for Approaches B and C decrease dramatically compared with those for Approach A at air concentrations <0.1 p.p.m. because of the sparseness of data in this range and by large proportions of negative values from background adjustment. Indeed, our analyses indicate that Approaches B and C effectively precluded any attempt at elucidating DSM of benzene in the range of 0.03 p.p.m. (Figure 5), a value that Price *et al.* weighted heavily in their calculations.

5. *Lung clearance.* Price *et al.* suggest that Kim *et al.*'s conclusion of enhanced benzene metabolism at sub-p.p.m. exposures is at odds with current knowledge about 'lung clearance.' However, their discourse on this matter (p. 2095, left column, par. 3 and p. 2098, right column, par. 1) is illogical because they confuse the concept of passive clearance of benzene from the lung (by exhalation) with absorption of benzene in the lung (following inhalation). The concept of clearance relates to removal of a chemical from the blood or plasma and has units of volume per unit of time (48). Clearance represents the sum of all removal processes, including saturable metabolism and passive first-order excretion via the lung (exhaled air) and kidney (urine). For volatile compounds like benzene, passive excretion by exhalation accounts for substantial proportions of the inhaled dose (49). In comparison, urinary excretion of benzene constitutes <2% of the benzene dose in humans exposed to tens to hundreds of p.p.m. (15). With this in mind, it is difficult to understand Price's statement (p. 2095, left column, par. 3) that 'There is a consensus that once absorbed, benzene is almost completely metabolized and that benzene's metabolites and any unreacted benzene are excreted in the urine (7,8). Indeed, neither of Price's reference 7 or 8 (both are reports of U.S. governmental agencies) offers such consensus.

To consider the relative contributions of passive and metabolic clearance of benzene, we invoke mass balance arguments that underlie physiologically based pharmacokinetic modeling of volatile organic compounds generally (49) and benzene in particular (28,29,50). At the beginning of exposure, we can assume that virtually all benzene entering the alveolar air is absorbed. Therefore, the ratio of the exhaled benzene concentration (C_{exh}) to the inhaled benzene concentration (C_{inh}) should be about ($1 - f_{alv}$), f_{alv} being the alveolar fraction of the lung volume (the rest being dead-space for gas exchange). For human benzene exposures, f_{alv} has been estimated to be 0.72 (50). After prolonged exposure, equilibrium is reached between the concentrations of benzene in air and blood. It follows from straightforward calculations (Supplementary Material, Section 3, available at *Carcinogenesis* Online) that the ratio C_{exh}/C_{inh} is rather insensitive to the fraction of benzene metabolized Q_{met}/Q_{inh}, where Q_{met} and Q_{inh} are the quantities of benzene metabolized and inhaled per unit time, respectively. Equation (S3) of Supplementary Material, available at *Carcinogenesis* Online, indicates that C_{exh}/C_{inh} ranges between 0.28 and 1. Thus, when Q_{met}/Q_{inh} doubles in magnitude from 0.4 to 0.8, the corresponding value of C_{exh}/C_{inh} only decreases by 40% (from 0.71 to 0.42). This suggests that a range of exhaled fractions would be compatible with a given metabolized fraction and vice versa. Nonetheless, the necessary interplay between C_{exh}/C_{inh} and Q_{met}/Q_{inh} contradicts Price *et al.*'s surprising suggestion (p. 2098, right column, par. 1) that 'lung clearance' can explain apparent increases in DSM without '… any change in the fraction of the absorbed dose that is metabolized.'

Despite insensitivity of the fraction exhaled to the fraction of metabolized benzene, Equation (S3) suggests that the relationship between C_{exh}/C_{inh} and Q_{met}/Q_{inh} can be investigated by examining inhaled and exhaled air from humans exposed to a range of air concentrations.

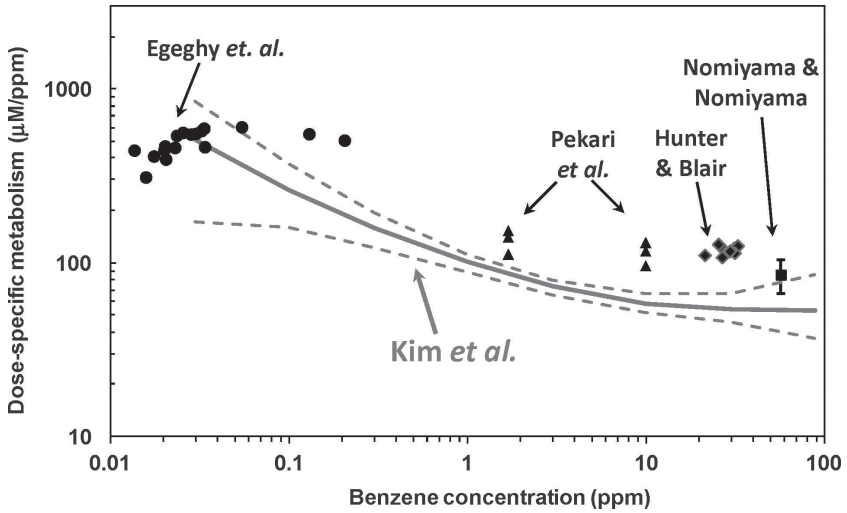

Fig. 6. Predictions of dose-specific metabolism based on measurements of benzene concentrations in inhaled and exhaled air from four published studies (51–54) juxtaposed with the modeling results of Kim *et al.* given in Supplementary Material (Table S.2), available at *Carcinogenesis* Online. The inhaled/exhaled air studies are described in Supplementary Material (Section 3) and the data and calculations are given in Table S.6, available at *Carcinogenesis* Online.

After exploring the literature, we extracted measurements from four human studies (51–54) that allowed us to estimate the ratio C_{exh}/C_{inh} and then used Equation (S3) to estimate Q_{met}/Q_{inh} over a wide range of benzene exposures. Three of the studies involved controlled exposures of volunteer subjects to benzene concentrations between 1.7 and 57 p.p.m. (52–54) and the fourth was an observational study of automobile mechanics exposed to air concentrations between 0.007 and 0.205 p.p.m. (median = 0.024 p.p.m.) (51).

Data from these four human studies are described and summarized in Supplementary Material (Section 3 and Table S.5), available at *Carcinogenesis* Online. Measurements of inhaled and exhaled benzene show that C_{exh}/C_{inh} increased with benzene exposure from about 0.5 to 0.7, whereas estimates of Q_{met}/Q_{inh} decreased concomitantly from about 0.7–0.4. To put the results of Kim *et al.* into perspective, predicted values of Q_{met}/Q_{inh} were used to estimate the corresponding values of DSM via Equation (S4) as described in Section 3 of Supplementary Material, available at *Carcinogenesis* Online. Overall, values of DSM decreased about 6-fold, from 509 µM/p.p.m. <0.2 p.p.m. (median value) to 86 µM/p.p.m. at 57 p.p.m. (mean value). As shown in Figure 6, these estimates of DSM are consistent with Kim's models of urinary metabolite levels (18,19) and suggest that Price's arguments regarding 'lung clearance' are scientifically unfounded.

Discussion

The work of Kim *et al.* (18–20) represents the most comprehensive analyses of human metabolism of benzene, an environmentally ubiquitous carcinogen. The molecular epidemiologic investigation that generated the Tianjin data was conducted with the utmost care regarding study design, selection of participating subjects, collection of air and biological specimens and measurement of biomarkers. The high quality of these data allowed Kim *et al.* to tease out low-dose metabolic effects that eluded other investigators. By describing their methods in detail, the authors maintained the transparency required for scientific work. Indeed, Price *et al.* (34) were able to successfully reproduce the NS modeling results of Kim *et al.* (19), which showed enhanced metabolism of benzene at low exposure levels.

Because virtually all humans are exposed to benzene from petroleum products and combustion processes, including tobacco smoking, Kim's finding of increased benzene metabolism at air concentrations <1 p.p.m. has public health implications. And even though Kim *et al.* did not estimate human health risks associated with sub-p.p.m. benzene exposures, their results suggest that these risks could be greater than expected from investigations of heavily exposed workers. Indeed, the recent report of increased risks of lymphohematopoietic cancers at average benzene exposures <1 p.p.m. (3) lends support to this argument.

After examining Price *et al.*'s reanalyses of the Tianjin data, we documented major shortcomings in the authors' rationale, methods and scientific rigor, the most serious of which are summarized as follows. First, Price *et al.* ignored the totality of evidence, which indicates that benzene is more efficiently metabolized at air concentrations <1 p.p.m. They did not mention that robust statistical analyses of the Tianjin data—published in *Carcinogenesis* (18)—reported a 14-fold reduction in DSM between 0.027 and 15.4 p.p.m. or that follow-up kinetic modeling pointed to a second metabolic pathway that was active at benzene concentrations <1 p.p.m. (32,33). They overlooked corroborating evidence for sub-p.p.m. metabolic effects from measurements of benzene-derived protein adducts (21–26) and from the only controlled exposures of human subjects <1 p.p.m. (31). Good science requires a fuller presentation of the literature. Second, Price *et al.* did not provide sufficient details concerning their NS modeling and uncertainty analyses to allow independent confirmation of their results. This lack of transparency and inconsistent results (Figure 4) make the findings of Price *et al.* unreliable. Third, Price *et al.* reanalyzed data in a manner that was virtually guaranteed to obscure low-dose effects of benzene exposure. When background adjustment with estimated mean metabolite levels from a sample of 60 subjects with demonstrably low benzene exposures (Approach A) recapitulated Kim's findings, Price *et al.* turned to alternatives (Approaches B and C) that magnified uncertainties and introduced misclassification errors (Figure 5). The authors fostered these alternatives in spite of subject-specific benzene measurements showing that background and modeled samples for Approaches B and C were unsuitable for discriminating metabolic changes at low air concentrations (Figure 2). Fourth, Price *et al.* promoted an illogical mechanistic argument to suggest that the apparent enhancement of low-dose benzene metabolism could be explained by 'lung clearance.' In fact, a careful examination of the published human literature on passive clearance of benzene from the lungs provides further evidence of enhanced low-dose metabolism of benzene, consistent with the findings of Kim *et al.* (Figure 6).

These shortcomings raise questions whether Price's reanalysis of Kim's work was motivated by scientific skepticism or by an effort to obfuscate the low-dose metabolism of benzene. In either case, we regard the above shortcomings as sufficient to justify retraction of Price *et al.* (34) from *Carcinogenesis* (http://publicationethics.org/).

Supplementary material

Supplementary Material, Tables 1–5 and Figure 1 can be found at http://carcin.oxfordjournals.org/

Funding

National Institute for Environmental Health Sciences (P42ES04705 to S.M.R.).

Conflicts of Interest Statement: S.M.R. has received consulting and expert testimony fees from law firms representing plaintiffs in cases involving exposure to benzene and has also received research support from the American Petroleum Institute and the American Chemistry Council. Other authors declare no conflicts of interest.

References

1. IARC. (2012) *Chemical Agents and Related Occupations, Volume 100 F, A Review of Human Carcinogens*. International Agency for Research on Cancer, Lyon, France.
2. Lan,Q. *et al.* (2004) Hematotoxicity in workers exposed to low levels of benzene. *Science*, **306**, 1774–1776.
3. Schnatter,A.R. *et al.* (2012) Myelodysplastic syndrome and benzene exposure among petroleum workers: an international pooled analysis. *J. Natl. Cancer Inst*. J Natl Cancer Inst. 2012 Nov 21;104(22):1724-37. doi: 10.1093/jnci/djs411. Epub 2012 Oct 30.
4. Weisel,C.P. (2010) Benzene exposure: an overview of monitoring methods and their findings. *Chem. Biol. Interact.*, **184**, 58–66.
5. Ross,D. (2000) The role of metabolism and specific metabolites in benzene-induced toxicity: evidence and issues. *J. Toxicol. Environ. Health Part A*, **61**, 357–372.
6. Snyder,R. (2004) Xenobiotic metabolism and the mechanism(s) of benzene toxicity. *Drug Metab. Rev.*, **36**, 531–547.
7. Smith,M.T. *et al.* (2011) Benzene, the exposome and future investigations of leukemia etiology. *Chem. Biol. Interact.*, **192**, 155–159.
8. Waidyanatha,S. *et al.* (2005) Investigation of cysteinyl protein adducts of benzene diolepoxide. *Chem. Biol. Interact.*, **153–154**, 261–266.
9. Waidyanatha,S. *et al.* (2004) Rapid determination of six urinary benzene metabolites in occupationally exposed and unexposed subjects. *Anal. Biochem.*, **327**, 184–199.
10. Bone,E. *et al.* (1976) The production of urinary phenols by gut bacteria and their possible role in the causation of large bowel cancer. *Am. J. Clin. Nutr.*, **29**, 1448–1454.
11. Carmella,S.G. *et al.* (1982) Quantitative analysis of catechol and 4-methylcatechol in human urine. *Food Chem. Toxicol.*, **20**, 587–590.
12. Deisinger,P.J. *et al.* (1996) Human exposure to naturally occurring hydroquinone. *J. Toxicol. Environ. Health*, **47**, 31–46.
13. McDonald,T.A. *et al.* (2001) Hypothesis: phenol and hydroquinone derived mainly from diet and gastrointestinal flora activity are causal factors in leukemia. *Leukemia*, **15**, 10–20.
14. Rothman,N. *et al.* (1998) Urinary excretion of phenol, catechol, hydroquinone, and muconic acid by workers occupationally exposed to benzene. *Occup. Environ. Med.*, **55**, 705–711.
15. Waidyanatha,S. *et al.* (2001) Urinary benzene as a biomarker of exposure among occupationally exposed and unexposed subjects. *Carcinogenesis*, **22**, 279–286.
16. Melikian,A.A. *et al.* (2002) Personal exposure to different levels of benzene and its relationships to the urinary metabolites S-phenylmercapturic acid and trans,trans-muconic acid. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **778**, 211–221.
17. Qu,Q. *et al.* (2000) Validation of biomarkers in humans exposed to benzene: urine metabolites. *Am. J. Ind. Med.*, **37**, 522–531.
18. Kim,S. *et al.* (2006) Using urinary biomarkers to elucidate dose-related patterns of human benzene metabolism. *Carcinogenesis*, **27**, 772–781.
19. Kim,S. *et al.* (2006) Modeling human metabolism of benzene following occupational and environmental exposures. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 2246–2252.
20. Kim,S. *et al.* (2007) Genetic polymorphisms and benzene metabolism in humans exposed to a wide range of air concentrations. *Pharmacogenet. Genomics*, **17**, 789–801.
21. Rappaport,S.M. *et al.* (2002) Non-linear production of benzene oxide-albumin adducts with human exposure to benzene. *J. Chromatog. B*, **778**, 367–374.
22. Rappaport,S.M. *et al.* (2002) Albumin adducts of benzene oxide and 1,4-benzoquinone as measures of human benzene metabolism. *Cancer Res.*, **62**, 1330–1337.
23. Lin,Y.S. *et al.* (2007) Albumin adducts of electrophilic benzene metabolites in benzene-exposed and control workers. *Environ. Health Perspect.*, **115**, 28–34.
24. Johnson,B.A. *et al.* (2005) Modeling exposure-biomarker relationships: applications of linear and nonlinear toxicokinetics for understanding genotoxic metabolism and carcinogenesis. *J. Agric. Biol. Environ. Stat.*, **10**, 440–459.
25. Johnson,B.A. *et al.* (2007) On modelling metabolism-based biomarkers of exposure: a comparative analysis of nonlinear models with few repeated measurements. *Stat. Med.*, **26**, 1901–1919.
26. Taylor,D.J. *et al.* (2008) Statistical methods for evaluating exposure-biomarker relationships. *J. Agric. Bio. Environ. Stat.*, **14**, 367–387.
27. Bois,F.Y. *et al.* (1991) Comparison of three physiologically based pharmacokinetic models of benzene disposition. *Toxicol. Appl. Pharmacol.*, **110**, 79–88.
28. Bois,F.Y. *et al.* (1996) Population toxicokinetics of benzene. *Environ. Health Perspect.*, **104 (suppl. 6)**, 1405–1411.
29. Travis,C.C. *et al.* (1990) Pharmacokinetics of benzene. *Toxicol. Appl. Pharmacol.*, **102**, 400–420.
30. Watanabe,K.H. *et al.* (1994) Benzene toxicokinetics in humans: exposure of bone marrow to metabolites. *Occup. Environ. Med.*, **51**, 414–420.
31. Weisel,C.P. *et al.* (2003) Use of stable isotopically labeled benzene to evaluate environmental exposures. *J. Expo. Anal. Environ. Epidemiol.*, **13**, 393–402.
32. Rappaport,S.M. *et al.* (2009) Evidence that humans metabolize benzene via two pathways. *Environ. Health Perspect.*, **117**, 946–952.
33. Rappaport,S.M. *et al.* (2010) Human benzene metabolism following occupational and environmental exposures. *Chem. Biol. Interact.*, **184**, 189–195.
34. Price,P.S. *et al.* (2012) A reanalysis of the evidence for increased efficiency in benzene metabolism at airborne exposure levels below 3 p.p.m. *Carcinogenesis*, **33**, 2094–2099.
35. Aitchison,J. *et al.* (1957) *The Lognormal Distribution*. Cambridge University Press, London.
36. Crow,E.L. *et al.*. (eds.) (1988) *Lognormal Distributions: Theory and Applications*. Marcel Dekker, New York.
37. Kromhout,H. *et al.* (1993) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Ann. Occup. Hyg.*, **37**, 253–270.
38. Lin,Y.S. *et al.* (2005) Air samples versus biomarkers for epidemiology. *Occup. Environ. Med.*, **62**, 750–760.
39. Rappaport,S.M. *et al.* (2008) *Quantitative Exposure Assessment*. Stephen Rappaport, El Cerrito, California.
40. Miller,D.M. (1984) Reducing transformation bias in curve fitting. *The American Statistician*, **38**, 124–126.
41. Harrell,F.E. (2001) *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York.
42. Ghittori,S. *et al.* (1993) Urinary excretion of unmetabolized benzene as an indicator of benzene exposure. *J. Toxicol. Environ. Health*, **38**, 233–243.
43. Ghittori,S. *et al.* (1995) Evaluation of occupational exposure to benzene by urinalysis. *Int. Arch. Occup. Environ. Health*, **67**, 195–200.
44. Fustinoni,S. *et al.* (1999) Headspace solid-phase microextraction for the determination of benzene, toluene, ethylbenzene and xylenes in urine. *J. Chromatogr. B Biomed. Sci. Appl.*, **723**, 105–115.
45. Fustinoni,S. *et al.* (2005) Urinary t,t-muconic acid, S-phenylmercapturic acid and benzene as biomarkers of low benzene exposure. *Chem. Biol. Interact.*, **153–154**, 253–256.
46. Neuman,S. (2003) Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment*, **17**, 291–305.
47. Durrleman,S. *et al.* (1989) Flexible regression models with cubic splines. *Stat. Med.*, **8**, 551–561.
48. Gibaldi,M. *et al.* (1982) *Pharmacokinetics.* 2nd edn, revised and expanded. Marcel Dekker, New York.
49. Andersen,M.E. (1981) Saturable metabolism and its relationship to toxicity. *Crit. Rev. Toxicol.*, **9**, 105–150.
50. Yokley,K. *et al.* (2006) Physiologically-based pharmacokinetic modeling of benzene in humans: a Bayesian approach. *Risk Anal.*, **26**, 925–943.
51. Egeghy,P.P. *et al.* (2002) Self-collected breath sampling for monitoring low-level benzene exposures among automobile mechanics. *Ann. Occup. Hyg.*, **46**, 489–500.
52. Hunter,C.G. *et al.* (1972) Benzene: pharmacokinetic studies in man. *Ann. Occup. Hyg.*, **15**, 193–201.
53. Nomiyama,K. *et al.* (1974) Respiratory retention, uptake and excretion of organic solvents in man. *Int. Arch. Arbeitsmed.*, **32**, 75–83.
54. Pekari,K. *et al.* (1992) Biological monitoring of occupational exposure to low levels of benzene. *Scand. J. Work. Environ. Health*, **18**, 317–322.

Received September 30, 2012; revised November 20, 2012; accepted November 30, 2012