Using Publicly Available Data, Physiologically-Based Pharmacokinetic Model and Bayesian Simulation to Improve Arsenic Non-Cancer Dose-Response

Zhaomin Dong[a,b], CuiXia Liu[b,c], Yanju Liu[a,b], Kaihong Yan[a,b], Kirk T. Semple[d] and Ravi Naidu*[,a,b]

[a] Global Centre for Environmental Remediation, the Faculty of Science and Information Technology, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

[b] Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC CARE), Mawson Lakes, SA 5095, Australia

[c] School of Environmental Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

[d] Lancaster Environment Centre, Lancaster University, LA1 4YQ Lancaster, United Kingdom

* Corresponding Author: Ravi Naidu

ATC Building, Global Center for Environmental Remediation, Faculty of Science and Information Technology, University of Newcastle, Callaghan, NSW 2308

Email: ravi.naidu@newcastle.edu.au
ravi.naidu@crccare.com
Phone : +61 02 4913 8705

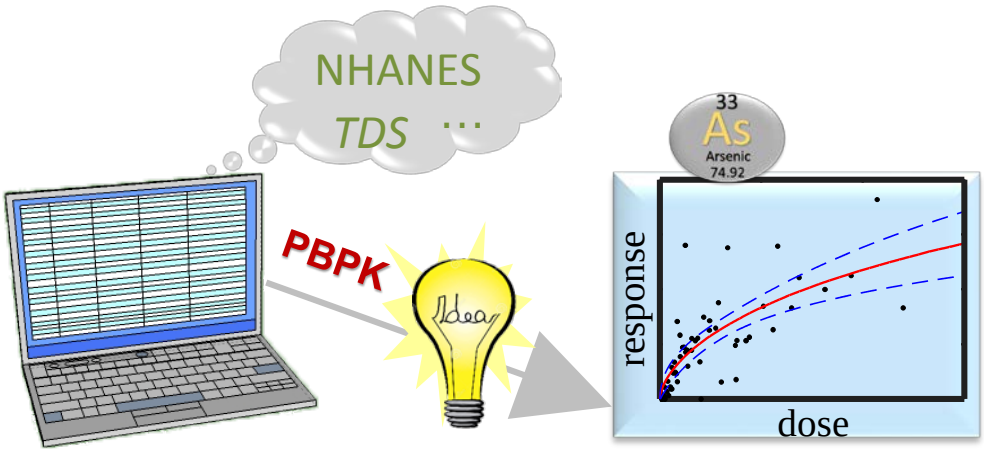
Contents

Abstract

Publicly available data can potentially examine the relationship between environmental exposure and public health, however, it has not yet been widely applied. Arsenic is of environmental concern, and previous studies mathematically parameterized exposure duration to create a link between duration of exposure and increase in risk. However, since the dose metric emerging from exposure duration is not a linear or explicit variable, it is difficult to address the effects of exposure duration simply by using mathematical functions. To relate cumulative dose metric to public health requires a lifetime physiologically-based pharmacokinetic (PBPK) model, yet this model is not available at a population level. In this study, the data from the U.S. total diet study (TDS, 2006-2011) was employed to assess exposure: daily dietary intakes for total arsenic (tAs) and inorganic arsenic (iAs) were estimated to be 0.15 and 0.028 μg/kg/day, respectively. Meanwhile, using National Health and Nutrition Examination Survey (NHANES, 2011-2012) data, the fraction of urinary As(III) levels (geometric mean: 0.31 μg/L) in tAs (geometric mean: 7.75 μg/L) was firstly reported to be approximately 4%. Together with Bayesian technique, the assessed exposure and urinary As(III) concentration were input to successfully optimize a lifetime population PBPK model. Finally, this optimized PBPK model was used to derive an oral reference dose (Rfd) of 0.8 μg/kg per day for iAs exposure. Our study also suggests the previous approach (by using mathematical functions to account for exposure duration) may result in a conservative Rfd estimation.

KEY WORDS： PBPK model; Dose response; Bayesian Simulation; Arsenic; Publicly Available data

66 **Graphical Abstract**



67

1. Introduction

Chronic exposure to elevated levels of arsenic (As) has resulted in many adverse effects appearing in humans (Maull et al. 2012; Naujokas et al. 2013). Epidemiological evidence provides opportunities to undertake a dose-response study, and furthermore to assist in assessment and management. For example, one study over a mean follow-up period of 9.7 years for 52,931 eligible participants suggested that the adjusted incidence rate ratios per 1 μg/L increment in arsenic levels in drinking water were 1.03 (95% confidence interval (CI): 1.01, 1.06) for all diabetes cases (Bräuner et al. 2014). Such epidemiological studies have convincingly linked the As exposure level and risk (Bräuner et al. 2014; U.S. EPA 1988).

Excepting exposure level, previous research has also demonstrated the incidence of diseases increases with exposure duration (Liao et al. 2008; Mazumder et al. 1998; U.S. EPA 1988). To quantify the exposure duration effects, mathematical functions (such as Weibull and Hill functions) have usually been employed, by parameterizing age factor to represent exposure duration effect (Liao et al. 2008; U.S. EPA 1988). For long-term chronic exposure, since the dose metric emerging from exposure duration is not a linear or explicit variable, it is difficult to address these effects simply based on mathematical parameterization (Hodgson and Darnton 2000; Philippe and Mansi 1998). The case study on dioxin has successfully illustrated how to use toxicokinetic model to convert external exposure level and exposure duration into a cumulative dose metric, which was further applied in dose-response study (Becher et al. 1998; Crump et al. 2003). To understand the influence of exposure duration to public health requires a toxicokinetic model to appropriately quantify the impact of exposure duration on delivered dose and ultimately risk in a quantitative dose-response framework.

Several toxicokinetic models have been previously developed (El-Masri and Kenyon 2008; Liao et al. 2008; Yu 1999). Based on short-term oral exposures, Yu (1999) developed a seven-compartment physiologically-based pharmacokinetic (PBPK) model for inorganic As (iAs). More recently, El-Masri and Kenyon (2008) published an individual PBPK model that traced the relationships among iAs, monomethylarsenic acid (MMA) and dimethylarsenic acid (DMA) for oral exposure. While these models offered an overview of the absorption, metabolism, distribution and excretion mechanisms in human systems, all such models were developed based on normal people at an individual level. To relate exposure to public health, a PBPK model needs to account for intrinsic heterogeneity at a

population and lifetime scale.

Publicly available data have the potential to support the optimization of population PBPK models for use in quantitative risk assessment (Bernillon and Bois 2000; Lyons et al. 2008), particularly in dose-response study. Specifically, the U.S. FDA has conducted a total diet study (TDS) program to monitor the levels of multiple elements, as well as As, in the country's food supply (Tao and Michael Bolger 1999). Also, the National Health and Nutrition Examination Survey (NHANES) program was initiated to assess the health and nutritional status of adults and children in the United States (Aylward et al. 2014). Fitting of PBPK models to available data using Bayesian methods such as Markov Chain Monte Carlo (MCMC), these publicly available data can be utilized to bridge As exposure and public health. To the best of our knowledge, this type of research has not previously been attempted and represents a novel interpretation of human health from existing data sets.

In this study, the aim is to illustrate how to integrate publicly available data, PBPK model and Bayesian simulation to refine human health risk assessment, using arsenic as a case study. In particular, the objectives include: 1) assessment of As exposure from U.S. TDS; 2) reporting As biomonitoring information based on the latest U.S. NHANES data (2011-2012); 3) optimizing an As population lifetime PBPK model; and 4) improving As non-cancer dose-response study. The newly proposed dose-response study has the potential to protect human health from arsenic exposure.

2. Materials and Methods

2.1. Procedure for Establishing Arsenic Dose Response.

As shown in Figure 1, the procedure for establishing As dose response consisted of three steps. In step 1, a national As exposure assessment was conducted based on TDS data. Then, the urinary As data was retrieved from NHANES database. The As exposure information and urinary As concentration were set as PBPK model input and output, respectively. Therefore a population, lifetime PBPK model was optimized by using Bayesian simulation (step 2). Finally, the optimized PBPK model assisted in As dose-response study (step 3).

2.2. Exposure Assessment.

The U.S. FDA has released analytical results for samples (all the samples in the TDS study were table-ready prior to analysis) collected during 2006-2011 for toxic and nutritional elements (U.S. FDA 2014). The total As concentrations (tAs) in 272 types of foods were

also measured. The foods were collected based on the food list representing the major components of American people's diet. In the meantime the U.S. FDA compiled food consumption data from 9 age subgroups (U.S. FDA 2009). Therefore, the daily tAs exposure (E_{As}) was estimated by multiplying arsenic concentration (C_{As}) and the age-specific consumption amount (A_{As}) for each TDS food:

$$E_{As}=C_{As}\times A_{As} \quad (1)$$

In this study, all 272 types of food were classified into six categories: seafood (exclude fish), rice/bread/wheat, fish, vegetables, meat, wine and others.

Only tAs was available in the current TDS study. Lynch et al. (2014) have evaluated the iAs fraction of tAs in food based on more than 6500 data points. To our knowledge, their research is the most comprehensive available analysis on arsenic forms in food. Thus, the fractions of iAs in different food categories were summarized in this study (Supplementary Material(SM) Table S1), and were used to estimate daily exposure for different forms of As in each food category.

Excepting diet exposure, drinking water was also deemed to be an important pathway for iAs exposure. Xue et al. (2010) have estimated that the daily iAs exposure from drinking water was 0.025±0.104 μg/kg.bw per day (median: 0.002 μg/kg.bw per day) for the U.S. population. Consequently this median value was considered to be geometric mean (GM) of drinking water exposure to help estimate arsenic exposure.

In this study, a log-normal distribution (*LN*) of daily intake was employed to account for population variability:

$$E_{As\text{-}Individual} \sim LN(GM, GSD) \quad (2)$$

The median value of daily arsenic intake (the sum of dietary exposure and drinking water) was used to represent the GM of this log-normal distribution, and the geometric standard deviation (GSD) of iAs intake was estimated to be 1.58, which was based on a previous survey of the general U.S. population (Yost et al. 2004).

2.3. Biomonitoring Data.

The urinary biomarker data, including the tAs, iAs, MMA, DMA, arsenobetaine and arsenocholine was derived from the NHANES (n=4794) (NHANES 2014). The detection rates for tAs, As(III), As(V), MMA, DMA, arsenobetaine, arsenocholine and

trimethylarsine oxide were 96%, 31%, 3%, 27%, 80%, 47%, 4% and 2%, respectively. A log-normal distribution was assumed for urinary concentrations (Aylward et al. 2014), and the maximum likelihood estimation (MLE) was performed to obtain the statistical parameters, including the GM and GSD.

2.4. PBPK Model.

Our PBPK model was derived from previous studies (El-Masri and Kenyon 2008; Liao et al. 2008; Yu 1999). Compared to prior models, this model allowed a lifetime exposure by specifying some age-dependent parameters (Table 1). This human PBPK was a five-compartment model consisting of four well-mixed tissue groups—liver, kidney, lung, and rest of the body—and a mixed blood compartment (Figure 1). A detailed description of the PBPK model and the programming code is available in the SM.

To account for the population variability of current lifetime PBPK model, three strategies were used: firstly, the Gaussian distribution families were assumed for physiological parameters (30% relative standard derivation (RSD) was used (Dong and Hu 2011), the means of theses Gaussian distributions have been provided in Table 1); secondly, a population optimization for sensitive parameters was performed based on a Bayesian hierarchical model (BHM) (Lyons et al. 2008; Wan et al. 2013): to select the sensitive parameters, a prior sensitivity analysis was carried out and the results were shown in Table S2. Consequently, three parameters, i.e. liver/blood partition coefficients for As(III), maximum metabolism rate constant for As(III)-MMA, urinary elimination constants for As(III), were outlined as the sensitive parameters for optimization. Thirdly, a log-normal distribution for daily exposure was adopted (as described in the earlier section).

2.5. Bayesian Simulation.

A BHM was established to estimate the sensitivity parameters as shown in Figure 1. Predicted urinary levels (PULs) through the PBPK model by inputting the sensitive parameters (P_s), exposure time (t) and other model parameters (Φ), and then PULs and observed urinary levels (MULs) were linked through a residual error model (log-normal distribution) with the mean (zero) and variance (σ^2) in the likelihood calculation (Yang et al. 2010). Considering reported As exposure information did not distinguish the specifications of organic arsenic (MMA, DMA, arsenobetaine and arsenocholine), correspondingly we

cannot employ the urinary organic arsenic (oAs) concentrations as the MULs in this study. In contrast, we selected urinary As(III) levels as the MULs as the detection rate for As(V) was too low (3%) to derive reliable statistics for As(V) (NHANES 2014).

Corresponding to Bayesian theory, the estimated posterior probability density function (PPDF) for target parameters was obtained from the product of the joint prior probability density function (pPDF) and the likelihood function. This was done based on the measurement model describing the difference between the model simulation and the observation (Lyons et al. 2008; Sohn et al. 2004). A joint prior probability distribution was encoded as $p(\sigma^2, P_s) = p(\sigma^2) \times p(P_s)$. Hence, the PPDF for σ^2 and P_s can be expressed by Equation (3):

$$P(\sigma^2, P_s | C_{MULs}) \propto p(C_{MULs} | \sigma^2, P_s) \times p(P_s) \times p(\sigma^2) \quad (3)$$

The prior distributions were non-informative for $p(\sigma^2)$ (a uniform distribution with boundary of 0.001-100) and $p(P_s)$ (normal distributions with 500% RSDs, the prior means of these normal distributions have been noted in Table 1). With the prior distributions for P_s and σ^2, the residual error model used in the likelihood function for different age groups (i=1~6) was expressed as Equation (4):

$$Ln(C_{MUT_{s-i}}) = Ln(f(P_s, t, \Phi)) + \varepsilon_i = Ln(C_{PUT_{s-i}}) + \varepsilon_i \quad (4)$$

where ε_i was the error in age group i, which was termed $\varepsilon_i \sim N(0, \sigma^2)$ and f expressed the PBPK model. With the prior distribution for σ^2, the posterior distribution for the parameters of interest was calculated by applying Equation (4) in the likelihood function (Equation (5)):

$$p(C_{MULs} | \sigma^2, P_s) \propto \prod_{i=1}^{6} p(C_{MULs-i} | C_{PULs-i}, P_s) \quad (5)$$

In this log-normal measurement model, 96 individuals (I) were chosen (Dong and Hu 2011) due to the computational time required (approximate 90mins for each individual). MCMC computation was used to optimize the parameters. The Gibbs and Metropolis Hastings (MH) samplers were used to update the object parameters (Xu et al. 2006): 1) the parameter, σ^2, was randomly drawn from the inverse gamma distribution by using the Gibbs sampler; 2) the conditional distributions for P_s have no specific form as the PBPK model is non-linear, and therefore we sampled the P_s by using the Metropolis algorithm.

2.6. Dose-Response Assessment.

Several symptoms have been considered to associate with ingested iAs (Bräuner et al. 2014; Liao et al. 2008), while the skin lesions are considered as the most common symptoms. Thus, the skin lesions of keratosis and hyperpigmentation were selected as the critical non-cancer effects for As exposure in this study. Between April 1995 and March 1996, a survey was conducted to investigate these two effects in West Bengal, India (Mazumder et al. 1998). In all, 7683 participants were examined and interviewed, and the As levels in drinking water were measured (Mazumder et al. 1998). The As levels and age were divided into eight groups (0-50, 50-99, 100-149, 150-199, 200-349, 350-49, 500-799, and 800+ μg/L) and seven age groups (<9, 10-19, 20-29, 30-39, 40-49, 50-59, 60+), respectively (Mazumder et al. 1998). Using the established PBPK model, drinking water iAs concentration (C_w) was converted into urinary iAs concentration (UC_{iAs}), and then cumulative urinary concentration (CUC) was estimated by integrating UC_{iAs} and age (t):

$$UC_{iAs} = f(C_w, P_s, \Phi, t) \quad (6)$$

$$CUC = \int_0^t UC_{iAs} dt \quad (7)$$

In our study, since previous studies indicated that the skin lesions were associated to As-contaminated drinking water and drinking water was considered as major exposure pathway (Bagla and Kaiser 1996; Mandal 1996; Mazumder et al. 1998), only drinking water pathway was included when conducting dose-response study.

Then, the benchmark dose (BMD) approach was employed to estimate the iAs BMD and the lower 95% confidence limit of BMD (BMDL) of the CUC (Davis et al. 2011; Wheeler and Bailer 2009), and finally the $BMDL_{CUC}$ was extrapolated as the reference dose (Rfd). To minimize model uncertainties, seven BMD models were included in this estimation, including Gamma, Dichotomous-Hill, Logistic, Log-logistic, Probit, Logprobit and Weibull models (Davis et al. 2011).

3. Results and Discussion

3.1. Exposure Estimation.

Of the 272 types of food, only the median value of 24 types were above the detection limit (U.S. FDA 2014). Together with consumption data (U.S. FDA 2009), the median of daily dietary tAs exposure was estimated to be 0.15 μg/kg/day (body weight was used as 70 kg). Specifically, the values for age groups 0 - 0.5, 2, 6, 10, 14 - 16, 25 - 30, 40 - 45, 60 - 65, 70+

were 0.24, 0.39, 0.19, 0.18, 0.15, 0.16, 0.15, 0.20 and 0.16 μg/kg/day, respectively (Figure 2, age-specific body weights as presented in Table 1 were used here). These age groups were identical to the classification of age groups by U.S. Food and Drug Administration (U.S. FDA 2009). For young children (<6 years of age), the tAs exposure from food was approximately 1.5 - 2 times higher than that shown for other groups. Figure 2 also identified that the highest contribution to tAs is made by seafood (58.46%, excluded fish), followed by rice/bread/wheat (22.82%) and fish (14.72%).

The median value of estimated daily dietary iAs intake was 0.028 μg/kg/day for all age groups. In particular the daily dietary intakes for As(III) and As(V) were estimated to be 0.020 and 0.0077 μg/kg/day, respectively. Thus, it was concluded that approximately 18.67% of tAs exposure originating from diet was toxic iAs. Specifically, the iAs percentages for age groups 0 - 0.5, 2, 6, 10, 14 - 16, 25 - 30, 40 - 45, 60 - 65, 70+ were 40.67%, 27.56%, 27.37%, 21.34%, 22.19%, 21.44%, 19.43%, 12.76% and 12.38%, respectively. Thus, age differences were observed (Figure 2): young children's daily intake of iAs daily can be up to 0.11 μg/kg/day, which was approximately 4 times higher than older age groups. However, the major contribution to iAs exposure arose from rice/bread/wheat (84% for As(III) and 72% for As(V)), while the rice/bread/wheat only contributed 23% to tAs exposure. Such differences can be explained by the iAs fraction in food commodities (Table S1): although seafood made the biggest contribution to tAs intake, the iAs fraction in seafood was approximately 1.2%. Thus, seafood only contributed 3.66% and 5.38% for As(III) and As(V), respectively. On the other hand, the fractions of As(III) and As(V) in the rice/bread/wheat were up to 49.01% and 15.99%, respectively.

Estimations in current study showed agreement with the urinary excretion: considering the median value for tAs in urine for the general U.S. population was reported to be 8.15 μg/L (Aylward et al. 2014), the daily intake should be approximately 12.68 μg/day. This is assuming the urine volume is about 1.4 L per day and 90% of excretion was estimated by urine (Pomroy et al. 1980): 8.15×1.4/90%=12.68 μg/day. Our dietary tAs exposure estimation was 10.5 μg/day (0.15 μg/kg/day × 70 kg=10.5 μg/day), which is close to the total intake amount of 12.68 μg/day. This estimation also indicated dietary is a major pathway, which has been demonstrated in previous studies (MacIntosh et al. 1996; Yost et al. 2004).

Previous studies have reported daily arsenic exposure in the general U.S. population

(MacIntosh et al. 1997; Xue et al. 2010; Yost et al. 2004). For example, MacIntosh et al. (1997) have reported that the mean dietary intake of tAs was 1.56 μg/kg/day (assuming an adult weight of 70 kg), which was much higher than the median value (0.15 μg/kg/day) in this report. Yost et al. (2004) estimated that for children the tAs exposure was 3.2 μg/day on average, with a range of 1.6~6.2 μg/day, which was similar to the estimations in this study. More recently, using data from 2003-2004 NHANES individual data, the median estimations for tAs and iAs were 0.08 μg/kg/day and 0.02 μg/kg/day, respectively (Xue et al. 2010). Comparing to the estimation for tAs in this study (median: 0.15 μg/kg/day), the estimation for tAs was 2~3 times lower in the previous study (Xue et al. 2010).

Another difference between current study and the previous one is the contributions of food commodities (Xue et al. 2010). One prior analysis showed that rice, wheat and related products contributed about 29% of the iAs intake (Xue et al. 2010), which was much lower than an estimation of 80.68% in this study. An estimation of only 24.32% iAs fraction in rice (Schoof et al. 1999), which was adopt in previous study (Xue et al. 2010) and then resulted in a low contribution being made by rice/bread/wheat. Contrastingly, a board array of literature indicated that iAs fraction in rice was up to 65% (Jorhem et al. 2008; Lynch et al. 2014; Torres-Escribano et al. 2008). Ancillary support provided by the European Food Safety Authority also suggested that on average iAs represents approximately 70% of the tAs content in rice, except for brown rice where on average iAs represents around 80% of tAs content (EFSA 2014). Thus, previous report may underestimate the contribution of iAs from rice/bread/wheat since it adopt a lower iAs fraction in rice/bread/wheat (Xue et al. 2010).

The iAs exposure from drinking water was estimated to be 0.002~0.004 μg/kg/day at a national survey previously (Xue et al. 2010), which was referred to include drinking water to calculate arsenic daily exposure in this study.

3.2. Urinary Arsenic Concentrations.

Of all the biomarkers examined for As exposure in the NHANES subjects (n=4794), the GM and GSD were estimated to be 7.75 μg/L and 3.14, respectively (Table 2). Specifically, DMA and arsenobetaine had relatively high concentrations, with GM of 3.85 μg/L and 1.66 μg/L, respectively, followed by MMA (0.55 μg/L) and As(III) (0.36 μg/L). The age trend for As(III) concentrations has also been statistically analysed (Table 2): the mean As(III) concentrations for age groups 6-9, 10-15, 16-29, 30-44, 45-64 and 65+ were 0.44, 0.49, 0.50, 0.50, 0.40 and 0.33 μg/L, respectively.

This study marks the first one to document As(III) concentration in the general U.S. population. For 2011-2012 NHANES data, the detection limit for As(III) declined sharply from 1.2 μg/L (2009-2010 NHANES) to 0.48 μg/L (2011-2012 NHANES). Thus, the detection rate increased from <5% to 31%, which provided an opportunity to estimate As(III) concentration in general population. Based on a log-normal assumption for As(III), the As(III) concentration using MLE methods was evaluated. A previous study stated that the MLE method has an acceptable error ratio (0.7%), and further simulation indicated that only when the detection rate fell below 25%, did the error ratio dose differ from zero (Croghan and Egeghy 2003). In this study the impact of low detection rate was also simulated (Matlab pseudocode is provided in the SM). The simulated results showed the error ratio was below 5% (detection rate>30%) when the size was 4794 (the population size in this study), which suggested our estimated As(III) may be reliable.

Aylward et al. (2014) observed that the secondary methylation index (SMI, ratio of urinary DMA to MMA) in the NHANES program likewise is much higher in people with measurable arsenobetaine than in those without, suggesting that direct DMA exposure is co-occurring with exposure to arsenobetaine. Such study indicated correlations among urinary DMA, MMA, and arsenobetaine may potentially characterize source exposure (Aylward et al. 2014). Figure 3(d) illustrates a relationship between DMA and MMA ($Ln(DMA)=1.04\times Ln(MMA)+1.83$, n=1280, $p<0.0001$), may indicate direct exposure to these species in seafood or the metabolism of organic arsenicals. Previous analyses did not correlate As(III) and organic arsenic at the national scale due to As(III) concentration was not available. In this study, Figure 3(a)-(c) stated there were significant log-log linear regressions between As(III) and tAs, MMA and DMA. The correlations between As(III) and MMA ($Ln(MMA)=0.55\times Ln(As(III))+0.48$; $r^2=0.35$, n=944, $p<0.001$) were apparently more significant than those between As(III) and DMA ($Ln(DMA)= 0.87\times Ln(As(III))+2.27$; $r^2=0.27$, n=1480, $p<0.001$). This can be explained by the metabolism from MMA to DMA, which would amplify the heterogeneities when addressing the relationship between As(III) and DMA. Such heterogeneities were also propagated when linking As(III) and tAs, which would reduce the fit (Figure 3(a), $Ln(tAs)= 1.04\times Ln(As3)+3.06$; $r^2=0.19$, n=1486, $p<0.001$). These correlations may help trace arsenic exposure in the future.

3.3. PBPK Model Optimisation.

Although nine age groups were used in the TDS, the youngest participant in the NHANES program was 6 years old. Therefore, the daily exposure estimations for the six age groups (as listed in Table 2) were identical to average exposures of (6 yrs, 10 yrs), (10 yrs, 14-16 yrs), (14-16 yrs, 25-30 yrs), (25-30 yrs, 40-45 yrs), (40-45 yrs, 60-65 yrs), and (60-65 yrs, 70 yrs), respectively. The Gelman-Rubin diagnostic method served to test the convergence of the objective parameters (Dong and Hu 2011) and was achieved in this study. The posterior distribution for the three sensitive parameters, including the liver/blood partition coefficient for As(III), maximum metabolism rate constant for As(III)-MMA, and the urinary elimination constant for As(III), were estimated to be 20.93 ± 11.33 (95%CI: 0.95 - 41.19), 5.68×10^{-7} ± 2.85×10^{-7} (95%CI: 0.68×10^{-7}- 1.12×10^{-6}) mol/min and 0.098 ± 0.046 (95% CI: 0.019-0.19) (min^{-1}) (as shown in Table 1), respectively. Comparison with the prior value from previous literature, increases of 26.79% and 9.23% were found for liver/blood partition coefficient and maximum metabolism rate constant for As(III)-MMA, respectively. The increase for liver/blood partition indicated the As(III) partitioned more in the liver, and the increase for maximum metabolism rate constant suggested arsenic is more able to achieve maximum metabolism. On another aspect, the posterior urinary elimination constant, a much higher with a value of 40% increases (comparing to prior value), suggesting that As (III) was excreted more readily in urine.

These parameter updates can be explained by the error between simulation results and observed values (SM Figure S1). Using prior information, the simulated GM±GSD of As(III) for the 6-9, 10-15, 16-29, 30-44, 45-64 and 65+ age groups were 0.19±1.91, 0.24±1.91, 0.41±1.86, 0.44±1.83, 0.39±1.85, 0.33±1.83 μg/L, respectively. The simulated concentrations for 6-9 and 10-15 were 40% lower than the observed values, while those values for other groups were 17%-50% higher than the observed values (corresponding As(III) levels, i.e. 0.32±2.24, 0.40±1.91, 0.35±2.31, 0.35±2.31, 0.28±2.34, 0.22±2.48 μg/L). By using the posterior information, the simulated values (GM±GSD) of As(III) for the six age groups were 0.20±2.34, 0.24±2.29, 0.36±2.19, 0.38±2.14, 0.29±2.16, 0.23±2.13 μg/L, respectively. Generally, the residual error was magnified with the cumulative probability increased due to the positive skewness of the lognormal distribution. Although the relative differences between the 6-9 and 10-15 age groups were still up to 0.38 and 0.40, the average difference for the following four age groups fell to only 0.053, and the overall relative residual sum of squares (RRSS) decreased from 0.85 to 0.31. Through Bayesian inference, crucial parameters in the PBPK model were updated based on the prior distributions, further, calibration of the PBPK

model improved the prediction of biomonitoring data. Therefore, the updated parameters under the constraints imposed by the model structure, model parameters, and the prior exposure, represent more responsible population parameters that can be used to better understand how exposure events are linked.

3.4. Dose Response Assessment.

The drinking water iAs concentration and age were inputted into the established PBPK model to estimate CUC (Equations 4-5). The data for females and males subjects were combined since the PBPK model did not treat the genders separately. Overall incidences of hyperpigmentation and keratosis were 4.56% and 2.01%, respectively. Both types of skin lesions demonstrated a positive age trend, as exemplified the hyperpigmentation incidences for the age groups <9, 10-19, 20-29, 30-39, 40-49, 50-59, >60 were 1.83%, 2.31%, 4.14%, 5.90%, 7.21%, 9.10% and 7.5%, respectively.

Table 3 showed the iAs BMD estimation for different models when BMR were set as 10% and 5%. The estimated iAs $BMDL_{10}$ ranked from 17.06 - 72.65 μg/kg per day, while the iAs $BMDL_5$ were estimated with a range of 8.29 - 46.37 μg/kg per day. Using the keratosis as the critical effect and BMR of 5%, the PoD (Point of Departure) was estimated to be 8.29 μg/kg per day (the lowest BMDL estimation was used). Since the data for dose-response was only stemmed from one report, an uncertainty factor of 10 was considered to account for population variability. Thus, the iAs Rfd was adjusted to be 0.8 μg/kg per day. As stated, current diet iAs daily intake was estimated to be 0.028 μg/kg/day, which suggested the hazard quotient (HQ) was only 0.035. Such a low HQ indicated an insignificant risk for skin lesions when the general U.S. population was exposed to iAs.

Previous studies also functionally parameterized exposure duration to create a link between risk increases and exposure duration (Liao et al. 2008; U.S. EPA 1988). Contrastingly, this study used a PBPK model to include the impacts from exposure duration. For comparisons, the dose-response data was also analysed using a generalized multistage function to parameterize exposure duration (U.S. EPA 1988):

$$p(duration, dose) = 1 - \exp(-(k_0 \times dose \times (duration - k_1)^{k2}) \quad (8)$$

where the parameters k_0, k_1, k_2 were skin lesion-specific best-fitted parameters, and model simulations were provided in SM Table S3, as well as risk-specific dose in SM Table S4.

Using a response (p in Equation 8) of 5%, the Rfd was estimated to be 0.40 μg/kg per day (for hyperpigmentation) when considering the intra-specific UF of 10. Thus, our analysis suggests the previous method may result in a conservative Rfd estimation, since one fold higher Rfd was obtained when using PBPK model. Moreover, using PBPK model to convert age into a dose metric not only took into account the cumulative effect, but also simplified the model fit since it involved fewer variables. A non-straightforward fit will emerge if the models used to fit dose-response model is too complicated. In fact the Weibull model (Liao et al. 2008), was also attempted to parameterize the age-effect in our study, however, the simulated results did not converge (data not shown).

Arsenic Rfd on humans from epidemiological data was previously evaluated by the U.S. EPA's IRIS (U.S. EPA 2012). Using the data from a Taiwanese farming population exposed to arsenic in well water, a chronic RfD of 0.3 μg/kg/day for inorganic arsenic was derived, based on a NOAEL of 0.8 μg/kg/day for skin effects and possible vascular complications. However, the Taiwanese dose-response data is not publicly available currently, which make it is impossible to implement the estimations and comparisons for this population group.

4. Limitations and Conclusions

Some limitations have been acknowledged in this study. The total exposures considered only diet and drinking water, since it was difficult to trace other pathways. This treatment may bring the bias since this value was used as input to optimise the PBPK model parameters. However, previous studies have demonstrated that diet and drinking water were the major exposures, and such estimations agree well with the biomonitoring in our analysis. Also, only As(III) was used for fitting the model parameters and the biomonitoring information for MMA and DMA was discarded: this modelling endeavour omits MMA and DMA. These arsenic species (MMA and DMA) have been known to have high activity and are likely the causes of many of even most of arsenic biological effects (Ahmad et al. 2002; Andrewes et al. 2003). This flaw resulted from that the details of exposure information on oAs is not available currently. Since oAs is much less toxic than the inorganic fraction, such a consideration may have limited impact on assessing toxicity. On another aspect, while cancer may drive the usual arsenic risk assessments, only Rfd based on non-cancer effect is estimated. This consideration is due to the dose-response data is available for hyperpigmentation and keratosis, but the raw data for cancer effects cannot be accessed based on our extensive literature review. Each of these limitations may result in some amount of error or bias into our study, and more available data promises to overcome these limitations.

One major aim of this study is to illustrate how to employ publicly available data inform environmental regulations. Toward the next generation (NexGen) of human health risk assessment strategies, new technologies are being used to collect and organize data streams that promise to reshape our understanding of chemical behaviour (Krewski et al. 2014). By exchanging such data, more hypotheses, methods and conclusions could benefit both researchers and stakeholders. For example, current publicly available datasets (such as ACToR, NHANES, National Morbidity, Mortality, and Air Pollution Study, IRIS) have largely advanced research on human exposure and health outcomes (Fowler 2013), especially when examining the links between public health and exposure to a certain chemical as shown in this study.

In conclusion, not only did we estimate dietary tAs and iAs exposures for the general U.S. population, our study is also the first to report that the fraction of As(III) levels in total arsenic was approximately 4%. Moreover, a population PBPK model was optimised to help derive iAs Rfd of 0.8 μg/kg per day for skin lesions. The framework presented here illustrates how to use publicly available data and computational techniques to help stakeholders make informed decisions.

5. Acknowledgements

We would like to thank the Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC CARE) for funding support, and the Global Centre for Environmental Remediation, University of Newcastle for use of its facilities.

6. Supplementary Materials Available

Information describing the PBPK model, pseudocode for PBPK model, pseudocode to address the impact of low detection rate, fractions of As(III) and As(V) in food, sensitivity analysis results for PBPK parameters, model fit results, risk-specific dose under generalized multistage function, contour of the residual error between the simulated urinary As levels and the observed urinary As levels here are provided.

7. References

Ahmad S, Kitchin KT, Cullen WR. 2002. Plasmid DNA damage caused by methylated arsenicals, ascorbic acid and human liver ferritin. Toxicol Lett 133:47-57.

Andrewes P, Kitchin KT, Wallace K. 2003. Dimethylarsine and trimethylarsine are potent genotoxins in vitro. Chem Res Toxicol 16:994-1003.

Aylward LL, Ramasamy S, Hays SM, Schoeny R, Kirman CR. 2014. Evaluation of urinary speciated arsenic in NHANES: Issues in interpretation in the context of potential inorganic arsenic exposure.

503 Regul Toxicol Pharmacol 69:49-54.
504 Bagla P, Kaiser J. 1996. India's spreading health crisis draws global arsenic experts. Science
505 274:174-175.
506 Becher H, Steindorf K, Flesch-Janys D. 1998. Quantitative cancer risk assessment for dioxins using an
507 occupational cohort. Environ Health Perspect 106:663.
508 Benramdane L, Accominotti M, Fanton L, Malicier D, Vallon JJ. 1999. Arsenic speciation in human
509 organs following fatal arsenic trioxide poisoning - a case report. Clin Chem 45:301-306.
510 Bernillon P, Bois FY. 2000. Statistical issues in toxicokinetic modeling: A Bayesian perspective.
511 Environ Health Perspect 108:883-893.
512 Bräuner EV, Nordsborg RB, Andersen ZJ, Tjønneland A, Loft S, Raaschou-Nielsen O. 2014.
513 Long-term exposure to low-level arsenic in drinking water and diabetes incidence: A prospective
514 study of the diet, cancer and health cohort. Environ Health Perspect 122:1059.
515 Brown R, Delp M, Lindstedt S, Rhomberg L, Beliles R. 1997. Physiological parameter values for
516 physiologically based pharmacokinetic models. Toxicol Ind Health 13407:484.
517 Croghan C, Egeghy P. 2003. Methods of dealing with values below the limit of detection using SAS.
518 Southeastern SAS User Group, St Petersburg, FL:22-24.
519 Crump KS, Canady R, Kogevinas M. 2003. Meta-analysis of dioxin cancer dose response for three
520 occupational cohorts. Environ Health Perspect 111:681.
521 Davis JA, Gift JS, Zhao QJ. 2011. Introduction to benchmark dose methods and US EPA's benchmark
522 dose software (BMDS) version 2.1. 1. Toxicol Appl Pharmacol 254:181-191.
523 Dong Z, Hu J. 2011. Development of lead source-specific exposure standards based on aggregate
524 exposure assessment: Bayesian inversion from biomonitoring information to multipathway
525 exposure. Environ Sci Technol 46:1144-1152.
526 EFSA. 2014. Dietary exposure to inorganic arsenic in the European population.
527 El-Masri HA, Kenyon EM. 2008. Development of a human physiologically based pharmacokinetic
528 (PBPK) model for inorganic arsenic and its mono-and di-methylated metabolites. J
529 Pharmacokinet Pharmacodyn 35:31-68.
530 Fowler BA. 2013. Computational toxicology: Methods and applications for risk assessment:Academic
531 Press.
532 Hodgson JT, Darnton A. 2000. The quantitative risks of mesothelioma and lung cancer in relation to
533 asbestos exposure. Annals of Occupational Hygiene 44:565-601.
534 Jorhem L, Åstrand C, Sundström B, Baxter M, Stokes P, Lewis J, et al. 2008. Elements in rice from the
535 swedish market: 1. Cadmium, lead and arsenic (total and inorganic). Food Addit Contam
536 25:284-292.
537 Krewski D, Westphal M, Andersen ME, Paoli GM, Chiu WA, Al-Zoughool M, et al. 2014. A
538 framework for the next generation of risk science. Environ Health Perspect 122:796-805.
539 Liao C-M, Lin T-L, Chen S-C. 2008. A weibull-PBPK model for assessing risk of arsenic-induced skin

lesions in children. Sci Total Environ 392:203-217.
Lynch HN, Greenberg GI, Pollock MC, Lewis AS. 2014. A comprehensive evaluation of inorganic arsenic in food and considerations for dietary intake analyses. Sci Total Environ 496:299-313.
Lyons MA, Yang RS, Mayeno AN, Reisfeld B. 2008. Computational toxicology of chloroform: Reverse dosimetry using Bayesian inference, Markov Chain Monte Carlo simulation, and human biomonitoring data. Environ Health Perspect 116:1040-1046.
MacIntosh D, Williams P, Hunter D, Sampson L, Morris S, Willett W, et al. 1997. Evaluation of a food frequency questionnaire-food composition approach for estimating dietary intake of inorganic arsenic and methylmercury. Cancer Epidemiology Biomarkers & Prevention 6:1043-1050.
MacIntosh DL, Spengler JD, Ozkaynak H, Tsai L-h, Ryan PB. 1996. Dietary exposures to selected metals and pesticides. Environ Health Perspect 104:202.
Mandal BK. 1996. Arsenic in groundwater in seven districts of west bengal, india-the biggest arsenic calamity in the world. Current Scienc 70:976-986.
Maull EA, Ahsan H, Edwards J, Longnecker MP, Navas-Acien A, Pi J, et al. 2012. Evaluation of the association between arsenic and diabetes: A national toxicology program workshop review. Environ Health Perspect 120:1658-1670.
Mazumder DNG, Haque R, Ghosh N, De Binay K, Santra A, Chakraborty D, et al. 1998. Arsenic levels in drinking water and the prevalence of skin lesions in west bengal, india. Int J Epidemiol 27:871-877.
Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, Thompson C, et al. 2013. The broad scope of health effects from chronic arsenic exposure: Update on a worldwide public health problem. Environ Health Perspect 121:295-302.
NHANES. 2014. NHANES 2011-2012 laboratory data, national health and nutrition examination survey. Available: http://wwwn.Cdc.Gov/nchs/nhanes/search/datapage.Aspx?Component=laboratory&cyclebeginyear=2011.
Philippe P, Mansi O. 1998. Nonlinearity in the epidemiology of complex health and disease processes. Theoretical medicine and bioethics 19:591-607.
Pomroy C, Charbonneau S, McCullough R, Tam G. 1980. Human retention studies with 74 As. Toxicol Appl Pharmacol 53:550-556.
Saady JJ, Blanke RV, Poklis A. 1989. Estimation of the body burden of arsenic in a child fatally poisoned by arsenite weedkiller. J Anal Toxicol 13:310-312.
Schoof R, Yost L, Eickhoff J, Crecelius E, Cragin D, Meacher D, et al. 1999. A market basket survey of inorganic arsenic in food. Food and Chemical Toxicology 37:839-846.
Sohn MD, McKone TE, Blancato JN. 2004. Reconstructing population exposures from dose biomarkers: Inhalation of trichloroethylene (TCE) as a case study. J Expo Sci Environ Epidemiol 14:204-213.

Tao SS-H, Michael Bolger P. 1999. Dietary arsenic intakes in the united states: FDA total diet study, September 1991-December 1996. Food Additives & Contaminants 16:465-472.

Torres-Escribano S, Leal M, Vélez D, Montoro R. 2008. Total and inorganic arsenic concentrations in rice sold in Spain, effect of cooking, and risk assessments. Environ Sci Technol 42:3867-3872.

U.S. EPA. 1988. Special report on ingested inorganic arsenic. Environmental Protection Agency: Washington, D.C. .

U.S. EPA. 2012. Arsenic, inorganic. Integrated risk information system (IRIS). Washington, DC: U.S. Environmental Protection AgencyWashington, DC: U.S. Environmental Protection Agency.

U.S. FDA. 2009. 2003 food list + 1994-96, 1998 CSFII data. Available: http://www.Fda.Gov/food/foodscienceresearch/totaldietstudy/ucm184232.htm.

U.S. FDA. 2014. Total diet study elements results summary statistics: Market baskets 2006 through 2011. Available: http://www.Fda.Gov/downloads/food/foodscienceresearch/totaldietstudy/ucm184301.pdf.

Wan Y, Zhang K, Dong Z, Hu J. 2013. Distribution is a major factor affecting bioaccumulation of decabrominated diphenyl ether: Chinese sturgeon (acipenser sinensis) as an example. Environ Sci Technol 47:2279-2286.

Wheeler MW, Bailer AJ. 2009. Comparing model averaging with other model selection strategies for benchmark dose estimation. Environ Ecol Stat 16:37-51.

Xu T, White L, Hui D, Luo Y. 2006. Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. Global Biogeochemical Cycles 20, GB2007.

Xue J, Zartarian V, Wang S-W, Liu SV, Georgopoulos P. 2010. Probabilistic modeling of dietary arsenic exposure and dose and evaluation with 2003-2004 NHANES data. Environmental health perspectives (Online) 118:345.

Yang Y, Xu X, Georgopoulos PG. 2010. A Bayesian population PBPK model for multiroute chloroform exposure. Journal of Exposure Science and Environmental Epidemiology 20:326-341.

Yost L, Tao S-H, Egan S, Barraj L, Smith K, Tsuji J, et al. 2004. Estimation of dietary intake of inorganic arsenic in US children. Human and Ecological Risk Assessment 10:473-483.

Yu D. 1999. A physiologically based pharmacokinetic model of inorganic arsenic. Regul Toxicol Pharmacol 29:128-141.

608 **List of Tables**

612 **TABLE 1.** PBPK parameters for arsenic

Parameters		Values			
Physiological Parameters (Brown et al. 1997)					
Body Weight (bw) (kg)		$0.00059\times age^3 -0.093\times age^2 +4.58\times age+2.96$			
Tissue volume fractions (%)					
Liver		2.57			
Kidney		0.44			
Lung		0.76			
Others		96.23			
Cardiac Output, QC (L/min)		$14.1\times bw^{0.75}$			
Tissue blood flow fractions (%)					
Liver		5.96			
Kidney		19.24			
Others		74.80			
Partition Coefficients (Benramdane et al. 1999; Saady et al. 1989)		As3	As5	MMA	DMA
Liver		20.92[a]	15.8	3.3	3.3
Kidney		11.7	8.3	4.4	3.8
Lung		6.7	2.1	1.3	1.3
Others		7.3	7.6	2.6	2.4
Metabolism Parameters (Yu 1999)[d]		As3 to MMA		As3 to DMA	MMA to DMA
Maximum metabolism rate constant, V_{max} (mol/min)	Liver	5.68×10^{-7} [b]		1.04×10^{-6}	7.41×10^{-7}
	Kidney	3.47×10^{-7}		4.63×10^{-7}	2.31×10^{-7}
Michaelis-Menten constant, K_m (mol/L)	Liver	1.00×10^{-4}		1.00×10^{-4}	1.00×10^{-4}
	Kidney	1.00×10^{-4}		1.00×10^{-4}	1.00×10^{-4}
The other Parameters (Yu 1999)		As3	As5	MMA	DMA
Uptake (min^{-1})		0.004	0.003	0.007	0.007
Urine elimination (min^{-1})		0.098[c]	0.07	0.3	0.13
Second-order rate ($mol^{-1}.min^{-1}$)		0.12			
Biliary elimination (min^{-1})		3.00×10^{-4}			
Absorption fraction (%)		90			
GSH concentration (mol/L)					
Liver		1.50×10^{-2}			
Kidney		5.00×10^{-3}			
Lung		5.00×10^{-3}			
Others		5.00×10^{-3}			

613 Most parameters were adopt from previous studies, except the parameters were optimized using Bayesian technique for: a)
614 liver/blood partition coefficients for As(III), prior mean is 16.5; b) maximum metabolism rate constant for
615 As(III)-MMA, prior mean is 5.2×10^{-7}; c) urinary elimination constants for As(III), prior mean is 0.07.
616
617 Note: d, the reference values are for 70kg adult, SM Equation 2.

Table 2. Statistical information for arsenic concentration in urine (n=4794)

Basic statistics (µg/L)				
As speciation	Detection limit	Detection rate (%)	GM (GSD)	Mean (STD)
tAs	1.25	96	7.75 (3.14)	14.91 (24.52)
As(III)	0.48	31	0.31 (2.36)	0.45 (0.47)
As(V)	0.87	3	NA	
MMA	0.89	27	0.55 (2.15)	0.74 (0.66)
DMA	1.80	80	3.85 (2.48)	5.82 (6.58)
Aresenobetaine	1.19	47	1.66 (5.00)	6.06 (21.29)
Arsenocholine	0.28	4	NA	
Trimethylarsine Oxide	0.25	2	NA	

Age-specific biomonitoring for As(III) (µg/L)		
Age	GM (GSD)	Mean (STD)
6-9	0.32 (2.24)	0.44 (0.42)
10-15	0.40 (1.91)	0.49 (0.36)
16-29	0.35 (2.31)	0.50 (0.50)
30-44	0.35 (2.31)	0.50 (0.50)
45-64	0.28 (2.34)	0.40 (0.41)
65+	0.22 (2.48)	0.33 (0.38)

Abbreviations. tAs: total arsenic; MMA: monomethylarsonic acid; DMA: dimethylarsinic acid.

622 **Table 3.** Benchmark dose (BMD) estimations (μg/kg/day) using various BMD models for
623 inorganic arsenic exposure (p>0.1)

		Gamma	Hill	Logistic	Loglogistic	Probit	Logprobit	Weibull
Hyperpigmentation	BMD_{10}	58.75	61.68	72.65	58.73	66.77	60.71	58.99
	$BMDL_{10}$	51.03	49.79	67.22	50.82	61.51	51.22	51.30
	BMD_5	30.34	27.17	53.61	30.06	47.77	28.29	30.43
	$BMDL_5$	26.96	22.87	49.58	26.66	44.02	24.86	27.03
Keratosis	BMD_{10}	27.10	20.34	51.02	26.53	46.37	25.08	27.10
	$BMDL_{10}$	24.86	17.06	47.72	24.19	43.32	22.63	24.86
	BMD_5	13.21	9.65	34.24	12.58	30.30	11.45	13.21
	$BMDL_5$	12.12	8.29	31.98	11.47	28.31	10.31	12.12

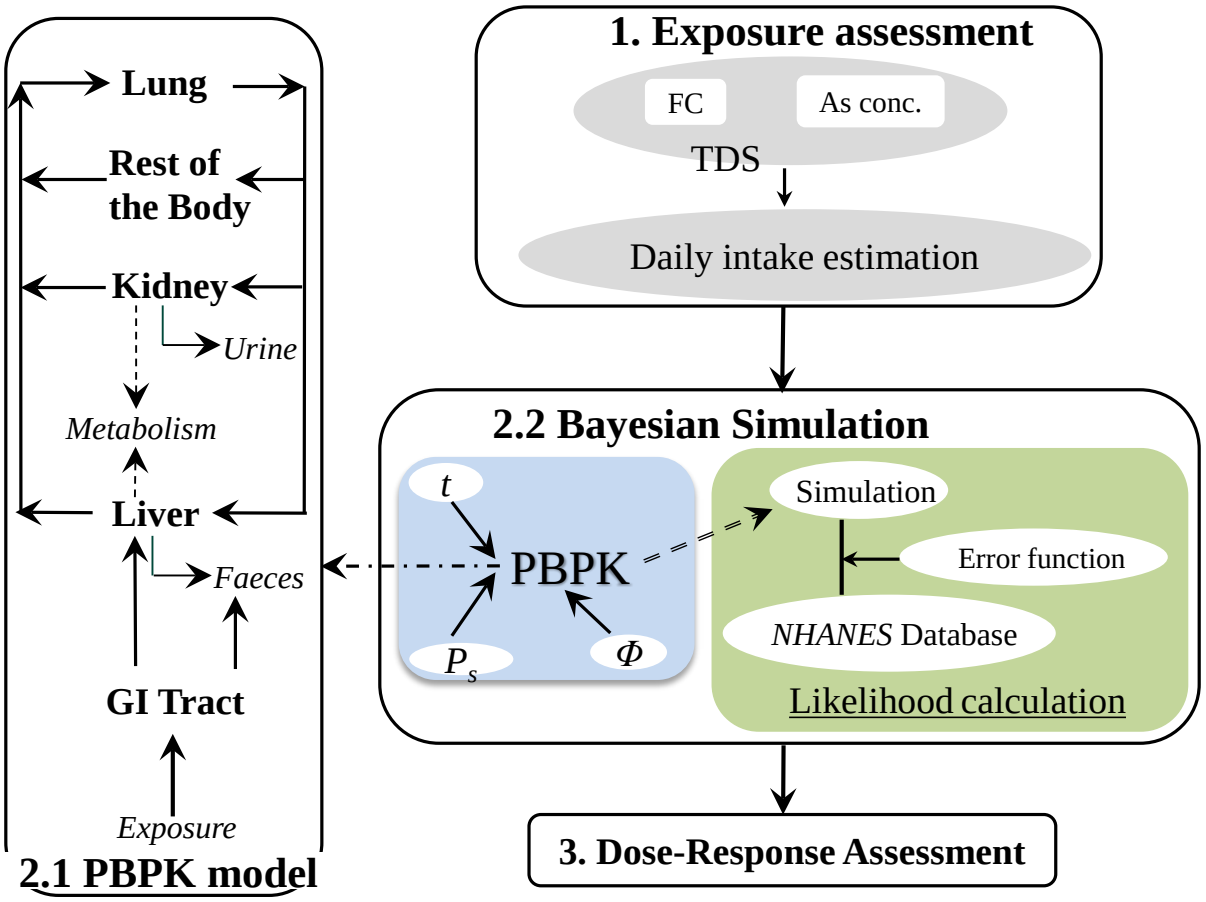
List of Figures.

Figure 1. Framework for establishing dose response.

Abbreviations. FC: food consumption; con.: concentration; TDS: total diet study; GI: gastrointestinal; PBPK: physiologically-based pharmacokinetic model; t: time; P_s: sensitive parameters; ϕ: other parameters; NHANES: national health and nutrition examination survey.

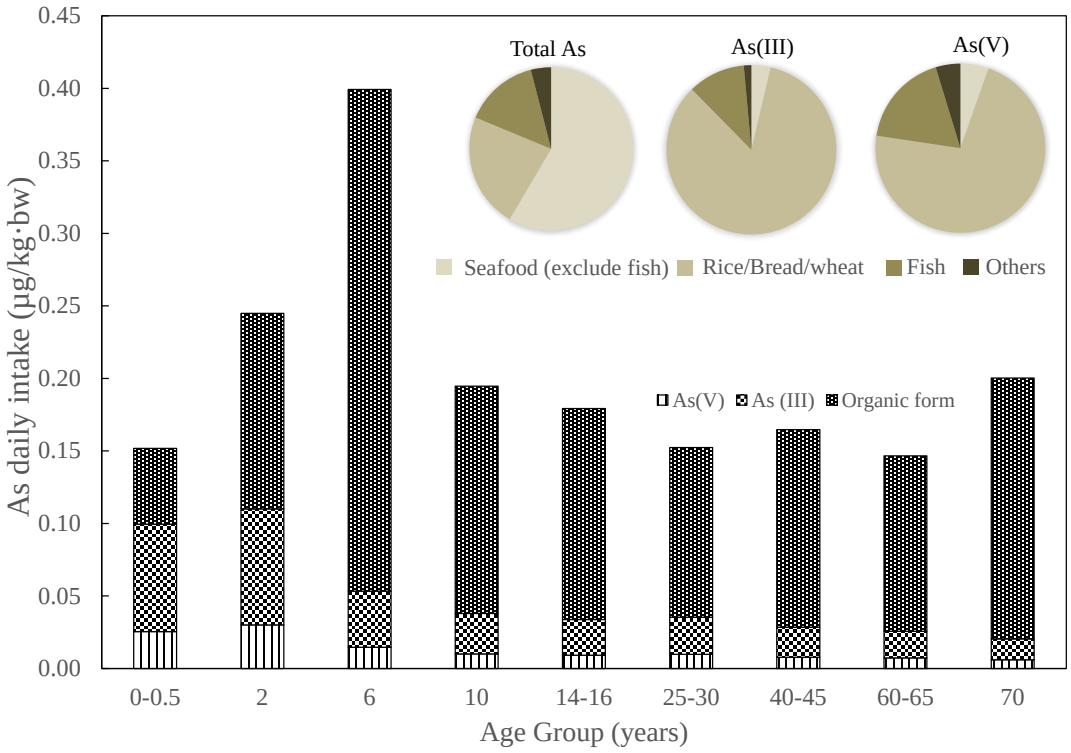
Figure 2. The daily intake for total Arsenic, As(III) and As(V), and contributions by foods.

Figure 3. Scatter plot for arsenic forms in urine: (a) total arsenic (y) and As^{III} (x); (b) monomethylarsonic acid (y) and As^{III} (x); (c) dimethylarsinic acid (y) and As^{III} (x); (d) dimethylarsinic acid (y) and monomethylarsonic acid (x). The data points in red color are considered to be outliers.



637

638 **Figure 1.** Framework for establishing dose response.
639 Abbreviations. FC: food consumption; con.: concentration; TDS: total diet study; GI:
640 gastrointestinal; PBPK: physiologically-based pharmacokinetic model; t: time; P_s: sensitive
641 parameters; ϕ: other parameters; NHANES: national health and nutrition examination
642 survey.



643
644 **Figure 2.** The daily intake for total Arsenic (tAs), As(III) and As(V), and contributions by
645 foods
646

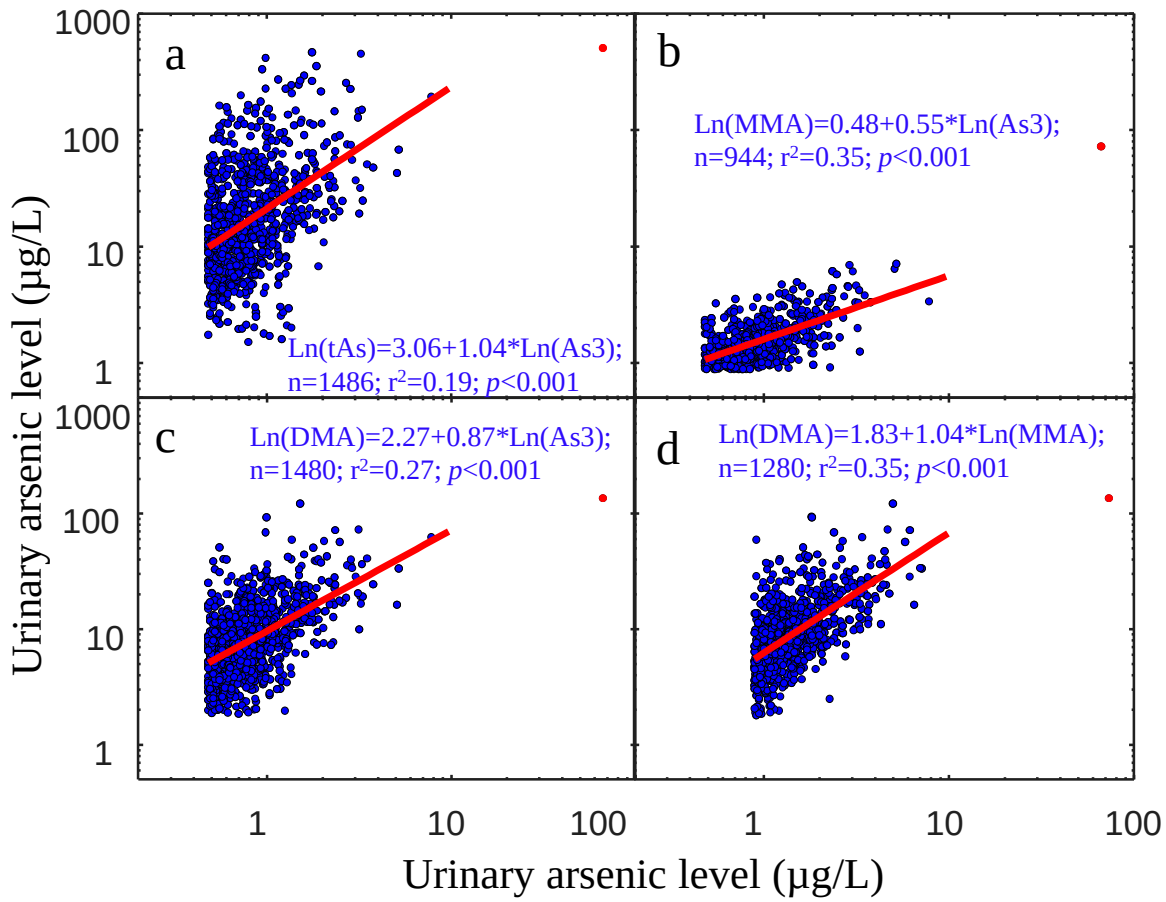


Figure 3. Scatter plot for arsenic forms in urine: (a) total arsenic (*y*) and As^{III} (*x*); (b) monomethylarsonic acid (*y*) and As^{III} (*x*); (c) dimethylarsinic acid (*y*) and As^{III} (*x*); (d) dimethylarsinic acid (*y*) and monomethylarsonic acid (*x*). The data points in red color are considered to be outliers.