BAYESIAN OPTIMIZATION OF PHYSIOLOGICALLY BASED PHARMACOKINETIC MODELS OF DIOXIN EXPOSURE IN CHINESE POPULATION

Tao Ying[1], Haitao Shen[2], S Wen[3], Jingguang Li[4], Yongning Wu[4]

[1]School of Public Health/Key Laboratory of Public Health Safety, Ministry of Education, Department of Nutrition and food science, Fudan University, Shanghai 200032, China, 20211020041@fudan.edu.cn. [2]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, 310051, China.[3]Hubei Provincial Key Laboratory for Applied Toxicology, Hubei Provincial Center for Disease Control and Prevention, Wuhan 430079, China. [4]Food Safety Research Unit of Chinese Academy of Medical Science (2019RU014), NHC Key Lab of Food Safety Risk Assessment, China National Center for Food Safety Risk Assessment, Beijing 100022, China.

Introduction

Polychlorinated dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs) and biphenyls (PCBs) are persistent organic pollutants that are omnipresent in the global environment, belonging to a family of organic compounds called 'dioxins' or dioxin-like chemicals (DLCs). More than 90% of human environmental exposure to dioxins is through the diet, and the majority of body burden of dioxins is contained in the liver and adipose tissues [1], and the ratio of dioxin concentrations in blood and breast to adipose tissue was close to 1[2]. Dioxins can efficiently bind to an intracellular receptor, aryl hydrocarbon receptor (AhR), and then activate transcription of target genes coding for xenobiotic-metabolizing enzymes. The elimination of dioxins is mainly through the hepatic metabolism, fecal excretion, and dilution by growth.

EFSA used the no observed adverse effect level (NOAEL) of the Russian Children's Study (median serum level of 7.0 pg TEQ/g fat for the sum of PCDD/F-TEQ in the lowest quartile) as reference point for the Health-Based Guidance Value and for derivation of the human exposure associated with this serum concentration at the age of 9 years[2]. Since more critical effects were observed in children exposed before the age of 10 years, a model considering dioxin levels in breast milk and the duration of breastfeeding influence is more suitable.

We reviewed and simulated previously published human dioxin toxicokinetic models. Models published by Ruiz et al. [3] and Emond et al. [4] are both inclusive of induction of liver CYP enzymes, dilution by growth and critical elimination pathway, and they have been verified based on human data collected in a variety of human exposure scenarios. Figure 1A showed the simulations of dioxin accumulation in children (repeated exposure to 0.5 pg TEQ kg^{-1} bw day^{-1} for 9 years). For Emond model and Ruiz model scenario 2, no breastfeeding is considered. In Ruiz model scenario 1, dioxin concentration in milk and breastfeeding duration was 5 pg TEQ g^{-1} lipid and 12 months, respectively. The results of the two models are very close when breastfeeding is not considered, while the estimated value is higher in Ruiz model when breastfeeding is considered. Considering the importance of breastfeeding to the dioxin body burden in children, we selected Ruiz model for Bayesian optimization. In addition, Ruiz model was optimized and then adopted to aid risk assessment of dioxins by EFSA in 2018 [2].However, dioxin models are based on European and American population data, whose characteristics differ from Chinese. Interindividual variability in anatomical and physiological properties results in differences in dioxin toxicokinetics. Bayesian probabilistic methodology with Markov chain Monte Carlo (MCMC) simulation has recently been recommended as a method to quantitatively address both variability and uncertainty in PBPK modeling, where available prior knowledge in the form of probability distributions is updated with information extracted from experimental data in the so-called posterior distribution. Thus, the objectives of this work were to optimize the dioxin PBPK model by MCMC based on Chinese datasets and provide basis for the establishment of dioxin exposure limit in China.

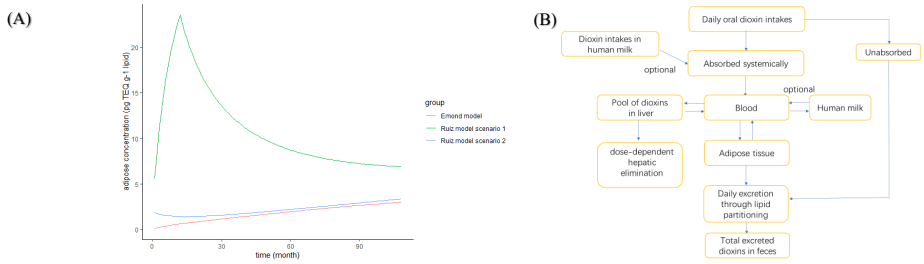

Figure 1. Comparisons of two dioxin PBPK model for humans in different scenarios (A) and structure of the dioxin PBPK model for humans (B).

Materials and methods

2.1 Model structure and physiological parameters

Model published by Ruiz et al., also known as The Concentration- and Age-Dependent Model (CADM), was developed by Carrier et al.[1], modified by Aylward et al.[5] and adapted by EFSA[2].As illustrated in **Figure 1B**, partitioning between fat and liver is concentration dependent, and the proportion of the body burden occurring in

the liver follows a Michalis-Menten relationship with body burden. The growth curves were from Emond et al. for men and women [4]. Table 1 showed important model parameters and definitions.

Table 1. Important CADM parameters and definitions.

Parameters	Description
fmin	Minimum proportion of body burden distributed to liver (%)
fmax	Maximum proportion of body burden distributed to liver (%)
K_half	Body concentration for half-maximum increase in liver distribution proportion (ng/kg)
fab	Absorption rate of dioxins (%)
ka	Rate constant for elimination based on partitioning from circulating lipids into large intestine (month-1)
ke	Rate constant for hepatic elimination (month-1)

2.2 Model evaluation

2.2.1 Model implementation and validation in R

To implement the Bayesian-MCMC method, we recoded the CADM model in R with package "mrgsolve".

We conducted the simulation of dioxin accumulation in the female after repeated exposure to 0.5 pg TEQ kg^{-1} bw day^{-1} for 30 years both in R and in Berkeley Madonna. Coefficient of determination (R^2) and mean absolute percentage error (MAPE) were calculated to evaluate the consistency of simulations from Berkeley Madonna to R. R version was highly consistent with Berkeley Madonna version (R^2=1.00, MAPE=0.2%).

2.2.2 Model training datasets

An extensive search of dioxin exposures datasets was performed to collect data for model calibration and evaluation. Li et al.[6 7] estimated dietary intake of PCDD/Fs in 12 districts in China in 2000 and analyzed human milk samples in the same districts in 2009. Another population exposure data in Beijing was obtained in assessments of dietary PCDD/Fs intake conducted by Zhang et al.[8] and PCDD/Fs milk level analyzed by Bao et al.[9]. In addition, we obtained unpublished data included 16 individual samples with dietary exposure and feces concentration measured.

2.3 Bayesian approach to optimize the CADM

2.3.1 Parameters prior distributions

The fitting results of literatures showed that ke and ka conform to the lognormal distribution with geometric mean and geometric standard deviation provided, and K_half conforms to the normal distribution with the mean and standard deviation provided. Fab, fmin, and fmax are assumed to be normally distributed and the mean, minimum and maximum values are from literatures, with the standard deviation is 2*max{(max-mean), (mean-min)}[1 3 5].

2.3.2 Bayesian-MCMC simulation

From Bayes' theorem, the joint posterior distribution of the parameter is proportional to the likelihood of data multiplied by the prior distribution. MCMC simulations were performed in the R software package “FME”, which was developed particularly for the non-linear model and MCMC simulations. The Delayed Rejection Adaptive Metropolis (DRAM) sampling was used to update parameters. Three Markov chains of 20,000 iterations were run with the first 5000 iterations as “burn-in” iterations. Three Markov chains use different parameter starting points, respectively: 1) The parameter nonlinearly fitted by the Nelder-Mead method; 2) The parameters used by EFSA; 3) The lower boundary of the parameters. Gelman-Robin method was used to perform a convergence test to obtain the potential scale reduction factors ($\widehat{R}$) and its 95% confidence interval, and the Gelman-Rubin-Brooks diagram visually displayed the convergence test results.

2.3.3 Model validation

Root-mean-squared-error (RMSE) was used to assess the goodness of model fitting. In addition to the datasets mentioned above for model parameter optimization, we obtained unpublished data included 3 districts in Hubei Province. A total of 46 men were selected from 3 districts (n>5 for every district) and serum PCDD/Fs concentrations were measured. Dietary dioxin exposures were obtained by multiplying dietary food consumptions by average PCDD/Fs levels in local food.

2.4 Application of optimized model

PBPK model can be used to estimate the intake leading to the critical serum levels or body burden. We simulated the original and optimized model for exposure of boys during the first 9 years in life based on the EFSA standards establishment process. Then, we used Monte Carlo simulation of human variation in intake and toxicokinetic with optimized model. We used the truncated posterior distribution of parameters, a range of average lifetime dioxin (0.1-1 pg TEQ kg^{-1} bw day^{-1}) and a range of duration of breastfeeding (0-12 months) to simulate the dioxin level in human body.

Results and discussion

3.1 Model training datasets

Li et al. [6] estimated dietary intake of PCDD/Fs in 12 districts (the first 4 rows in Table2) in China in 2000 and daily intake among regions ranged 0.09–0.59 pg TEQ kg^{-1} body eight day^{-1}. In the same 12 districts, human milk samples were analyzed for PCDD/Fs in 2009[7] and the range of milk concentration ranged 3.0-4.7 pg TEQ g-1 bw. The published datasets used in this study were listed in Table 2 with area, mean age, dietary exposure, human milk level received at birth and human milk level themselves. In unpublished individual datasets, there were 8 men and 8 women, evenly distributed in the four age groups of 10, 20, 40 and 60 years old. The 14-day duplicate diets method was used to evaluate the dietary intake of total TEQ (PCDD/Fs and dl-PCBs) and 14-day fecal samples was collected to measure the total TEQ. The average dietary intake was 0.93 pg TEQ kg^{-1} bw day^{-1}, and the average fecal concentration is 0.55 pg TEQ kg^{-1} day^{-1}.

Table 2. Chinese population-based dioxin datasets.

Area	Mean age	Milk concentration* (pg TEQ kg-1 bw)	Dietary dioxins intake (pg TEQ kg-1 bw day-1)	Milk Concentration* (pg TEQ g-1 bw)
Heilongjiang, Liaoning, Hebei	26.0		0.41	3.8
Shanxi, Ningxia, Henan	25.3		0.09	3.0
Shanghai, Fujian, Jiangxi	25.5	2.6	0.59	4.7
Hubei, Sichuan, Guangxi	25.5		0.22	3.6
Beijing	28.5		0.20	4.5

3.2 Convergence analysis

Figure 3 showed the Gelman-Rubin-Brooks plots of three chains, suggesting good convergences among chains for each parameter. Corrected Scale Reduction Factors ($\widehat{R}$) were calculated for the three chains based on the method of Brooks and Gelman and the values were between 1.0 and 1.02, and $\widehat{R}$ <1.2 was correspond to equilibrium posterior parameter distribution.

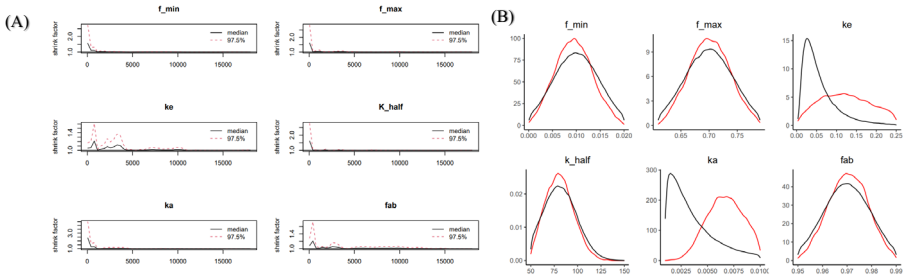

Figure 3. Model Bayesian-MCMC optimization with (A) Gelman-Rubin-Brooks plots of convergence of the Markov chains and (B) densities of posterior parameter uncertainty distributions. In plot (A), it shows the evolution of Gelman and Rubin's shrink factors as the number of iterations increase and the shrink factors and their 95% CI are close to 1, indicating three Markov chains converged. In plot (B), the x-axis represents the value of each parameter and the y-axis represents the densities of the parameter uncertainty distributions. The black lines refer to the densities of the prior distribution. The red lines refer to the densities of the posterior distribution.

3.3 Estimation of posterior parameter distributions

The prior and posterior distributions of the means and standard deviations for the estimated parameters were shown in Table 3. The medians of posterior distributions for K_half, fmin, fmax and fab were close to prior estimates and the standard deviation were substantially lower than prior distributions. In particular, the posterior distribution of ke and ka, rate constants for dioxin elimination, were substantially greater and had flatter distributions compared to prior information. Ruiz et al. also found the model predicts dioxin serum lipids levels much better using a ke of 0.074 per month rather than 0.05[3]. This finding indicated that Chinese may have higher dioxin elimination rates than European and American.

Table 3. The prior and posterior distributions of the estimated parameters.

Parameters	Prior distribution type	Prior distribution		Posterior distribution		Min	Max
		Mean	SD	Mean	SD		
ke	log-normal	0.046*	2.14*	0.126	0.06	0.0034	0.25
ka	log-normal	0.0028*	2.31*	0.0066	0.0017	0.001	0.1

K_half	normal	80.1	18.9	80.16	15.1	50	150
fmin	normal	0.01	0.005	0.0096	0.004	0.0001	0.02
fmax	normal	0.7	0.045	0.70	0.036	0.061	0.079
fab	normal	0.97	0.01	0.97	0.008	0.95	0.99

*geometric means and geometric standard deviations

3.3 Comparisons of model predictions with train data

The RMSE of the optimized parameters were lower than original parameters for population-based datasets (39.9% improvement), for individual datasets (10.6% improvements) and for independent population-based datasets (66.6% improvement). Figure 4 showed comparisons of model predictions (y-axis) with observed data (x-axis) from the original model (red dots) and model after optimization (green dots) with 16 individual datasets and independent population-based datasets. In Figure 4B, serum PCDD/Fs concentration were obviously overestimated in original model while the optimized model predicted the data reasonably well.

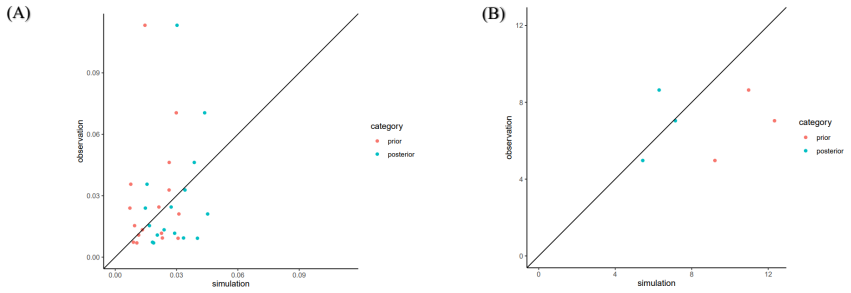

Figure 4. Comparisons of model predictions (y-axis) with observed data (x-axis) from the original model (red dots) and model after parameter optimization (green dots) with (A) 16 individual datasets and with (B) independent population-based datasets. The solid black diagonal line represents the unity line where the observed value and the predicted value are equal.

3.4 Application of optimized model

EFSA decided to use a NOAEL of 7.0 pg TEQ g-1 bw in blood sampled at age 9 years based on PCDD/F-TEQs [2]. We used original and optimized model for derivation of the human exposure associated with this serum concentration at the age of 9 years (shown in Figure 5A). In the original model, the simulations indicate that following breastfeeding for 12 months with milk containing 5 pg TEQ g-1 bw, the intake should be below 0.6 pg TEQ kg-1 bw day-1 in order not to reach a serum concentration of 7.0 pg TEQ g-1 bw. In the optimized model, the intake should be below 1.05 pg TEQ kg-1 bw day-1. And in scenarios without breastfeeding, the recommended value were 1.15 and 1.4 pg TEQ kg-1 bw day-1 for original model and optimized model, respectively.

We run 20 Monte Carlo simulations for both men and women and the results were showed in Figure 5B. Adipose dioxin concentrations in population-based datasets (dots) and the 30-year concentration versus time profiles for optimized model simulations were based on a range of average dietary exposure levels (0.1-1 pg TEQ kg-1 bw day-1) and variation of other model parameters (from posterior distributions) in the Monte Carlo exercise. These uncertainties compel consideration when applying these toxicokinetic models to human epidemiological studies.

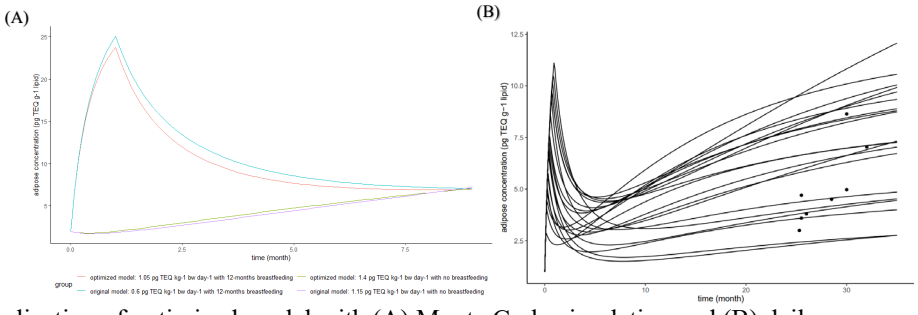

Figure 5. Application of optimized model with (A) Monte Carlo simulation and (B) daily exposure estimation from critical serum dioxin concentration.

This study was the first to adapt PBPK models of dioxin based on Chinese data using Bayesian method. The parametrically modified models are more in line with the metabolic characteristics of dioxin in the Chinese population. The parameters fitting methods could be used under different exposure scenario to achieve accurate prediction, and the optimized models will help establish Chinese dietary dioxin exposure Health-Based Guidance Value and improve the food safety risk assessment.

Acknowledgements
This study was supported by the National Key R&D Program of China of Ministry of Science and Technology of the People's Republic of China (No. 2017YFC1600500).

References
1. Carrier G, Brunet RC, Brodeur J.(1995) *Toxicol Appl Pharmacol.* 131(2):253-266.
2. Chain EPoCitF, Knutsen HK, Alexander J, et al.(2018) *EFSA Journal.* 16(11):e05333.
3. Ruiz P, Aylward L, Mumtaz M.(2014) *SAR QSAR Environ Res.* 25(11):873-890.
4. Emond C, Ruiz P, Mumtaz M.(2017) *Toxicol Appl Pharmacol.* 315:70-79.
5. Aylward LL, Brunet RC, Carrier G, et al.(2005) *J Expo Anal Environ Epidemiol.* 15(1):51-65.
6. Li J, Wu Y, Zhang L, et al.(2007) *Food Addit Contam.* 24(2):186-193.
7. Li J, Zhang L, Wu Y, et al.(2009) *Chemosphere.* 75(9):1236-1242.
8. Zhang L, Yin S, Wang X, et al.(2015) *Chemosphere.* 137:178-184.
9. Bao Y, Zhang L, Liu X, et al.(2020) *Ecotoxicol Environ Saf.* 194:110339.