ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Ecole Doctorale ABIES
Agriculture Agriculture Alimentation Food Biologie Biology Environnement Environment Santé Health

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement

(AgroParisTech)

Spécialité : Santé publique – Mathématiques appliquées

présentée et soutenue publiquement par

Camille BÉCHAUX

le 24 novembre 2014

Modélisation dynamique de l'exposition intégrant la toxicocinétique pour améliorer l'évaluation du risque à des substances chimiques

Directeur de thèse : **Stéphan CLÉMENÇON**
Co-encadrement de la thèse : **Amélie CRÉPET**

Jury

M. Frédéric BOIS, Professeur, Unité Modèles pour l'Ecotoxicologie et la Toxicologie, INERIS	Rapporteur
M. Claude LAMBRÉ, Directeur de recherche, INSERM (retraité)	Rapporteur
M. Philippe VERGER, Professeur, department of food safety zoonoses and foodborne diseases, WHO	Examinateur
M Rémy SLAMA, Directeur de recherche, U823, INSERM	Examinateur
M. Georges KASS, Professeur, Food Ingredients and Packaging, EFSA	Examinateur
Mme. Sophie DONNET, Chargé de recherche, Département mathématiques et informatique appliquées, INRA	Examinatrice
M. Stéphan CLÉMENÇON, Professeur, Département Traitement du Signal et des Images, TelecomParisTech	Directeur
Mme. Amélie CRÉPET, Chef de projet, Direction d'évaluation des risques, ANSES	Co-directrice

AgroParisTech
Unité Méthodologies et études relatives aux risques physico-chimiques (MER-PC)
Agence Nationale de Sécurité Sanitaire (ANSES)
27 avenue du Général Leclerc, 94700 Maisons-Alfort

Remerciements

En premier lieu, je tiens à remercier mes deux encadrants de thèse, Stéphan Clémençon et Amélie Crépet, qui m'ont accompagnée pendant ces trois années. Je remercie très chaleureusement Amélie, pour la confiance qu'elle m'a accordée en permettant mon recrutement à l'ANSES puis dans le montage de ce projet de thèse. Un grand merci aussi pour la qualité de son encadrement et son soutien continu, du début à la fin, dans les nombreux bons moments mais aussi dans les quelques moins bons. J'ai particulièrement apprécié nos nombreuses discussions, parfois animées, toujours enrichissantes, et suis très reconnaissante pour la diversité des connaissances et compétences que j'en ai tirées. Je remercie également Stéphan pour avoir permis cette thèse en acceptant d'en assurer l'encadrement et ensuite pour sa confiance, ses conseils et son soutien durant son déroulement.

Je remercie aussi chaleureusement Jean-Lou Dorne pour avoir permis mon passage à l'EFSA, pour son accueil remarquable et sa disponibilité mais surtout pour avoir fait de ce séjour, par le partage de ses connaissances et de ses idées, un moment très enrichissant.

Mes remerciements vont ensuite à Philippe Verger, Georges Kass, Rémy Slama et Sophie Donnet pour l'intérêt qu'ils ont porté à mon travail en acceptant d'être membre du jury. Je remercie plus particulièrement Frédéric Bois et Claude Lambré de me faire l'honneur d'être rapporteurs de cette thèse et ainsi d'apporter leur regard d'expert sur mon travail.

Je tiens également à remercier Jean-Charles Leblanc et Fanny Héraud pour la confiance qu'ils m'ont témoignée en me recrutant à l'ANSES, pour avoir permis le montage de cette thèse et surtout pour avoir continué à m'apporter une aide précieuse y compris après leur départ. Je n'oublie pas non plus l' « équipe PCB » et à sa tête, Mathilde Merlo, pour leur implication dans ce travail, plus particulièrement à son début. Merci aussi à Laurent Bodin pour son aide précieuse sur les modèles pharmacocinétiques. Un grand merci également à mes autres collègues de l'ANSES pour leur aide dans ce projet, Jean-Luc, Sébastien, Gilles et particulièrement à Véronique, Céline, Lynda, Nawel, Aurélie, Virginie, Marie, Guillaume, Alexandra, Fabrice, Carine, Karine, Sandrine, Sophie, Maëlle, Sabine, Josselin, Alexandre pour m'avoir si souvent supportée (dans tous les sens du terme) et pour les très bons moments partagés ensemble à l'ANSES, aux petits déjeuners, aux déjeuners, aux goûters et parfois même aux diners. Je remercie plus spécialement Jean-Cédric et Marion pour leur incroyable soutien durant ces années.

Je remercie également mes collègues de l'unité SCER et des autres unités de l'EFSA pour leur accueil chaleureux et tout ce que j'ai appris à leur contact. Je remercie particulièrement Tobin

Robinson pour m'avoir accueillie dans son unité. Je n'oublie pas non plus les anciennes collègues retrouvées à Parme et mes collègues « trainees » pour tous ces bons moments passés ensemble en Italie.

Je souhaite également remercier chaleureusement Marco Zeilmaker pour avoir partagé si pédagogiquement ses connaissances en toxicologie, sa confiance infaillible dans notre travail et nos échanges très enrichissants. Je remercie également Jessica Tressou et Cyril Feidt pour leurs conseils avisés.

Enfin, force est de constater qu'il aurait été impossible de survivre à cette thèse sans la présence de ma famille et de mes amis. Je remercie donc mes parents, Magali, Grégoire, Rémi, Marianne, Noémie, Léopold, Gaspard et mes grands-parents pour leur confiance en mes projets mais surtout pour tous ces bons moments ensemble et marques d'affection qui ont été autant de soutiens indispensables. Un grand merci aussi à ceux avec qui j'ai partagé les angoisses et les joies de la condition de doctorant : Thomas, Benoit, Bastien, Manu, Félix, Anne-Cha, ceux avec qui j'ai partagé d'indispensables moments de détente et de réinvention du monde après le boulot : Benoit, Manon, Andres, Suzie, Max, Elise, Rémy, Marion, Fanny, Anne-Lo, Quentin, mes compagnons de cordées qui m'ont permis de m'évader: Sacha, Sylvain, Marc, Aurélie, Marie-Fanny, Claire, Emilie, François et al., sans oublier les expat' avec qui j'ai partagé de belles vacances: Axel, Charlène et FX.

Valorisations scientifiques

Articles scientifiques issus de la thèse

Publiés :

Béchaux, C., Zetlaoui, M., Tressou, J., Leblanc, J. C., Héraud, F., & Crépet, A. (2013). Identification of pesticide mixtures and connection between combined exposure and diet. *Food and Chemical Toxicology*, *59*, 191-198.

Béchaux, C., Crépet, A., & Clémençon, S. (2014a). Improving Dietary Exposure Models by Imputing Biomonitoring Data through ABC Methods. *The international journal of biostatistics*. In press.

Béchaux, C., Zeilmaker, M., Merlo, M., Bokkers, B., & Crépet, A. (2014b). An integrative risk assessment approach for persistent chemicals: a case study on dioxins, furans and dioxin-like PCBs in France. *Regulatory Toxicology and Pharmacology*. 70, 261-269

Béchaux, C., Bodin, L., Clémençon, S., & Crépet, A. (2014c). PBPK and population modelling to interpret urine cadmium concentrations of the French population. *Toxicology and applied pharmacology, 279, 364–372.*

En préparation :

Béchaux, C., Amzal, B., Quinot, N., Crépet, A., & Dorne JLM. Human variability in polymorphic metabolism and related uncertainty factors: focus on inter-ethnic and inter-phenotype differences in CYP2D6 metabolism.

Articles scientifiques en relation avec le sujet de thèse

Publiés :

Crépet, A., Tressou, J., Graillot, V., **Béchaux, C**., Pierlot, S., Héraud, F., & Leblanc, J. C. (2013). Identification of the main pesticide residue mixtures to which the French population is exposed. *Environmental research*, *126*, 125-133.

Crépet, A., Héraud, F., **Béchaux, C.,** Gouze, M. E., Pierlot, S., Fastier, A., J.Ch. Leblanc, L. Le Hégarat, N. Takakura, V. Fessard, J. Tressou, R. Maximilien, G. de Sousa, A. Nawaz, N. Zucchini-Pascal, R. Rahmani, M. Audebert, V. Graillot, & Cravedi, J. P. (2013). The PERICLES research program: An integrated approach to characterize the combined effects of mixtures of pesticide residues to which the French population is exposed. *Toxicology*, *313*(2), 83-93.

Rapport scientifique

Participation à l'écriture du rapport EFSA: Modern methodologies and tools for human hazard assessment of chemicals. *EFSA Journal* 2014;12(4):3638 [87 pp.]. doi:10.2903/j.efsa.2014.3638

Résumés

Béchaux, C., Merlo, M., Bokkers, B., Zeilmaker, M., & Crépet, A. (2012). Risk assessment for dioxins in France based on a dynamic modeling of exposure. *Toxicology Letters*, *211*, S123-S124.

Béchaux, C., Crépet, A., & Clémençon, S. (2013, July). Approximate Bayesian Computation to Improve Dynamical Modelling of Dietary Exposure. In *of the 29-th European Meeting of Statisticians* (p. 44).

Béchaux, C., & Dorne, J. L. (2014). Mechanistic approaches for human and ecological risk assessment of chemical mixtures. *Toxicology Letters*, *229*, S18.

Amzal, B., Quignot, N., Dorne, J. L., Bois, F., & **Béchaux, C.** (2014). New tools for evidence-based risk assessment of chemical mixtures. *Toxicology Letters*,*229*, S19.

Communications scientifiques

Communications orales :

Béchaux, C., Crépet, A., & Clémençon, S. (2013, July). Approximate Bayesian Computation to Improve Dynamical Modelling of Dietary Exposure. In *of the 29-th European Meeting of Statisticians* (p. 44).

Béchaux, C., & Dorne, J. L. (2014). Mechanistic approaches for human and ecological risk assessment of chemical mixtures. *Toxicology Letters*, *229*, S18.

Posters :

Béchaux, C., Merlo, M., Bokkers, B., Zeilmaker, M., & Crépet, A. (2012). Risk assessment for dioxins in France based on a dynamic modeling of exposure. *Toxicology Letters*, *211*, S123-S124. ECETOC - Young Scientist Awards

Béchaux C, Merlo M, Leblanc JC, Heraud F, Favrot MC, Volatier JL, Crepet A (2011), Pharmacokinetic Modelling of PCB Body Burden. First Approach on the French Freshwater Fishermen Families. http://www.dioxin20xx.org/pdfs/2011/3718.pdf

Encadrement de stage

Co-encadrement de Rozenn Gazan, Stage de Master 2 (6 mois) pour l'obtention du diplôme d'ingénieur de l'institut supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage, spécialité Statistiques Appliquées.
Sujet : *Construction d'une typologie des consommations alimentaires de la population française adulte – Comparaison de deux méthode d'analyse factorielle*

Dynamic modeling of exposure with toxicokinetic integration to improve chemical risk assessment

Table of content

1 Introductory chapter 9

1.1. The risk assessment framework: toward an exposure-response continuum 9

1.1. Exposure, Toxicokinetic and Bayesian statistics for risk assessment: concepts and related issues 13

1.1.1. Exposure assessment 13

1.1.2. Toxicokinetic modeling 18

1.1.3. Bayesian approach and Approximate Bayesian computation 21

1.3 Framework of the thesis 26

2 Combining kinetic modeling, external exposure assessment and biomonitoring data to improve risk assessment of persistent chemicals 28

2.1 Improving dietary exposure models by imputing biomonitoring data through ABC methods 28

2.1.1 Summary of the objectives, methodological developments, main results and limitations of the work 28

2.1.2 Publication 30

2.2.1 Summary of the objectives, methodological developments, main results and limitations of the work 41

2.2.2 Publication 41

2.3.1 Summary of the objectives, methodological developments, main results and limitations of the work 51

2.3.2 Publication 52

3 Human variability in polymorphic metabolism and related uncertainty factors: focus on inter-ethnic and inter-phenotype differences in CYP2D6 metabolism. 61

4 Discussion 96

4.1 Main results and limits 96

4.1.1 Toxicokinetic modeling 97

4.1.2 Inter-individual variability 98

4.1.3 Modeling of lifelong exposure 101

4.2 Is it possible to extrapolate the proposed methods to mixtures? 103

4.2.1 Combined exposure 103

4.2.2 Toxicokinetic/dynamic interactions 112

4.2.3 Proposition of a case study to be investigated 113

5 References 115

6 Supplementary Material 123

6.1 ABC-MCMC algorithm and related functions written in R language used in the work “Improving dietary exposure models by imputing biomonitoring data through ABC methods” 123

6.2 ABC-SMC algorithm and related functions written in R language used in the work “PBPK and population modeling to interpret urine cadmium concentrations of the French population” 131

6.3 Supplementary material related to the work “Toward an integrative approach for cumulative risk assessment of persistent chemicals: a case study on dioxins, furans and dioxin-like PCBs in France” 138

6.4 Bayesian hierarchic model for meta-analysis in JAGS language 144

Résumé du manuscrit 146

1 Introductory chapter

1.1. The risk assessment framework: toward an exposure-response continuum

Humans are exposed to a wide range of natural and synthetic chemicals through their food, beverages and environment. Consequently a number of questions have been raised to determine if such exposure could potentiate adverse health effects and how identify, characterize and take action against this risk. In the recent decades, many organizations were born worldwide at different geographical and/or political scales to answer these questions (e.g. the US Environmental Protection Agency: US-EPA, the US Food and Drug Administration: FDA, the European Food Safety Authority: EFSA, the French Agency for food, environmental and occupational Health: ANSES). They aim to conduct risk analysis in order to ensure food and/or environmental safety and to protect public health, but also in a context of trade marked by globalization, to comply with international and national standards and market regulations. A formal framework for risk analysis was defined for the purpose of the Codex Alimentarius Commission which was created in 1963 by the Food Agriculture Organization (FAO) and the World Health Organization (WHO) to develop food standards, guidelines and related texts.

In this formal framework, risk analysis is defined as “a process consisting of three components: risk management, risk assessment, and risk communication" (WHO, 1995). Figure 1 illustrates the relationship between the three components of risk analysis. Risk management is defined as "The process, distinct from risk assessment, of weighing policy alternatives, in consultation with all interested parties, considering risk assessment and other factors relevant for the health protection of consumers and for the promotion of fair trade practices, and, if needed, selecting appropriate prevention and control options” (WHO, 1995). These options can be of several forms: introduction of maximum contaminant concentrations in certain foods (e.g. Maximal Residue Limit for pesticides), withdrawal of certain products from the market, recommendations in industrial process, consumer information and prevention, etc… Risk communication is defined as "The interactive exchange of information and opinions throughout the risk analysis process concerning hazards and risks, risk-related factors and risk perceptions, among risk assessors, risk managers, consumers, industry, the academic community and other interested parties, including the explanation of risk assessment findings and the basis of risk management decisions." (WHO, 1995). Risk assessment is the scientific evaluation of known or potential adverse health effects resulting from human exposure to dietary and/or environmental hazards.

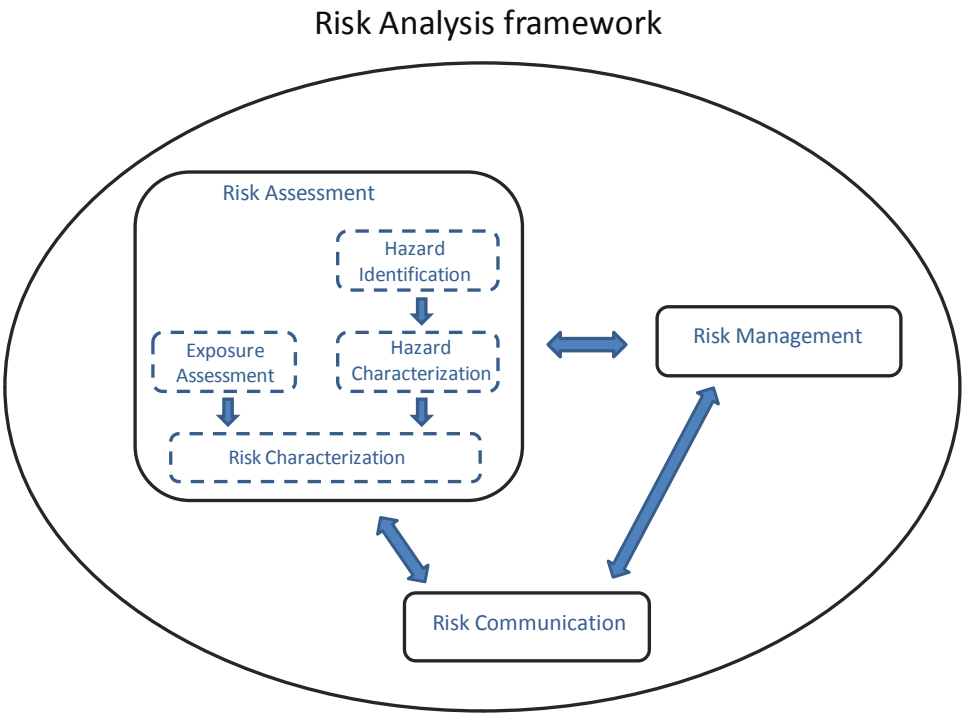


Figure 1: The components of risk analysis as defined by Codex Alimentarius Commission.

Traditionally, the risk assessment process has been divided into four sequential steps: hazard identification, hazard characterization, exposure assessment and risk characterization. Relationships between these four steps of traditional risk assessment are illustrated in Figure 1. The first step in the risk assessment procedure is thus to identify the substance of concern and address whether it would be a hazard to human health (adverse effects) and under what circumstances it might express its toxicity (target populations, conditions of exposure, etc…) (WHO, 1995). Hazard identification can be based on the results of different studies and data. Human data from epidemiological studies can provide the most relevant health effect related to the presence of the hazard in the environment as well as associated risk factors (diet, smoking etc …) and would avoid extrapolation from animals to man. However, such human data are rare and toxicological studies on animals (mainly mice, rat, rabbit and dog) thus play a critical role in hazard identification. Structure-activity relationship data based on the chemical properties can also provide useful insights into the prediction of the hazard for groups of structurally related chemicals. The second step of risk assessment involves the characterization of the dose-response relationship and the quantification of the incidence of adverse health effects after exposure to the agent. Similarly to the hazard identification step, the lack of human data does not allow direct establishment of the dose response and usually requires the use of dose response data from animal studies. However, with the increasing ethical concerns about animal experimentation, efforts was put into replacing the use of animals and in vitro systems from human

or animal tissues (cell culture, isolated tissues and organs) have been developed as alternative methods. For example, two research programs, TOX-21 in the US (Schmidt, 2009) and SEURAT-1 in Europe (Sholz et al., 2013), both deal with alternatives to animal testing such *as in vitro* methods and other integrated testing strategies (ITS) and have provided useful data. Exposure assessment is defined as the qualitative and/or quantitative evaluation of the degree of intake likely to occur. This third step aims to quantify the amount of an agent that has reached the individual (external dose) or has been absorbed by the individual (internal dose, absorbed dose). Three main approaches can be used to assess exposure: point of contact or personal assessment, where the exposure is measured from the contact point(s) of the body and both the concentration and time of exposure, are integrated; scenario evaluation where the concentration of the chemical is measured in the media (e.g. soil, water, food) and exposure is modeled by combining this data with data on people behavior (e.g. consumption habits, time spend indoor) and physiological data (e.g. weight, respiratory volume) using different exposure models; biological monitoring (or biomonitoring) which uses internal markers (biomarkers) of exposure (e.g. blood levels, urinary levels) to provide an estimate of the intakes and uptakes. Risk characterization is the last step of risk assessment which constitutes the synthesis of critically evaluated information and data from hazard identification, dose response assessment, exposure assessment into an estimation of the adverse effects likely to occur in a given population, including attendant uncertainties.

Recently, the concepts of Mode of Action (MoA) and Adverse Outcome Pathway (AOP) appeared in risk assessment (Boobis et al., 2006; Meek et al., 2013). The current WHO definition for MoA is “a biologically plausible sequence of key events leading to an observed effect supported by robust experimental observations and mechanistic data”. Therefore, MoA involves information about Toxicokinetics (TK) which describes the processes leading to the internal concentrations of a chemical or its metabolites(s) through knowledge of absorption, distribution, metabolism and excretion (ADME) and Toxicodynamics (TD) which describes the processes that lead to the toxic effects of a chemical or its metabolites(s) once it has reached the organ(s) or tissue(s). A number of modern *in vivo, in vitro* methodologies and *in silico* tools have been developed to investigate Mode of Action (MoA) at different levels of biological organization (organism, organ, cellular and molecular level). The concept of AOP first emerged from the field of ecotoxicology (Ankley et al., 2010) but AOPs are regarded by the WHO as equivalent to MoA in the human health context and an AOP has been defined as “a sequence of events from the exposure of an individual or population to a chemical substance through a final adverse (toxic) effect at the individual level (from a human health perspective) or population level (from an environmental perspective)” (Meek et al., 2013; OECD, 2013). In parallel to the MoA/AOP approaches, mathematical tools have also been developed in the

same spirit. Indeed, biologically based dose-response (BBDR) models were described by Shuey et al. (1994) as the mechanistic description in mathematical form of the sequence of causal events which intervene between administered dose and adverse outcome. In practice, BBDR models are often divided into two submodels, one linking exposure with tissue dose (TK model) and one linking the tissue dose with toxic effect (TD model). Currently, a number of TK and/or TD models are available and vary in level of complexity from simple data-based non-compartmental models to physiologically-based toxicokinetic (PBTK) or physiologically-based toxicokinetic/toxicodynamic (PBTK/TD) models.

All these methods provide the opportunity to move towards a more integrated approach of risk assessment. Indeed, while the four steps of risk assessment are still relevant, in a modern approach, they should not be regarded as separated steps anymore but considered together in an exposure-response continuum. Figure 2 describes the several steps between exposure and response and the corresponding models. Considering such a continuum allows a more mechanistic understanding of the impact of a chemical on human health. Such approach can also lead to combine several types of data which are often considered independently in a traditional approach, and can constitute a true strengthening for risk assessment. Focusing on exposure, the assessment is usually made either from external exposure by combining contamination levels and population behaviors, or from internal exposure that can be provided by biomonitoring data, whereas both are of great importance in the risk assessment process. In order to consider a true continuum, it is also suggested to take into account variability and uncertainty throughout the continuum. Consequently variability in external exposure but also in toxicokinetic and toxicodynamic should be taken into account with all the data that can inform it.

The work presented in this thesis is in line with the exposure-response continuum. The objective of the work is to propose methods that allow to connect external exposure with internal exposure integrating toxicokinetic of the chemical(s), intra and inter individual variability. This aims at using all the available data for exposure assessment in order to improve risk assessment related to several chemicals and/or to exposure from different sources.

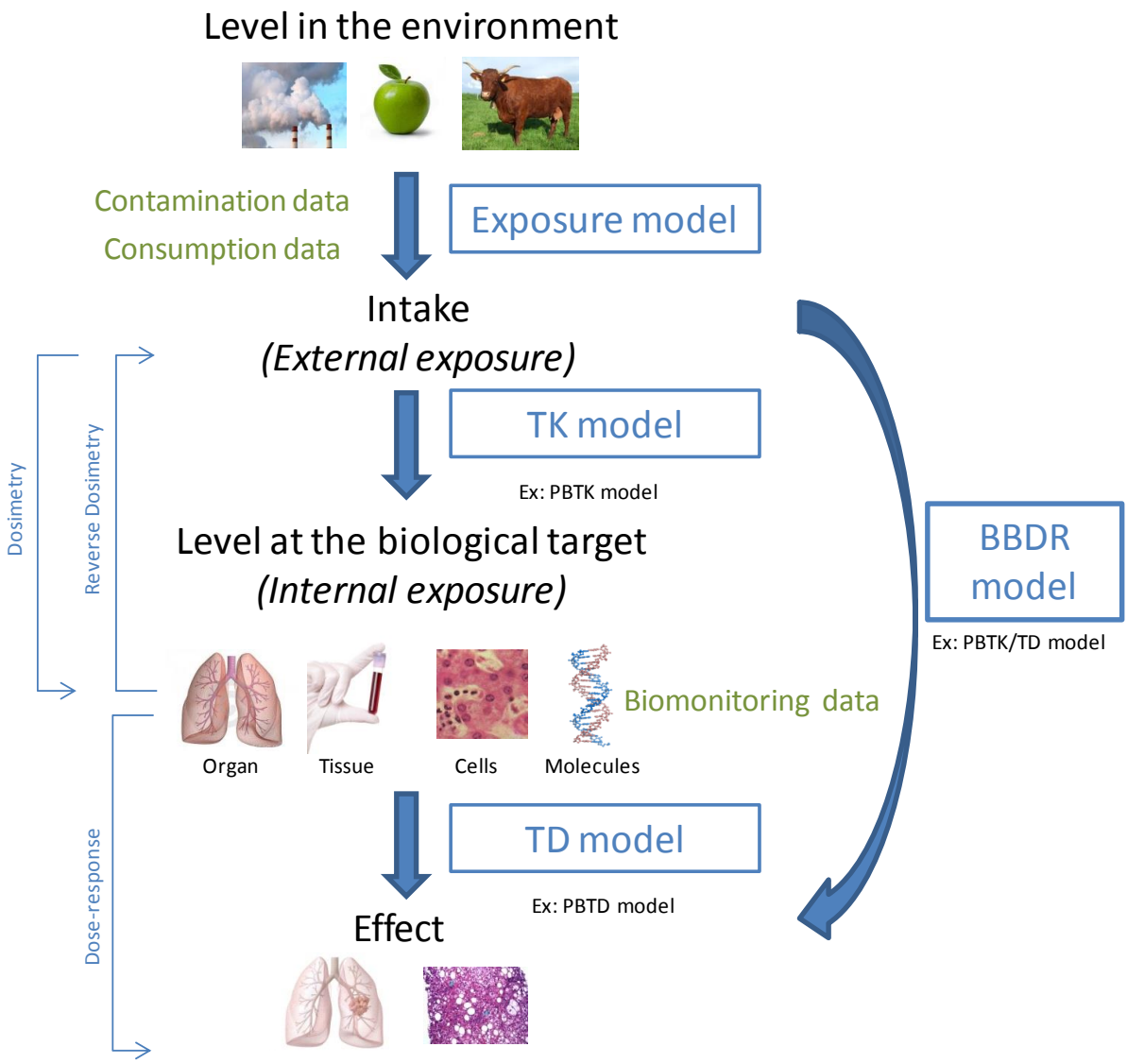


Figure 2 : Schematic representation of events leading from chemical level in the environment to a specific effect and associated models.

1.1. Exposure, Toxicokinetic and Bayesian statistics for risk assessment: concepts and related issues

1.1.1. Exposure assessment

Exposure assessment aims to quantify the amount of an agent that reaches the individual. The European community commission defined exposure assessment as "the determination of the emission pathways and rates of movement of a substance and its transformation or degradation, to

estimate the concentrations/doses to which human populations or environmental spheres (water, soil, air) are or may be exposed" (EC, 1996).

1.1.1.1.External exposure assessment

Different approaches are available to estimate chemical exposure in humans. In an ideal case, the personal exposure of the individual is quantified from the contact point of the body (skin, mouth, nostrils) and both concentration and time of exposure are measured and integrated. For example, a badge for radiation exposure or skin patches can be used. Such exposure measurement is not used in most cases. Therefore, exposure is most of the time assessed from indirect data and mathematical model. In this case, exposure models combining data on the concentration of the chemical measured in the media (e.g. soil, water, food…) and data on individual behavior (e.g. food consumption, time spend in an area…) are used. These models allow for an assessment of the external exposure which refers to a measure of environmental chemical concentration in contact with the body (Lyons et al., 2008).

Several adjectives can specify the kind of exposure assessment that is conducted (Crépet et al., 2012). Depending on the chemical, different pathways of exposure can be considered. When more than one pathway or more than one source of exposure is considered, it is referred as aggregated exposure. Regarding adults, the most common pathways are exposure via inhalation which may result from outdoor or indoor air contamination and from house dust contamination and oral exposure which is mainly determined by contamination of food and drinking water. Exposure resulting from a skin contact can also be considered in some cases. Regarding children, ingestion of dust and soil due to hand-to-mouth activities may also contribute to the exposure for some chemicals. When the simultaneous exposure to more than one chemical is considered, from the same or from different pathway(s) of exposure, it is referred as combined exposure or co-exposure (WHO, 2009; Meek et al. 2011) and the associated chemicals form a mixture. When several chemicals are shown to share a similar Mode/Mechanism of Action (MOA), they can be placed in cumulative assessment groups (CAG). Exposure to these chemicals belonging to a same CAG is thus referred as cumulative exposure (Boobis et al., 2008). Some examples of cumulative risk assessment are: Dioxins, furans and dioxin-like polychlorinated biphenyls (Van der Berg et al., 2005), organophosphates (EPA, 2006a) or triazines pesticides (EPA, 2006b).

In order to assess external exposure, different approaches can be used which range from semi-quantitative to probabilistic exposure assessment. With few data, semi-quantitative estimates of exposure can be implemented using theoretical exposure scenario and for example volume of

production, usage of the chemicals, etc. This exposure assessment is often viewed as a screening tool in order to rank/prioritize mixtures to be considered in a more refined assessment (WHO, 2009). When available, deterministic point estimates for both the behavior of the population (e.g. food consumption) and chemical levels in the media (e.g. food) can be combined in generic exposure scenario. Focusing on dietary exposure for example, information about the consumption of individuals in a certain population is generally provided by the results of surveys, where data related to the dietary behavior of a large sample of the population of interest over short periods of time are collected. For example, the Comprehensive European Food Consumption Database gathers a total of 32 different dietary surveys carried out in 22 countries (Efsa, 2011). Occurrence for the chemicals is estimated from measurements (monitoring data/experimental studies) or from reported usages levels. When data are sufficient to estimate distributions, a full probabilistic exposure assessment can be conducted by combining distributions of chemical occurrence and individual behavior. This assessment can also be refined according to other information available like a spatial temporal dimension (for example, a seasonal variability associated with some consumption behavior, or differences in levels of chemicals in food present according to the region).

Related issues and improvements:

Even when assessed from a high tier approach (i.e. full probabilistic approach), external exposure assessment still present some limitations. First, as external exposure assessment very often do not take the absorption into account; it does not represent the dose that is really taken up from the environment. Considering a 100% absorption rate for every route of exposure may conduct to a high over estimation of the exposure. Second, external exposure does not take the kinetic of the chemical in the body into account. Consequently, the potential accumulation in the body due to successive external exposures is not considered. Moreover, for chemicals with very long half-lives, external exposures which occurred years ago still participate to current individual burden. Even if it is very rarely considered and known, the evolution of the past exposure is of interest for the risk assessment of these chemicals. To be able to take these points into account in the exposure assessment, there is a need to move from a static to a dynamical modeling of the exposure. Bertail et al. (2008) proposed a dynamic modeling of the exposure that precisely takes the accumulation into account: the Kinetic Dietary Exposure Model (KDEM). External exposure (only dietary intake in this case) was modeled by a marked point process that is determined by the value of the intake and the time between two successive intakes. The KDEM approach was used to assess exposure to dietary methylmercury in the French adult female population (Verger et al., 2007). However, even if it

constitutes a new and dynamical way to model exposure, it still presents several limitations. First, the model was only proposed for a population and therefore do not account for inter and intra individuals variability. Second, the modeling does not make it possible to take a potential trend of exposure into account. Third, the toxicokinetic is described from a single compartment first order toxicokinetic model which is considered as sufficient to describe kinetic of some chemicals (for example chemicals that accumulate mainly in lipids like dioxins or methylmercury) but not for all chemicals. Finally, the model was only built for dietary exposure to a single chemical and thus need to be modified when considering aggregated and/or combined exposure. Moreover, KDEM outputs have never been compared to measured body burdens provided by biomonitoring data in order to validate or inform the model, whereas most of this information is included in the body burden or internal dose that can be informed by biomonitoring data.

Therefore, one of the objectives of this work is to propose a dynamic modeling of exposure that integrates toxicokinetic of the chemical(s), intra and inter individual variability, potential evolution of the exposure through the individual life and which could integrate information from biomonitoring studies.

1.1.1.2. Biomonitoring data and interpretation of internal exposure

Biomonitoring is defined as a method for assessing human exposure to chemicals by measuring the chemicals or their metabolites in human tissues or specimens, such as blood or urine (NRC, 2006). Biomonitoring data thus refers to analytical measurements of biomarkers in these specified body tissues or products (Albertini et al., 2006) where biomarkers can be any substance, structure or process that indicates an exposure or a susceptibility to disease. Consequently, two different types of biomonitoring data can be defined: biomarker of internal exposure (e.g. pg/g fat in blood, ug/g creatinine in urine) and biomarker of biological effect (e.g. AChE inhibition). This part is focused on biomarker of internal exposure. Over recent years, advances in analytic methods have allowed the measure of more chemicals, in smaller concentrations, using smaller samples of blood or urine. As a result, biomonitoring has become more widely used as a useful tool for public health research and risk assessment. For example the USA implemented the National Health and Nutrition Examination Survey NHANES (Centers for disease Control & Prevention, 2000), Germany the German Environmental Survey (Schulz et al., 2007) and France the French National survey on nutrition and health (ENNS, Falqa et al., 2011).

As highlighted in Albertini et al. (2006), biomonitoring data has many advantages for exposure assessment. First, it represents an integration of exposure from all sources and routes, which provides important information for risk assessment in case of aggregated exposure. Second, it takes into account the accumulation of the chemical in the body due to successive or long external exposure. For example, body burden is considered as a very reliable way to assess the risk related to chemicals known to be eliminated slowly from the body (Van Leeuwen and Younes, 2000). Third, a potential collection of serial biomonitoring samples over an extended period of time can provide information regarding variability and trends in exposure. Fourth, it represents direct measurements of the dose of the chemical substance that is really taken up from the environment i.e. internal dose and therefore includes the bioavailability (Angerer et al. 2007).

Although the availability of human biomonitoring data for many chemicals can greatly improve the risk assessment of these chemicals, their interpretation often remains their main weakness. Some methods have been proposed for the interpretation of Biomonitoring Data in the context of risk assessment. First, a simple description of the data can be achieved. This may suffice to highlight some groups of individuals with higher levels than others (e.g. smoker/non-smoker, male/female, children/adults/elderly, people living closed to a contaminated area). Then the explanations for these differences may be sought: higher exposure? Differences in kinetic? A comparison of successive biomonitoring data can also inform on a potential trend in exposure (decreasing or increasing) and on the effectiveness of public health efforts to reduce exposure. Second, in order to evaluate the risk associated with the amount of chemical in the body, biomonitoring data can be related to measures of toxicity. When an internal value associated with toxicity is available, a direct comparison can provide the risk which is thus the easiest case. However, internal critical values are rarely available since traditionally risk is rather made from external exposure. Starting from an external value associated with toxicity, biomonitoring equivalents (BEs) can be used to interpret internal exposure. BEs use toxicokinetic modeling to extrapolate the concentration of a chemical or its metabolite in a biological medium which corresponds to an existing exposure guidance value such as a tolerable daily intake, minimal risk levels, or reference dose. This extrapolation can also be referred as a dosimetry approach. Epidemiologic data can also be used to link some measured amount of chemical in the body with adverse effects. Finally, biomonitoring data can be related to external exposure in order to explain the amount of chemical in the body. The reverse dosimetry approach uses toxicokinetic modeling to estimate the intake dose or external environmental concentration based on a measured tissue concentration (Mumtaz et al. 2012). Establishing this relationship involves the reconstruction of past external exposure or dose, from biomonitoring data collected at some earlier time (Lyons et al., 2008).

Related issues and improvements:

In term of risk, biomonitoring data can be interpreted directly from an internal critical value when available or by deriving an internal critical value from an external one using toxicokinetic modeling. The main weakness of biomonitoring data thus lies in the inability to identify the different sources of exposure, their relative contribution to the total exposure or to interpret it in terms of individual behaviors. Consequently, biomonitoring data needs to be linked with external exposure for interpretation. Some reverse dosimetry approaches have already been proposed to investigate the external exposure that could explain an internal exposure. However, they do not integrate information from the external exposure previously assessed. Indeed, usually, external exposure and internal exposure are considered separately, whereas they both represent one event along the exposure-response continuum. Moreover, most of the time, the reverse-dosimetry approaches are limited to a single source of exposure.

Therefore, one of the objectives of this work is to propose a method which makes it possible to combine all the information available to assess the exposure and to link external and internal exposure. Very often individual biomonitoring data is available and sometimes information about the individual behavior (e.g. consumption habits, smoking status) and anthropometric measure (e.g. weight) are recorded. Such data can allow to assess an individual external exposure when combined with contamination data. One objective of this work would be also to investigate, in this case, whether it is possible to model the link between external and internal exposure integrating individual information at both level.

1.1.2. Toxicokinetic modeling

Toxicokinetic (TK) models are defined as “mathematical descriptions simulating the relationship between external exposure level and chemical concentration in biological matrices over time” (WHO, 2010). TK models take into account the Absorption, Distribution in the body, Metabolism and Excretion (ADME) of the chemical and its metabolites.

TK models vary in level of complexity and can be classified into two main categories: data-based (non-compartmental or compartmental) models and Physiologically-based Toxicokinetic (PBTK) models (US EPA, 2014). Data-based models (also called classical models) mathematically describe the temporal change in chemical concentration in chosen tissue of the body. The purpose

of these models is the description of the data and they do not consider *a priori* information on the physiological characteristics of the organism or the biological attributes of the response. Consequently, body is often considered as a single homogenous or multi-compartment system with elimination occurring in a specific compartment and the characteristics of the compartments (number, volume, etc.) are hypothetical. A classic example of these models is the single compartment first order TK model which describes the evolution of the concentration of the chemical in the considered compartment (e.g. lipids, whole body) through time using a first order differential equation. This simple model is entirely determined by the elimination rate (or the half-life), the fraction of absorption and the volume of distribution. These parameters can be easily assessed by combining different data. For example, Aylward et al. (2013) estimated apparent elimination half-life for dioxins congeners based on repeated blood sampling collected from former chlorophenol workers. The single compartment first order TK model is often used assuming the steady state and then a simple equation links external exposure with internal exposure (Lorber 2007; Sirot et al., 2008; Fromme et al. 2009).

The integration of physiological parameters into TK models results in a toxicologically based-TK model which is defined as "a model that estimates the dose to target tissue by taking into account the rate of absorption into the body, distribution and storage in tissues, metabolism and excretion on the basis of interplay among critical physiological, physicochemical and biochemical determinants" (WHO, 2010). PB-TK models are based on a compartmental approach that separates the organism's body into a series of biologically relevant anatomical compartments of defined volumes. The number of compartments varies from model to model depending on data quality, tissue compartments of interest (i.e. site of pharmacological or toxicological activity), purpose of the model, and the physico-chemical properties and behavior of the chemical in the organism (lipophilic/hydrophilic). All compartments are in general connected in anatomical order based on the blood circulatory system to form an integrated model. The transfer of chemicals between compartments is thus governed by blood flow rates and tissue solubility or partition coefficients (EFSA, 2014).

These models have been first developed for pharmaceuticals (PB-PK) and for hazard assessment of industrial chemicals (solvents) but they are increasingly developed in food safety for substances with well-known TK which include regulated substances (mostly pesticides (organophosphates, fungicides…), food contact materials (bisphenol A) and contaminants (persistent organic pollutants), acrylamide, heavy metals (cadmium, lead, mercury…), metalloids (arsenic). They have been mostly used in the context of human hazard assessment due to their ability to conduct route-to-route, high-to-low dose or interspecies extrapolation as made by Doerge et al. (2010) and Fisher et al. (2011) for BPA. They are also a very accurate tool to investigate and predict human

variability in TK. Bois et al. (2010) reviewed the prediction and sources of inter-individual variability in ADME parameters and highlighted three main contributing sources to the total variability: the variation across a population of 'normal' individuals at the same age, e.g. young adults; the variation across the population resulting from their different ages, e.g. infants or the elderly; and the variation resulting from the existence of subpopulations that differ in some way from the 'normal' population, e.g. due to genetic polymorphisms. Based on PBTK modeling, Valcke and Krishnan (2014) derived chemical specific adjustment factors (CSAFs) for a number of chemicals in specific subgroups of the population taking into account variability in physiological parameters and TK processes. PBTK models are also particularly suited to explore and model interaction in case of chemical mixtures (WHO, 2010). For example, Sasso et al. (2010) developed a generalised PB-TK model for mixtures using a chemical independent approach based on modules approach that can be directly 'mapped' to individual TK models for specific chemicals, while maintaining physiological consistency across different chemicals. Their model was applied to the mixture: methylmercury, cadmium, lead, arsenic (and metabolites), toluene, and benzene. In a reverse way, as mentioned previously (part…), PB-TK models are also used for reverse-dosimetry to estimate the external dose or exposure concentration needed to achieve given target organ concentrations, measured for example using biomonitoring data (Bernillon and Bois, 2001; Lyons et al., 2008; Verner et al., 2009).

Related issues and improvements:

TK is the link to connect external exposure to internal exposure in one way (dosimetry) or another (reverse-dosimetry) and therefore is an essential tool to consider exposure as a continuum and to take all the available information and data into account. Many TK models are now available and vary in level of complexity. The simplest ones (e.g. one compartmental single order models) have the advantage to depend on few parameters and thus to be easy to develop and implement. The more complex ones (e.g. PBTK models) propose a more accurate understanding of the mechanisms that explains differences between individuals, species, and chemicals and thus a better integration or inter and intra individual variability. They also allow to focus on a specific organ or tissue that is of particular interest for risk assessment (e.g. kidneys for cadmium) and to combine several exposure pathways.

Thanks to TK models, it is now possible to integrate the kinetic of the chemical and to link external exposure with internal dose. However, considering the kinetic leads to move the external exposure assessment from a static to a dynamic perspective. Indeed, internal dose depends on the

succession of exposures or the duration of exposure and not on a single exposure anymore, especially for chemicals with long half-life. Consequently, these models should be used with scenario of exposure on a period of time that varies with the bioaccumulation properties of the chemical of interest (e.g. one day, one month, lifelong exposure). Since TK modeling makes it possible to take intra and inter individual variability into account, it becomes essential to include it also in the dynamical exposure modeling. For example, the external exposure is very often shown as varying according to the age which is of importance when modeling a lifelong exposure. To consider an external exposure on a long period of time leads also to investigate the evolution of exposure in an area of interest. Therefore, there is a need to develop dynamical model for exposure assessment on long period of time that can integrate variability and match to kinetic modeling.

1.1.3. Bayesian approach and Approximate Bayesian computation

Considering a statistical model where x are the observed data, θ the parameter (or a vector of parameters) and $x \sim p(x|\theta)$ the sampling distribution (i.e. the probability distribution of the observed data conditional on its parameters), the Bayesian approach can be described as follow. θ is unknown (by nature) and considered as described by a prior probability distribution $p(\theta)$ which is the distribution of θ before any data is observed. The objective of the Bayesian inferences is to update the *a priori* distribution $p(\theta)$ from the data x to determine an *a posteriori* probability distribution $\theta \sim p(\theta|x)$. Sometimes previous results or expert knowledge (Albert et al. 2012) allow to speculate that θ has a high probability to be in a specific region of space or conversely cannot take some values. In this case, the information can be included using an informative *a priori* distribution. When no information is available on θ, uninformative *a priori* distribution are used. The Bayesian approach can be formalized as follows.

The joint distribution of (x, θ) is calculated as:

$$p(x, \theta) = p(x|\theta)p(\theta)$$

Using the Bayes formula, it can be written:

$$p(x, \theta) = p(\theta|x)p(x)$$

Then, the posterior probability is derived as a combination of the prior probability $p(\theta)$ and the "likelihood function" $p(x|\theta)$:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Where $p(x)$ is the marginal distribution of x, independent from θ, which can be written as:

$$p(x) = \int_{\Theta} p(x|\theta)p(\theta)d\theta$$

And thus

$$p(\theta|x) \propto p(x|\theta)p(\theta) .$$

From a practical point of view, calculations needed to determine the posterior distribution are often very complex or unsolvable. Indeed, analytical resolution of the integral for the marginal distribution of x is very hard when the posterior distributions $p(\theta|x)$ are not in the same family as the prior probability distribution $p(\theta)$ (i.e. the prior is not a conjugate prior for the likelihood function) or in case of hierarchical models with multiple variables. Such models are very often used the context of risk assessment (see below). Consequently, most of the time, posterior distributions are approximated using numerical integration. Many algorithms based on Markov Chain Monte Carlo (MCMC) methods have been developed for this purpose. MCMC algorithms aim at sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a large number of steps is then used as a sample of the desired distribution. The two most commonly used algorithms are Hasting-Metropolis algorithm (Metropolis et al., 1953; Hastings, 1970) and Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990).

The methods described above depend on knowing the likelihood function. However, sometimes in complex probability models, the likelihood can be very hard to calculate explicitly and is either impossible or computationally prohibitive to obtain. In these cases, the use of MCMC is limited by the difficulty of computing the likelihood function (Csillery et al. 2003). Beaumont et al. (2002) proposed a method, particularly suited to these cases, which bypasses the exact evaluation of the likelihood function to estimate the posterior distribution: the Approximate Bayesian Computation (ABC). The principle of standard ABC relies on the basic rejection algorithm: a candidate value $\theta^*$is generated from a prior distribution $p(\theta)$. Based on the numerical model defined by θ^*a simulated dataset x^* is generated. It is next reduced to a vector of summary statistics s^*. One then computes a "distance", $\rho(s^*, s)$, between the simulated and the observed summary statistics and

compares it to a tolerance level δ: if $\rho(s^*, s) \leq \delta$, $\theta^*$is accepted. After a chosen number of iterations, the accepted θ^*'s form the posterior distribution of θ.

Several algorithms have been proposed to compute ABC and they can be classified into three broad categories (Csillery et al. 2003). A first category includes the algorithms directly based on the basic rejection algorithm (described above) with some improvements like describing the relation between θ and s through different regressions (Beaumont et al., 2002; Blum et al., 2009). Some ABC-MCMC algorithms were also proposed (Marjoram et al. 2003; Wegmann et al. 2009) and form a second category. In these algorithms, a path of a Markov chain whose equilibrium distribution is the approximate posterior distribution is simulated. The parameter space is thus explored iteratively using $\rho(s^*, s)$ to update the current parameter values. It makes it possible to visit preferentially parameters values that produce simulation close to observed data and therefore to reduce the number of simulations required to estimate the posterior distribution. The last category relies on Sequential Monte Carlo methods (SMC) and algorithms are thus called ABC-SMC algorithms. They use large sample of proposed parameters values called "particles" which are propagated by sampling mechanisms or rejected according to $\rho(s^*, s)$. The main advantage of these algorithms relies on efficient sampling strategies that drive particles toward regions of high posterior probabilities and therefore allow a more efficient parameter space exploration (Sisson et al., 2007; Toni et al., 2009; Beaumont et al., 2009).

In order to compute ABC methods, several choices have to be made. First, the choice of relevant summary statistics and their dimension is paramount. Indeed, as highlighted in Csillery et al. (2003), carrying out inferences based on summary statistics instead of the full dataset inevitably implies discarding potentially useful information. Consequently, if the summary statistics are not sufficient, there is no guarantee that the posterior distribution based on the latter equals to the exact posterior distribution (i.e. based on the likelihood function) (Marjoram et al., 2003). On another hand, increasing the amount of information in summary statistics may reduce the accuracy of the procedure insofar as Beaumont et al. (2002) showed that the probability of accepting a θ decreases exponentially as dimensionality increases. The choice of the tolerance δ is also of importance and corresponds to a trade-off between the biais and the variance. Indeed, as highlighted in Beaumont et al. (2002), increasing δ reduces variance thanks to a larger sample size but also increases biais arising from θ that provide simulated data far from the observed data. In practice, the choice of δ is also a trade-off between computability and accuracy. Indeed, increasing δ decreased also the time of computation since a larger amount of θ would satisfy $\rho(s^*, s) \leq \delta$. The tolerance is most of the time set to be a quantile P_δ of the empirical distribution function of the distance between observed and simulated data (Beaumont et al., 2002; Blum et al., 2012; Wegmann

et al., 2009). For example, choosing $P_\delta = 0.05$ means that 5% of simulated x^* that are closest to observed x are kept. According to the chosen algorithm, other choices have to be made like transition kernel when implementing ABC-MCMC and ABC-SMC or number of particles and number of time the particles are perturbed when implementing ABC-SMC (Toni et al., 2009; Filipi et al., 2012). In term of practical implementation, some improvements have also been proposed, especially in Wegmann et al. (2009). For example, they proposed to run some preliminary iterations where values of proposed θ^* are randomly sampled from the prior distribution $p(\theta)$. This calibration step allows to define the tolerance as a quantile (as explained above). Moreover any of these simulations for which satisfy $\rho(s^*, s) \leq \delta$ can be used as a starting point for the chain. This removes the issue of the large number of iterations that are necessary to first satisfy this condition. These preliminary simulations can also be used to adjust the proposal range by choosing an appropriate transition kernel in order to ensure a thorough exploration of the whole parameter space.

Related issues and improvements:

Due to some specific features compared with frequentist approach, the Bayesian approach seems very well adapted to risk assessment context. First, Bayesian methods natively consider the uncertainty associated with the parameters of a probability model since a probability distribution is determined for each parameters rather than a single value. Second, Bayesian hierarchical modeling also provides a natural framework to take into account the various sources of variability present in a context of risk assessment. Indeed, in these models the probability distribution of the parameters can be interpreted in term of variability whereas probability distribution of the hyper-parameters can be interpreted in term of uncertainty (Pouillot et al., 2003). In a frequentist approach, this can be done using two-dimensional Monte Carlo simulations to account for uncertainty associated with the parameters of a probability model, whereas Bayesian methods handle it natively. Third, thanks to the prior distribution, Bayesian methods are often recommended as the proper way to make formal use of information from previous studies or expert opinion (Albert et al., 2012).

Focusing on TK modeling and link between external and internal exposure, Bayesian approach suits even better and have been proposed in many works to (re)estimate half-lives for single compartment TK models (Amzal et al., 2009, Albert et al., 2010), calibrate PBTK models (Loizou and Hoggs, 2011; Bois et al. 2010) or for reverse-dosimetry approach (Lyons et al. 2008, Bernillon and Bois 2010). Indeed, as addressed in Bois et al. (2010), there are many sources of inter-individual variability in the toxicokinetic of chemical that have to be taken into account. In order to model it

properly but also to improve our understanding of the determinants of population heterogeneity and differential susceptibility, hierarchical Bayesian model are very well suited. Moreover, very often biomonitoring data are available at the individual level whereas some parameters are available at an aggregated level (population or sub population). Therefore, when considering the link between biomonitoring data and external exposure, the hierarchical modeling is very accurate. In the specific context of dose reconstruction or reverse-dosimetry, such modeling has been proposed and is often called Bayesian population model (Gelman et al., 1996; Bernillon and Bois, 2000; Lyons et al. 2008). Another advantage of Bayesian approach in this context relies on the prior distribution. Indeed, a significant advantage of PBTK models is that the parameters have a physical or biological interpretation. Consequently, as highlighted in Lyons et al. (2008), the use of prior knowledge regarding possible ranges, central values and measure of dispersion as well as specific data from separate studies through informative prior distribution is very relevant.

In conclusion, many works already demonstrated the relevance of Bayesian approach for TK modeling; consequently this work will also be placed in a Bayesian context. However, as explained previously, one objective of the work is to link external exposure with internal exposure from biomonitoring data using a dynamical way of modeling exposure. Such a modeling would make it possible to take intra-individual variability (in addition to inter-individual variability) which can be related for example to changes in individual behaviors with age and/or environmental evolution of the exposure through time. Taking this variability into account in a dynamical modeling of exposure will lead us to consider individual trajectories of exposure that evolve in time according to intake values and inter-intake time as proposed in Verger et al. (2007) but also according to the date and the age, the intakes occur. As a consequence of this new dynamical modeling, the likelihood of such process becomes very complex and difficult to compute. Therefore, we propose to implement an ABC method as described above. ABC methods have been shown to be very suitable in case of inference problem with complex likelihood. However, they require to find “the good” summary statistics and distance and this choice can be very tricky since wrong statistics and/or associated distance can lean to fail to fit the model (Marjoram et al. 2003). Some statistics and distances have already been proposed in several previous works. However, they are closely linked to the particular inference question raised and the nature of the data and to the best of our knowledge, ABC methods were never implemented in the specific context of risk assessment or exposure/biomonitoring data. Thus, methodological developments are needed in order to implement ABC in a multi-level (or hierarchical) population modeling that would link external and internal exposure through TK modeling taken intra and inter-individual variability into account.

1.3 Framework of the thesis

The manuscript is divided into four parts. This first part presented the main available methods and the relative issues for exposure assessment (external exposure assessment and biomonitoring data) and toxicokinetic models. We also quickly presented the Bayesian approach and its advantages to answer methodological issues highlighted in risk assessment with a special focus on Approximate Bayesian Computation.

In a second part, we propose methods to combine kinetic modeling, external exposure modeling and biomonitoring data to improve risk assessment of persistent chemicals. This part is constituted by three different works. In a first work, an Approximate Bayesian Computation (ABC) method is developed to link the biomonitoring data with the external exposure taking the accumulation and the past exposure into account. A new dynamic modeling of the dietary exposure which takes the past exposure and its evolution into account is proposed. The method is applied to cumulative exposure to dioxins, furans and dioxin-like polychlorinated biphenyls (DL-PCBs). In a second work, the previous dynamic modeling of the dietary exposure is used but specific historic evolution of exposure and kinetics were considered for each congener (or group of congeners) in order to investigate cumulative risk assessment of persistent chemicals with very different half-lives. This work is also applied on dioxins, furans and DL-PCBs to study the current and future cumulative risk related to these substances. The risk for the general adult population is assessed from body burdens indicative for various types of toxicity through a probabilistic approach using distributions for BMDL and for predicted population body burdens. In a third work, an integrative approach is proposed to interpret the urine cadmium concentrations of the French population through dietary and cigarette smoke exposure. External exposure was assessed combining contamination data and individual data on consumption and smoking habits. A PBPK model is used to take the two sources of exposure into account and combined with the ABC approach to provide an accurate interpretation of the current biomonitoring data. Then, the related risk was assessed from external and internal critical values corresponding to kidney effects.

In a third part, we investigate human variability in TK resulting from the existence of subpopulations that differ in some way from the ‘normal’ population due to genetic polymorphisms. Indeed, conventionally, a default uncertainty factor is used to cover the human variability aspects. However, it has been suggested that regarding some polymorphic metabolic pathway, this factor would not provide sufficient protection (Dorne 2010, Martin et al, 2013). It is thus proposed to investigate the human variability in phase I enzyme from the example of CYP2D6. A review of

pharmacokinetic (PK) data available for substrates of this enzyme was conducted. A Bayesian hierarchical model is proposed to meta-analyse the data in order to be able to properly separate and quantify inter-study, inter-substrate and inter-individual variability. The predictability of human variability is also investigated from this example.

In a fourth and last part, we first discuss the strengths and limitations of the proposed work and how it could fit into a new risk assessment, especially in light of a better integration of human variability. Then, we propose to discuss how this work could be extended for mixtures and what are the limitations and the issues raised in this context. A statistical method is proposed to identify the main mixtures to which a population is exposed and to connect these combined exposures with the diet. The non-negative-matrix factorization (NMF) is used to extract the main consumption patterns from individual consumption data of the French population and the exposure to associated mixtures. Afterward, discussion proceeds on kinetics interactions and models to include it in risk assessment of chemical mixtures. Finally, a general discussion concludes this manuscript with.

2 Combining kinetic modeling, external exposure assessment and biomonitoring data to improve risk assessment of persistent chemicals

In this part, three different works investigate how external exposure and biomonitoring data can be combined using dynamic modeling of exposure and toxicokinetic models. Such works first aims at providing an accurate interpretation of biomonitoring data in term of exposure level from different sources and/or several chemicals. It also aims at improving the exposure assessment and thus the whole risk assessment of persistent chemicals. The developed methods are applied to cumulative exposure to dioxins, furans and DL-PCBs and aggregated exposure to cadmium. For both case studies, exposure and toxicokinetic models developed in this work are combined with toxicodynamic models provided by other studies in order to conduct a full risk assessment.

2.1 Improving dietary exposure models by imputing biomonitoring data through ABC methods

2.1.1 Summary of the objectives, methodological developments, main results and limitations of the work

The main objective of this work is to propose a method that integrates information provided by biomonitoring data in order to inform and improve exposure models.

Several methodological developments are proposed in this work. First, the Kinetic Dietary Exposure Model (KDEM) initially proposed by Bertail et al. (2008) was modified in order to take into account a potential historical variation in exposure through time. This was shown as a crucial issue when dealing with persistent chemicals. KDEM was also modified in order to better model the inter-individual and intra-individual variability trough the simulation of individual trajectories that take into account individual intake and half-life values and evolution of exposure and half-lives with the age. Secondly, in order to integrate information providing by biomonitoring data to infer on parameters of the dynamic model of exposure, it was proposed to implement Approximate Bayesian Computation (ABC) methods. Implementation of such method on the exposure model required to provide relevant summary statistics and distances for risk assessment. Given the complexity of the simulations, an efficient algorithm had to be found and an ABC-MCMC algorithm has been implemented in the R software (code of the algorithm is provided in SM). Moreover, a second stage

has been added to the algorithm in order to perform model selection simultaneously with the estimation of the parameters.

The promoted method was applied to measurements of blood concentration of dioxins in a group of French fishermen families. A cross validation showed that the model successfully fit to dioxins data. The outputs were an estimation of the body burden distribution from observed dietary intakes and the evolution of dietary exposure to dioxins in France between 1930 and today.

However, some limitation remains and this work can still be improved. Indeed, despite the implementation of a MCMC-algorithm, the time of computation remains long. Therefore, the investigation of other algorithm could be useful especially to deal with more complex modeling or a bigger population. The modeling of the kinetic i.e. a simple first order one compartment model was relevant for dioxins but could not fit for all chemical. Moreover, such modeling, do not make it possible to integrate anatomical nor physiological aspects. Consequently, look at more complex kinetic models like physiologically based models could be interesting. Regarding the application on dioxins, furans and DL-PCBs, it has been shown that some congeners have very different kinetics. In this work this difference in kinetic was not considered and thus integrate it in the model could improve the modeling. Finally, this work could be used to conduct a full risk assessment based on body burden.

2.1.2 Publication

DE GRUYTER The International Journal of Biostatistics 2014; aop

Research Article

Camille Béchaux*, Amélie Crépet, and Stéphan Clémençon

Improving Dietary Exposure Models by Imputing Biomonitoring Data through ABC Methods

Abstract: New data are available in the field of risk assessment: the biomonitoring data which is measurement of the chemical dose in a human tissue (e.g. blood or urine). These data are original because they represent direct measurements of the dose of chemical substances really taken up from the environment, whereas exposure is usually assessed from contamination levels of the different exposure media (e.g. food, air, water, etc.) and statistical models. However, considered alone, these data provide little help from the perspective of Public Health guidance. The objective of this paper is to propose a method to exploit the information provided by human biomonitoring in order to improve the modeling of exposure. This method is based on the Kinetic Dietary Exposure Model which takes into account the pharmacokinetic elimination and the accumulation phenomenon inside the human body. This model is corrected to account for any possible temporal evolution in exposure by adding a scaling function which describes this evolution. Approximate Bayesian Computation is used to fit this exposure model from the biomonitoring data available. Specific summary statistics and appropriate distances between simulated and observed statistical distributions are proposed and discussed in the light of risk assessment. The promoted method is then applied to measurements of blood concentration of dioxins in a group of French fishermen families. The outputs of the model are an estimation of the body burden distribution from observed dietary intakes and the evolution of dietary exposure to dioxins in France between 1930 and today. This model successfully fit to dioxins data can also be used with other biomonitoring data to improve the risk assessment to many other contaminants.

Keywords: chemical risk assessment; approximate Bayesian computation; biomonitoring data; kinetic model; exposure model

DOI 10.1515/ijb-2013-0062

1 Introduction

Many chemicals are known to accumulate in the environment and to be very slowly eliminated from the human body such as polychlorinated biphenyls, dioxins, organochlorine pesticides, chlorobenzenes, and polybrominated diphenyl ethers. They are of great concern for the public health and are listed by the Stockholm convention as *Persistent Organic Pollutants* [1]. To assess the risk related to these chemicals, the use of the body burden is now considered as the best way [2]. Recently, new data allow for a direct assessment of the body burden: the biomonitoring data. However, as highlighted in Lyons et al. [3], understanding the effects on public health of exposure to environmental chemicals requires establishing relationships among events along an exposure–health evaluation–risk assessment continuum. Therefore, biomonitoring data need to be linked to external exposure to be interpreted in terms of source of exposure

***Corresponding author: Camille Béchaux,** Anses – DER, 27 Avenue du général Leclerc, Maisons-Alfort 94700, France, E-mail: camille.bechaux@anses.fr; camille.bechaux@gmail.com
Amélie Crépet, Anses – DER, 27 Avenue du général Leclerc, Maisons-Alfort 94700, France, E-mail: amelie.crepet@anses.fr
Stéphan Clémençon, Telecom ParisTech – LTCI UMR, Paris, France, E-mail: stephan.clemencon@telecom-paristech.fr

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

and for risk management purposes. External exposure is usually assessed from statistical models considering current contamination levels of the different exposure media (e.g. food, air, water, etc.) and behaviors of individuals (consumption habits, activity pattern, etc.). However, these classical exposure models do not take the accumulation in the body into account. Some reverse-dosimetry or steady-state approaches aim at taking this elimination time into account through toxicokinetic models [4]. However, they do not account for the level of the past exposures and especially its evolution. While because of the long half-life of persistent chemicals, body burden measured today also reflects the past exposure of the individuals. It is well highlighted that exposure to environmental contaminants has significantly changed over the last decades due to changes in production, in use, and in successive regulations [5] but data on past exposure are rarely available. Therefore, to be able to interpret biomonitoring data when assessing the risk related to persistent chemical, a comprehensive method that takes lifelong exposure and accumulation into account is needed. It is precisely the objective of this paper to develop a method to link the biomonitoring data with the external exposure taking the accumulation and the past exposure into account.

Different statistical points are developed and combined in this paper to propose a comprehensive approach. First, we modified the *Kinetic Dietary Exposure Model* (KDEM) which was introduced in the context of dietary risk assessment by Bertail et al. [6]. This dynamic model takes into account the accumulation phenomenon of a given chemical in the body and a single-compartment pharmacokinetics model governing its elimination. A simple pharmacokinetics model is suitable for many environmental contaminants that accumulate mainly in lipids (e.g. MeHg, polychlorinated biphenyls, polybrominated diphenyl ethers, dioxins, etc.). The KDEM approach was used to assess exposure to dietary methylmercury (MeHg) in the French adult female population [7]. However, these results have never been validated with measured body burdens provided by biomonitoring data [8]. Moreover, the model does not take the trend of exposure into account. That is why, a modified model is developed which integrates a scaling function λ to describe the evolution of the exposure. Data on past contamination or past exposure are rarely available but it is proposed here to use biomonitoring data to estimate the parameters of this function. Given the very high complexity of the likelihood function of the new model and the computational difficulties faced when trying to optimize it, Approximate Bayesian Computation (ABC in short, Beaumont et al. [9]) is proposed to fit the numerical model from the biomonitoring data available. This approach bypasses the exact evaluation of the likelihood function to estimate the posterior distributions by means of simulations and well-chosen summary statistics. ABC has recently become popular for statistical inference in complex models such as those considered in biosciences, in population-genetics [10], and in mathematical epidemiology [11] for instance. In order to perform model selection simultaneously with the estimation of the parameter, an original two-stage hierarchical ABC-MCMC algorithm is implemented following in the footsteps of Toni et al. [12]. Specific summary statistics, especially relevant for risk assessment, are proposed based on Cramer–Von Mises statistic.

The paper is structured as follows: Section 2 outlines the key ingredients governing the modified KDEM which accounts for the possible temporal evolution in exposure. Section 3 specifies the principles of the ABC method used to fit the new dynamic model of exposure from punctual biomonitoring data. Relevant specific summary statistics for risk assessment are proposed, and the model selection for the parametric modeling of the function λ is described. Section 4 is devoted to the application of the methodology to blood concentration in dioxins measured in a population of French fishermen families.

2 Dynamic modeling of dietary exposure

The KDEM was introduced by Bertail et al. [6] in order to describe the temporal evolution of the total body burden of a chemical present in a variety of foods. Its dynamic is ruled by two components, a *marked point process* (MPP) which models dietary behavior and governs the accumulation of the chemical in the human body and a linear differential equation accounting for its physiological elimination.

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

The accumulation due to successive dietary intakes is described by the MPP $(T_n, U_n)_{n\in N}$, where the T_n's are the successive times when an intake of value U_n occurs (with $T_0 = 0$ by convention). The U_n's are independent from the T_n's and the inter-intake times $\Delta T_n = T_n - T_{n-1}$, $n \geq 1$.

The elimination process between two successive intakes is described by a *single-compartment* pharmacokinetic model. The total body burden of the chemical $x(t)$ decreases with time t according to the linear differential equation:

$$\frac{dx}{dt}(t) = -\gamma \times x(t), \tag{1}$$

where $\gamma > 0$ denotes the elimination rate. The parameter γ is related to the biological half-life of the chemical in the body $t_{1/2} = \log(2)/\gamma$, namely the time needed for the total body burden x to decrease by half in the absence of any additional intake.

In order to integrate the evolution of the exposure into the model, a scaling function of time $\lambda(t)$ is introduced in the KDEM model. This function describes the possible evolution of the dietary exposure through time and is assumed to be of a parametric form, fully determined by a vector θ. The shape of the exposure evolution is unknown. Therefore, the function λ must be represented in a flexible manner. That is why, a linear combination of functions from a B-spline basis of degree $m \geq 1$: $\lambda(t) = \sum_{i=0}^{m} \beta_i B_i^m(t)$ was proposed.

Finally, exposure is described by a piecewise deterministic process $X(t)$ of initial value x_0 and evolving according to eq. (1) between intake times. The total body burden values $X_n = X(T_n)$ at intake times T_n are defined through the recurrence equation:

$$\begin{cases} X_{n+1} = X_n e^{-\gamma \Delta T_{n+1}} + \lambda_\theta(T_{n+1}) U_{n+1} \\ X_0 \quad = x_0 \end{cases}. \tag{2}$$

3 An ABC method to fit the model to biomonitoring data

We propose to implement an ABC method [9] to fit the parameter θ of the scaling function λ by comparing the distribution of the measured body burdens from data to values simulated from KDEM.

The principle of standard ABC relies on the basic rejection algorithm: a candidate value θ^* is generated from a prior distribution $\Pi(\theta)$. Based on the numerical model defined by θ^*, a simulated data set X^* is generated. It is next reduced to a vector of summary statistics s^*. One then computes a "distance", $\rho(s^*, s)$, between the simulated and the observed summary statistics and compares it to a tolerance level δ: if $\rho(s^*, s) \leq \delta$, θ^* is accepted. After a chosen number of iterations, the accepted θ^* s form the posterior distribution of θ.

3.1 Specific summary statistics and distances for risk assessment

The choice of the vector of summary statistics s and the corresponding distance ρ is paramount in ABC methods. This aspect has been widely studied in the case of population genetic data [9, 13, 14]. Regarding biomonitoring data, specific summary statistics must be defined. Here, the summary statistics are built from distances between the observed and the simulated distributions of body burdens of different age classes of individuals. Individuals with ages in the same range, $[a_k, a_{k+1}]$, at time t are grouped, in order to form $K \geq 1$ (disjoint) groups of comparable sizes. Denote by $I_k \subset \{1, \ldots, n\}$ the corresponding subsets of indexes and then form the empirical distributions of the corresponding exposures: $F_k = (1/\#I_k) \sum_{i\in I_k} \delta_{x_i(t-\tau_i)}$, denoting by $\#I_k$ the cardinality of I_k, $1 \leq k \leq K$. The choice of the number K of age classes is critical since it determines the dimensionality of the vector of summary statistics

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

$s = (F_1, \ldots, F_K)$. Indeed, when the summary statistic is not sufficient, there is no guarantee that the posterior distribution based on the latter equals to the exact posterior distribution, i.e. based on the likelihood function) [15]. However, increasing the number K of summary statistics, thereby increasing the amount of information available may reduce the accuracy of the procedure insofar as the probability of accepting a value decreases exponentially as dimensionality increases [9].

For each age class k in $\{1, \ldots, K\}$, the empirical distribution $\tilde{F}_k(dx)$ of the observed body burdens, $\{\tilde{X}_i(t-\tau_i),\ i \in I_k\}$, is compared to the simulated body burdens $\{X_i^*(t-\tau_i),\ i \in I_k\}$ with the model defined by the parameter value θ^*, $F_k^*(dx)$. Various quantities can be considered to evaluate the discrepancy between two distributions [16]. A very popular measure of global discrepancy is the two-sample Cramer–Von Mises statistic which is widely used in nonparametric hypothesis testing [17]. In the present context, it corresponds to the quantity:

$$\rho_{VM}(\tilde{F}_k,\ F_k^*) = \frac{\#I_k}{2}\int_{-\infty}^{+\infty} \left(\tilde{F}_k(x) - F_k^*(x)\right)^2 F_k(dx), \tag{3}$$

where $\tilde{F}_k(x) = (1/\#I_k)\sum_{i\in I_k} I\{\tilde{X}_i(t-\tau_i) \le x\}$, $F_k^*(x) = (1/\#I_k)\sum_{i\in I_k} I\{X_i^*(t-\tau_i) \le x\}$, and $F_k = (F_k^* + \tilde{F}_k)/2$.

In the field of risk assessment, one generally focuses on the right tail of the exposure distribution. Thus, it may be relevant to consider the truncated version:

$$\tilde{\rho}_q(\tilde{F}_k,\ F_k^*) = \frac{\#I_k}{2}\int_{q}^{+\infty} \left(\tilde{F}_k(x) - F_k^*(x)\right)^2 F_k(dx), \tag{4}$$

choosing q as a specific quantile of the distribution $\tilde{F}_k$. We consider the following dissimilarity measure between the distribution of observed data $\tilde{X}$ and that of simulations X^*:

$$\rho(\tilde{X},\ X^*) = \sum_{k=1}^{K} \tilde{\rho}_q(\tilde{F}_k,\ F_k^*). \tag{5}$$

3.2 A two-stage hierarchical ABC-MCMC algorithm for model selection

Many algorithms can generate observations from a posterior distribution without computing likelihoods explicitly, including basic rejection algorithms. In order to reduce the number of simulations required to estimate the posterior distribution, an ABC-MCMC algorithm is used [15]. Thereby, a path of a Markov chain whose equilibrium distribution is the approximate posterior distribution $\Pi(\theta|\rho(\tilde{X}, X^*) \le \delta)$ is simulated. In this way, the values of the parameter θ providing simulations close to the observed data are visited preferentially.

Here, a model selection is performed simultaneously with the estimation of the parameter θ of the scaling function λ. A two-stage hierarchical algorithm is proposed to consider m as an additional parameter [12]. It makes it possible to choose the best form for the scaling function λ among a finite number of models, indexed by $m \in \{1, \ldots, M\}$ and sorted by increasing complexity. The two-stage hierarchical algorithm is based on an ABC-MCMC procedure and implemented as follows.

1. If now at m, propose a move to m^* according to a transition kernel $q(m \to m^*)$.
   (i) if now at θ_m^*, propose a move to $\theta_{m^*}^*$ according to a transition kernel $q'(\theta_{m^*} \to \theta_{m^*}^*)$
   (ii) Generate X^* from model m^* and $\theta_{m^*}^*$
   (iii) Compute $\alpha' = \min\left\{1, \frac{\Pi(\theta^*|m^*)q'\left(\theta_{m^*}^* \to \theta_{m^*}\right)}{\Pi(\theta_{m^*})q'\left(\theta_{m^*} \to \theta_{m^*}^*\right)} 1\left(\rho(X^*, \tilde{X}) \le \delta\right)\right\}$
   (iv) Accept $\theta_{m^*}^*$ with probability α' and otherwise stay at θ_{m^*}, then return to **1**.
2. Compute $\alpha = \min\left\{1, \frac{\Pi(m^*)q(m\to m^*)}{\Pi(m)q(m\to m^*)} 1\left(\rho(X^*, \tilde{X}) \le \delta\right)\right\}$
3. Accept m^* with probability α, stay at m otherwise and then return to **1**.

This algorithm was implemented in the software R, and the code is available in the supplementary material. It can be noticed that the higher the posterior probability, the greater the number of accepted θ. This

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

ensures a good estimate of the posterior distribution for the selected model. However, for the models which are poorly represented in the marginal posterior distribution of m, the small number of accepted θ does not provide an accurate estimation of the posterior distribution of θ[12]. It should be noticed that this model selection implicitly penalizes the models with the highest numbers of parameters since the probability for θ to be accepted decreases with the dimension.

4 Application to dioxin biomonitoring data in France

4.1 Data and exposure assessment

In the French National study on dioxin blood levels in French consumers of freshwater fish (ICAR study, Anses and InVS [18]) conducted in 2009, the dioxin concentrations in blood (pg/g fat) were recorded for 606 adults between 18 and 75 years of age. All individuals were from fishermen families. In order to compare these measurements with the estimated body burdens (pg/kg body weight), the concentrations were converted into total body burdens using the equation proposed by Deurenberg et al. [19]. This conversion assumes that the balance between all lipid compartments in the body and the fat content of the human body varies with age, sex, weight, and height. Consumed quantities of 115 food items covering the main diet of each individual taking part in the study were collected by means of a food frequency questionnaire and a photograph manual of portions' size.

Contamination data in food are provided by three French studies. Mean concentrations in edible parts of the 26 most consumed species of freshwater fish are provided by the ICAR study. Contamination data of seafood come from the CALIPSO study [20]. This study provides the mean contamination in 30 species of marine fish, 17 species of molluscs and crustaceans, and 14 other products. The mean concentration of other foods known to contribute to the exposure to dioxins (24 food items including the different kinds of meat, cooked meat, ready meals, dairy products, eggs, and fat such as butter and oil) comes from the Second French Total Diet Study [21].

The reference distribution of the dioxins intake U in 2009 in adulthood is estimated by combining the distribution of the reference contamination of foods with consumption provided by the food questionnaire of the ICAR study. Since the consumption is known to be subject to wide fluctuations during childhood, reference intakes U_n at age under 18 are simulated by combining the consumption distribution of 1,455 French children aged 3–17 years recorded in the INCA2 study [22] with the distribution of the reference contamination vector.

4.2 Modeling λ for dioxin exposure

The purpose of the scaling function λ is to correct the intake according to the date on which it occurred. It thus describes the historic evolution of the exposure over time. Dioxins are naturally found in the environment [23], but the emissions due to human activities are suspected of having significantly changed since the 1930s [24]. Thereby we set $\lambda(t \leq 1930) = 0.1$ the level of exposure related to the background contamination of the environment, as calculated in Van Der Molen et al. [24]. Since the data were collected in 2009, individual reference intake U corresponds to this date and the correction factor is equal to 1 in 2009, i.e. $\lambda(2009) = 1$. The scaling function λ is assumed to be a continuous function in the interval $[1930; 2009]$. Therefore, the linear combination of functions from the B-spline basis of degree $m \geq 1$ has only two knots (1930 and 2009 namely). In this situation, the number of parameters of the vector θ to be estimated corresponds to the number $m + 1$ of basis functions. In the present study, $M = 4$ models are tested, each model $m \in \{1, \ldots, , 4\}$ corresponding to a B-spline basis of degree m.

The evolution of environmental contamination by dioxins has already been studied in certain countries by Lorber [4], Van Der Molen et al. [24], and Aylward and Hays [25]. Their results were used to set

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

appropriate prior distributions for the parameters governing the different candidate model. Typical scaling functions λ obtained from these prior distributions for θ are drawn in red in Figure 2.

4.3 ABC implementation

Summary statistics of Section 3 were calculated using $K = 5$ age classes. This number of classes guarantees a sufficient number of individuals in each class (Table 1). Observed body burdens and their distributions by age classes are depicted in Figure 1. In risk assessment, a focus is made on the most exposed individuals. Therefore, the threshold q of the distance ρ (eq. (4)) is set to 0.5. Thereby, the 50% of the most exposed individuals of each age class are used to fit the model.

Table 1 Comparison of observed and simulated body burden (mean and 95% credibility interval) distributions for each age class

Age (years)			Observed and predicted body burden (pg\kg bw)		
	Size		25th p	Mean	97.5th p
18–37	124	Obs.	1,022	4,097	8,695
		Sim.	959 [854;1,152]	4,501 [4,018;5,387]	9,072 [8,463;11,615]
38–46	123	Obs.	3,253	6,358	13,881
		Sim.	1,571 [1,396;1,905]	6231 [5515;7567]	15,408 [13,662;18,685]
47–54	130	Obs.	4,254	11,709	22,327
		Sim.	1,388 [1,220;1,697]	11,397 [9,998;13,951]	25,881 [21,048;32,601]
55–61	112	Obs.	3,539	14,916	30,932
		Sim.	1,806 [2,076;2,550]	21,760 [18,945;26,729]	32,742 [26,084;45,422]
62–75	117	Obs.	9,958	20,213	43,491
		Sim.	4,686 [4,038;5,814]	22,092 [19,110;27,325]	56,144 [42,070;68,530]

The choice of tolerance δ corresponds to a trade-off between the bias and the variance of the estimate [9] and, in practice, between computability and accuracy. The tolerance is often set to be a quantile P_δ of the empirical distribution function of the distance between observed and simulated data [9, 11, 26]. In this work, $P_\delta = 0.05$ which means that the 5% of simulated X that are closest to $\tilde{X}$ considering all the models together are kept.

As proposed in Wegmann et al. [26], a series of 10,000 simulations have been preliminarily performed, where values for the parameter θ are each time randomly drawn from the prior distribution $\Pi(d\theta)$. This calibration step allows to conveniently define the tolerance δ such that $P_\delta = 0.05$. Moreover any of these simulations for which the condition $\rho(\tilde{X}, X^*) \leq \delta$ is true can be used as a starting point for the chain. This removes the issue of the large number of iterations that are necessary to first satisfy the condition $\rho(\tilde{X}, X^*) \leq \delta$. These preliminary simulations are also used to adjust the proposal range by choosing an appropriate transition kernel in order to ensure a thorough exploration of the whole parameter space.

The θ which satisfy $\rho(\tilde{X}, X^*) \leq \delta$ are used as a starting point for the chains of the ABC-MCMC algorithm. Thus, there is no initial burn-in period. The convergence of the chains is checked by testing the goodness-of-fit between successive intermediate posterior distributions of θ for each parameter, using a two-sample Kolmogorov–Smirnov test [27]. This test indicates that the chains have converged after 150,000 iterations. The acceptance rate is close to 38% which ensure a thorough exploration of the parameter space. The posterior distributions of each parameter are built from the accepted θ of the 150,000 iterations of two different chains. The function λ obtained from these posterior distributions for the four models are shown in black in Figure 2.

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

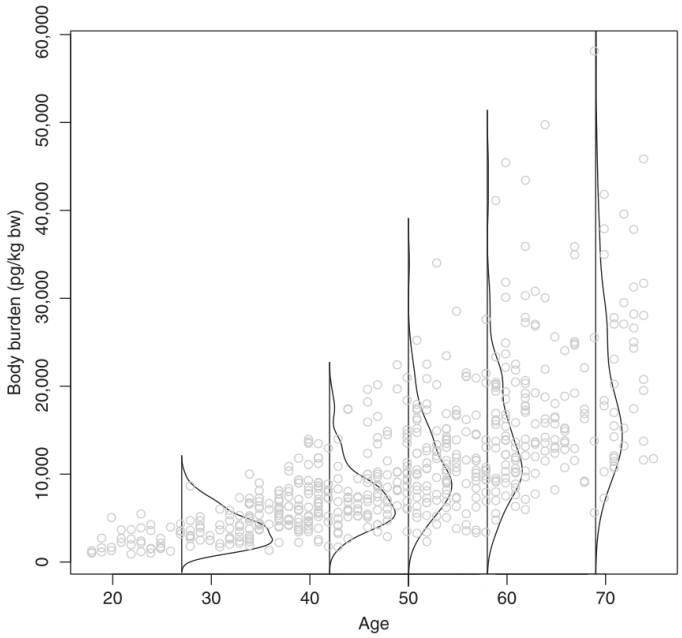


Figure 1 Observed body burdens $\tilde{X}$ and their distributions by age classes ($K = 5$)

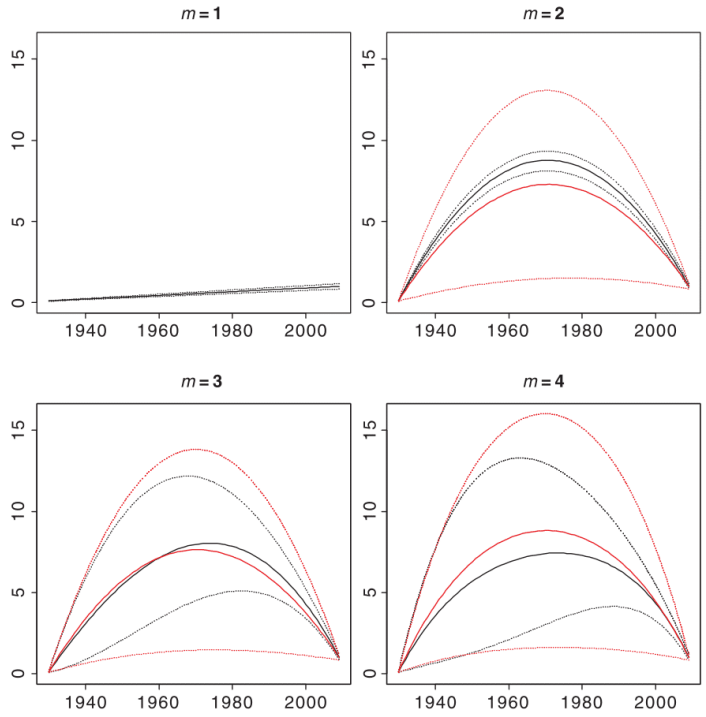


Figure 2 Function λ obtained from the prior distributions of the parameters (p5, mean, p95) in red and from posterior distributions (p5, mean, p95) in black according to the four models

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

4.4 Results and model validation

For the model $m = 2$, one reaches a rate of 41% of accepted θ (against 0% for model 1, 35% for model 3, and 24% for model 4). This model is thus considered as the best regarding the selection method proposed in Section 3. To confirm this result, the Root Mean Square Error of Prediction (RMSEP) of each model is calculated. The data set is randomly divided into three subsets respecting the proportion of individuals in each age class. Then the model is fitted from two-thirds of the data set, and the RMSEP is calculated on the remaining third using the posterior distribution of θ for each model. This procedure is repeated three times, and the mean values of the RMSEP are: 13,599 for $m = 2$, 13,535 for $m = 3$, and 13,452 for $m = 4$. Since, no value of θ was accepted for $m = 1$, no RMSEP is calculated. The RMSEP decreases when the number of parameters increases. However, the values obtained remain quite comparable, and the gain of prediction provided by an increasing number of parameters does not seem significant enough. This confirms the selection of model $m = 2$.

The scaling functions provided by each of the four tested models are shown in black in Figure 2. Since, no value of θ was accepted for $m = 1$, there is no scaling function associated with this model. The scaling function provided by the models $m = 2$, $m = 3$, and $m = 4$ are closed regarding the mean. However, the interval around the mean is much more bigger for models $m = 3$ and $m = 4$. The previous selection model concluded that model $m = 2$ was the best model. The scaling function provided by this model shows that exposure has increased since the introduction of dioxins in the 1930s to a peak in the 1970s corresponding to a seven times higher exposure than the current exposure. Then, the exposure has decreased to the current exposure values.

From the obtained scaling function, the body burdens of the individuals were predicted. Thereby, the goodness-of-fit of the model is checked by comparing the distributions of the predicted body burdens and the observed body burdens. Figure 3 shows the distributions of the observed body burdens, the body burdens predicted from KDEM without scaling function and the body burdens predicted from the corrected KDEM with the scaling function considering the mean estimation of the $m = 2$ model parameters. This comparison shows a much better fit of the corrected KDEM model to the biomonitoring data regarding the whole population.

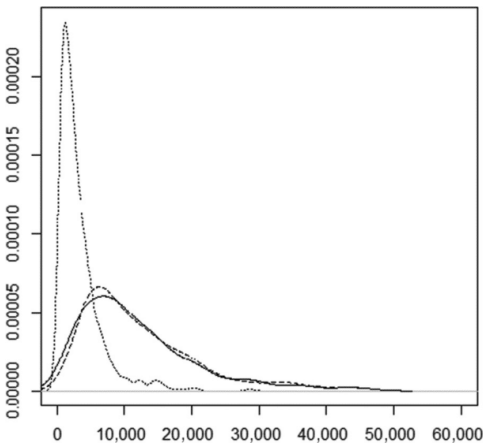


Figure 3 Distributions of the observed body burden (solid line), the body burdens predicted by KDEM (dotted line), and the body burdens predicted by the corrected KDEM $m = 2$ (dashed line) for the ICAR population (from the mean estimation of the model parameters)

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

In order to be relevant to risk assessment, the model also needs good prediction abilities. As made for the model selection, the quality of the model prediction is checked by performing a cross-validation. Body burdens for one-third of individuals are simulated with the model fitted from the remaining two-thirds. The distributions of these simulated body burdens are summarized and compared with the observed body burdens of the corresponding individuals. Results of these comparisons are given in Table 1. A distribution is considered as well predicted if the observed percentiles are included in the credibility interval of the predicted percentiles. Table 1 shows that the mean and the high percentiles are well estimated for each age class whereas for the low percentiles intervals do not contain the observed percentiles. This can be explained by the choice of specific summary statistics that only used the 50% that represent the most exposed individuals to fit the model.

5 Discussion

This paper proposes an original method based on an ABC approach to calibrate a new dynamic exposure model using biomonitoring data. This method improves chemical exposure and risk assessment as well as risk management in many ways. First, the combination of biomarker measurements with contamination levels usually exploited separately makes it possible to refine exposure assessment. The comparison of KDEM outputs with measured body burdens highlighted that the current contamination and consumption data even used in a dynamic exposure model were not sufficient to explain the actual accumulation of dioxins in the body. Indeed, environmental contamination by dioxins in France has changed with time. Individuals were thus exposed to higher intakes in the past. Unfortunately, the value of these intakes are unknown in France since no past data are available. However, the biomonitoring data from fishermen families and the modified KDEM made it possible to estimate the scaling function parameters without past data. The obtained scaling function highlights that exposure to dioxins in France peaked in the 1970s. This is fully consistent with observed evolutions in the emission of dioxin-like PCBs in the environment [28].

In France, regarding the general population, no biomonitoring data on dioxins are available. Risk assessment for the general population is thus conducted using dietary intakes provided by the Second French Total Diet Study [21]. This risk estimate does not take into account the exposure evolution and the accumulation phenomenon. This could lead to underestimate the dioxin body burdens, as observed for the Fishermen families. The KDEM corrected by the obtained scaling function can be applied to the general population intakes to predict body burdens. Therefore, assuming the exposure evolution to be the same between the fishermen population and the general population, the risk related to dioxins can be refined. The method also makes it possible to propose new risk indicators which depend on time [29]. Indeed, as individual exposure over life is described as a process $X(t)$, the maximum exposure over a given period of time: $\max_{t\in[0,T]}X(t)$ or the time spent above a reference threshold u: $\int_{t=0}^{T} I_{X(t)\geq u}dt$ can be calculated. A new time dimension can therefore be introduced in risk assessment.

The proposed method also improves risk management. The weakness of biomonitoring data lies in its inability to identify exposure sources and their related part to the overall exposure. Indeed, these data alone do not provide any information on the exposure route (inhalation, ingestion, etc.) or for example on the food the most involved in the body burden. Therefore, linking dietary exposure with blood concentration data constitutes therefore a useful tool for the interpretation of biomonitoring data by the risk managers. The promoted method can be used to interpret biomonitoring data from cohort studies of the US National Health and Nutrition Examination Survey [30] or the French National survey on nutrition and health [31] for instance. This can be done for dioxins but also for other substances like polychlorinated biphenyls, polybrominated diphenyl ethers, arsenic, cadmium, mercury, and pesticides.

An ABC was proposed to fit the corrected KDEM to the biomonitoring data. This application confirms that ABC methods are particularly suitable in case of complex likelihood. Indeed, the computational difficulties faced when trying to optimize the likelihood are bypassed in this method. However, ABC

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

methods require the choice of sufficient distances and summary statistics as highlighted by Joyce and Marjoram [14] in case of high-dimensional data in genetics. This choice can be tricky, and a wrong choice can lead to fail to fit the model [15]. The distance and the summary statistics defined here were proven able to correctly predict the body burdens of the population. They were specifically defined for risk assessment and can be used for other chemicals. An original two-stage hierarchical ABC-MCMC algorithm is proposed to perform the ABC method. By considering the model as an additional parameter, the parameter estimation for each model is performed simultaneously with the model selection. However, as highlighted by Toni et al. [12] this selection procedure may penalize/favor some models because of the choices made for the prior distribution (which is also inherent in the standard Bayesian model selection) and the tolerance. For this reason, the proposed model selection was validated by a usual criterion, the RMSEP. In our application, the RMSEP confirmed the model selection made from the hierarchical algorithm.

Using the ABC-MCMC algorithm proposed by Marjoram [15], the number of proposed values to obtain a sufficient rate of accepted θ seems reasonable since after 150,000 iterations, around 57,000 θ were accepted. Otherwise when a more complicated model of exposure is used or if it is necessary to be more restrictive regarding the tolerance δ, then it would be relevant to consider alternative algorithms. Toni et al. [12] showed that the sequential Monte Carlo method, first developed by Sisson et al. [32] and derived from a sequential importance sampling algorithm from Del Moral et al. [33], may yield much better computational performance.

Regarding dioxins, a single-compartment pharmacokinetic model, determined by a single parameter, i.e. the half-life, has been proved relevant to describe the elimination process by the human body [4, 24]. This makes it possible to link exposure with biomonitoring data from a simple model with reliable statistical properties, as described by Bertail et al. [8]. Such pharmacokinetic models are suitable for many other chemicals like polychlorinated biphenyl, polybrominated diphenyl ethers, and methyl mercury, which allows for a direct use of the method proposed in this paper with these chemicals. However, for certain other chemicals, a model which more effectively describes the anatomical, physiological, physical, and chemical phenomena involved in the absorption, distribution, metabolism, and excretion of chemicals in the human body might be necessary. Physiologically based pharmacokinetic models, providing a multi-compartment description, may be a relevant choice in these situations [34].

References

1. Jones K, Voogt P. Persistent organic pollutants (pops): state of the science. Environ Pollut 1999;100:209–21.
2. Van Leeuwen F, Younes M. Consultation on assessment of the health risk of dioxins: re-evaluation of the tolerable daily intake (TDI): executive summary. Food Addit Contam 2000;17:223–40.
3. Lyons M, Yang R, Mayeno A, Reisfeld B. Computational toxicology of chloroform: reverse dosimetry using Bayesian inference, Markov chain Monte Carlo simulation, and human biomonitoring data. Environ Health Perspect 2008;116:1040–6.
4. Lorber M. A pharmacokinetic model for estimating exposure of americans to dioxin-like compounds in the past, present, and future. Sci Total Environ 2002;288:81–95.
5. Hays S, Becker R, Leung H, Aylward L, Pyatt D. Biomonitoring equivalents: a screening approach for interpreting. Regul Toxicol Pharmacol 2007;47:96–109.
6. Bertail P, Clémençon S, Tressou J. A storage model with random release rate for modeling exposure to food contaminants. Math Biosci Eng 2008;35:35–60.
7. Verger P, Tressou J, Clémençon S. Integration of time as a description parameter in risk characterisation: application to methyl mercury. Regul Toxicol Pharmacol 2007;49:25–30.
8. Bertail P, Clémençon S, Tressou J. Statistical analysis of a dynamic model for dietary contaminant exposure. J Biol Dyn 2010;4:212–34.
9. Beaumont M, Zhang W, Balding D. Approximate Bayesian computation in population genetics. Genetics 2002;162: 2025–35.
10. Csilléry K, Blum M, Gaggiotti O, François O. Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 2010;25:490–1.

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

11. Blum M, Chi Tran V. HIV with contact tracing: a case study in approximate Bayesian computation. Biostatistics 2010;11:644–60.
12. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J Royal Soc Interface 2009;6:187–202.
13. Blum M. Approximate Bayesian computation: a nonparametric perspective. J Am Stat Assoc 2010;105:1178–87.
14. Joyce P, Marjoram P. Approximately sufficient statistics and Bayesian computation. Stat Appl Genet Mol Biol 2008;7:1544–6115.
15. Marjoram P. Markov Chain Monte Carlo without likelihoods. Proc Natl Acad Sci USA 2003;100:15324–8.
16. Rachev S. Probability metrics and the stability of stochastic models, 1991.
17. Lehmann E, Romano J. Testing statistical hypotheses, 2005.
18. Anses and InVS. Etude Nationale d'Imprégnation Aux Polychlorobiphényles des Consommateurs de Poissons d'Eau Douce (ICAR-PCB). Rapport d'Étude Scientifique et Synthèse, 2011.
19. Deurenberg P, Weststrate J, Seidell J. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br J Nutr 1991;65:105–14.
20. Leblanc J. CALIPSO, fish and seafood consumption study and biomarker of exposure to trace elements, pollutants and omega3, 2006.
21. Sirot V, Tard A, Venisseau A, Brosseaud A, Marchand P, Le Bizec B. Dietary exposure to polychlorinated dibenzo-p-dioxins, polychlorinated dibenzofurans and polychlorinated biphenyls of the French population: results of the second French total diet study. Chemosphere 2012;88:492–500.
22. Lioret S, Dubuisson C, Dufour A, Touvier M, Calamassi-Tran G, Maire B, Volatier J, Lafay L. Trends in food intake in french children from 1999 to 2007: results from the INCA (Étude Individuelle Nationale Des Consommations Alimentaires) dietary surveys. Br J Nutr 2010;103:585–601.
23. Sinkkonen S, Paasivirta J. Degradation half-life times of PCDDs, PCDFs and PCBs for environmental fate modeling. Chemosphere 2000;40:943–9.
24. Van Der Molen G, Kooijman S, Slob W. A generic toxicokinetic model for persistent lipophilic compounds in humans: an application to TCDD. Fundam Appl Toxicol 1996;31:83–94.
25. Aylward L, Hays S. Temporal trends in human TCDD body burden: decreases over three decades and implications for exposure levels. J Expo Anal Environ Epidemiol 2002;12:319–28.
26. Wegmann D, Leuenberger C, Excoffier L. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 2009;182:1207–18.
27. Robert G, Gelman A, Gilks W. Weak convergence and optimal scaling of random walk metropolis algorithms. Ann Appl Probability 1997;7:110–20.
28. INERIS. Données technico économiques sur les substances chimiques en France: les polychlorobiphenyles (PCB), 2011. Available at: http://www.ineris.fr/substances/fr/.
29. Clémençon S, Tressou J. Exposition aux risques alimentaires et processus stochastiques: le cas des contaminants chimiques. J De La Société Française De Statistique 2009;150:3–29.
30. Centers for disease Control & Prevention. National Health and Nutrition Examination Survey, 2000. Available at: http://www.cdc.gov/nchs/about/major/nhanes/nhanes99_00.htmS.
31. Falqa G, Zeghnouna A, Pascala M, Vernayb M, Le Stratc Y, Garnierd R, Olichone D, Bretina P, Castetbonb K, Fréry N. Blood lead levels in the adult population living in France the French nutrition and health survey (ENNS 2006–2007). Environ Int 2011;37:565–71.
32. Sisson S, Fan Y, Tanaka M. Sequential Monte Carlo without likelihoods. Proc Natl Acad Sci USA 2007;104:1760–5.
33. Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. J Royal Stat Soc Ser B 2006;68:411–32.
34. Reddy M, Yang R, Andersen M, Clewell H. Physiologically based pharmacokinetic modeling: science and applications. Hoboken, New Jersey: John Wiley and Sons, 2005.

Note: Supplemental Material: The online version of this article (DOI: 10.1515/ijb-2013-0062) offers supplementary material, available to authorized users.

Authenticated | camille.bechaux@gmail.com author's copy
Download Date | 7/4/14 9:16 AM

2.2.1 Summary of the objectives, methodological developments, main results and limitations of the work

Starting from the previous work that proposes to integrate biomonitoring information in the modeling of exposure, the objective of this work is to develop an integrative approach to assess the risk through a body burden approach taking into account the long-term time trend of the exposure to persistent chemicals, their bioaccumulation in the human body and the corresponding toxicity. This methodology is applied on the long-term exposure of dioxins, furans and dl-PCBs in France.

The dynamic model of exposure developed in the previous work is used and parameters are also estimated by Bayesian inference using the same biomonitoring data (i.e. of blood concentration of dioxins in a group of French fishermen families). However, on contrary to the previous work, specific historic evolution of exposure and kinetics were considered for each congener (or group of congeners) in order to investigate cumulative risk assessment of chemicals with very different half-lives. These adjusted models are then used to predict the body burden of the French population combining consumption data on this population and French contamination data. The associated toxic risk is assessed from body burdens indicative for various types of toxicity through a probabilistic approach using distributions for BMDL and for predicted population body burdens.

This work provides an estimation of the historic evolution of the exposure to dioxins and furans in France between 1930 and today and shows that evolution of dl-PCBs would not have been significant. The risk related to thyroid, reproductive and hepatic toxicity in 4 different age groups of the French population are also provided. Finally, as an example of application of such approach to test specific monitoring scenario and effectiveness of risk management scenario, body burden of the French population in 2030 was estimated assuming that exposure would remain stable in the future.

Such approach highlights that combining several methods and data can improve the risk assessment of persistent chemicals. However, the simplicity of the kinetic modeling used in this work on one side allowed to consider the difference in kinetic between congeners, on the other still do not allow to integrate more mechanistic information. It also does not allow to consider in a single model several pathways of exposure.

2.2.2 Publication

Regulatory Toxicology and Pharmacology 70 (2014) 261–269

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

An integrative risk assessment approach for persistent chemicals: A case study on dioxins, furans and dioxin-like PCBs in France

Camille Béchaux [a,*], Marco Zeilmaker [b], Mathilde Merlo [a], Bas Bokkers [c], Amélie Crépet [a]

[a] ANSES, French Agency for Food, Environmental and Occupational Health Safety, 27-31 Avenue du Général Leclerc, 94701 Maisons-Alfort, France
[b] RIVM, National Institute for Public Health and the Environment, Centre for Nutrition, Prevention and Health, PO Box 1, 3720 BA Bilthoven, The Netherlands
[c] RIVM, National Institute for Public Health and the Environment, Centre for Safety of Substances and Products, PO Box 1, 3720 BA Bilthoven, The Netherlands

ARTICLE INFO

Article history:
Received 28 October 2013
Available online 16 July 2014

Keywords:
Cumulative risk assessment
Dioxins
PCBs
Biomarker
Pharmacokinetic
Benchmark dose

ABSTRACT

For persistent chemicals slowly eliminated from the body, the accumulated concentration (body burden), rather than the daily exposure, is considered the proper starting point for the risk assessment. This work introduces an integrative approach for persistent chemical risk assessment by means of a dynamic body burden approach. To reach this goal a Kinetic Dietary Exposure Model (KDEM) was extended with the long term time trend in the exposure (historic exposure) and the comparison of bioaccumulation with body burden references for toxicity. The usefulness of the model was illustrated on the dietary exposure to PolyChlorinatedDibenzo-p-Dioxins (PCDDs), PolyChlorinatedDibenzoFurans (PCDFs) and PolyChlorinated Biphenyls (PCBs) in France. Firstly the dietary exposure to these compounds was determined in 2009 and combined with its long term time trend. In order to take differences between the kinetics of PCDD/F and dl-PCBs into account, three groups of congeners were considered i.e. PCDD/Fs, PCB 126 and remaining dl-PCBs. The body burden was compared with reference body burdens corresponding to reproductive, hepatic and thyroid toxicity. In the case of thyroid toxicity this comparison indicated that in 2009 the probability of the body burden to exceed its reference ranged from 2.8% (95% CI: 1.5–4.9%) up to 3.9% (95% CI: 2.7–7.1%) (18–29 vs. 60–79 year olds). Notwithstanding the decreasing long-term time trend of the dietary dioxin exposure in France, this probability still is expected to be 1.5% (95% CI: 0.3–2.5%) in 2030 in 60–79 olds. In the case of reproductive toxicity the probability of the 2009 body burden to exceed its reference ranged from 3.1% (95% CI: 1.4–5.0%) (18–29 year olds) to 3.5% (95% CI: 2.2–5.2%) (30–44 year olds). In 2030 this probability is negligible in 18–29 year olds, however small though significant in 30–44 year olds (0.7%, 95% CI: 0–1.6%). In the case of hepatic toxicity the probability in 2009 even in 60–79 year olds already was negligible. In conclusion this approach indicates that in France dioxin levels in food form a declining, though still present, future health risk with respect to thyroid and reproductive toxicity.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Chemicals like PolyChlorinatedDibenzo-p-Dioxins (PCDDs), PolyChlorinatedDibenzoFurans, (PCDFs), PolyChlorinated Biphenyls (PCBs), PolyBrominated Diphenyl Ethers (PBDEs), DichloroDiphenyl-Trichloroethane (DDT), cadmium and methyl mercury are known to be slowly eliminated from the body. The successive exposures to these chemicals therefore results in bioaccumulation, i.e. increasing levels in the body. For this reason, the accumulated concentration in the body (body burden), rather than the daily exposure, is considered as the proper starting point for the risk assessment of these chemicals (Van Leeuwen and Younes, 2000; US EPA, 2010).

In comparison with a traditional risk assessment a body burden approach requires various methodological adaptations (NRC, 2006; Gies et al., 2007; Sirot et al., 2012). Firstly, the accumulation in the human body results from the net effect of repeated intakes and elimination. The incorporation of such accumulation in the risk assessment procedure needs the explicit incorporation of chemical kinetics. In this context several kinetic models ranging from single compartment toxico-kinetic (TK) models to complex multi-compartment physiologically-based toxicokinetic (PBTK) models may serve the purpose. Secondly, the exposure to bioaccumulating agents may not be limited to one single chemical, but may consist of the exposure to a mixture of related chemicals. Here PCDD/Fs and dl-PCBs are a typical example (Van den Berg et al., 2006).

* Corresponding author. Fax: +33 (0)1 49 77 38 92.
E-mail address: camille.bechaux@anses.fr (C. Béchaux).

http://dx.doi.org/10.1016/j.yrtph.2014.07.004
0273-2300/© 2014 Elsevier Inc. All rights reserved.

Thirdly, the accumulation kinetics of persistent chemicals may be impeded by their complex historic exposure pattern. For example, it is well highlighted that the exposure to environmental contaminants such as PCDD/Fs have changed significantly over the last decades due to changes in production, in use and successive regulations (Hays and Aylward, 2003; De Mul et al., 2008). Such changes may be different for closely related chemicals such as PCDDs, PCDFs and dl-PCBs (Consonni et al., 2012). In the case of PCDD/Fs previous studies already have addressed the historic exposure pattern of PCDD/Fs (Van der Molen et al., 1996; Pinsky and Lorber, 1998; Aylward and Hays, 2002). To the best of our knowledge, none of these studies performed a full quantitative risk assessment integrating the historic, long term, exposure of PCDD/Fs and dl-PCBs, the resulting bioaccumulation and the corresponding toxic risk. Therefore, the objective of this study is to develop an integrative risk assessment methodology which takes into account the long-term time trend of the exposure to persistent chemicals, their bioaccumulation in the human body and the corresponding toxicity and to apply this methodology on the long-term exposure of PCDD/Fs and dl-PCBs in France.

As dietary intake represents more than 90% of the human exposure to PCDD/Fs and dl-PCBs (WHO, 2002) only this route of exposure was considered. The current dietary exposure in France was assessed by means of a food questionnaires and contamination data in food and scaled over time in concordance with the method of Van der Molen et al. (1996), Van der Molen (1998). The resulting long-term intake was used as input for the Kinetic Dietary Exposure Model (KDEM) as proposed by Verger et al. (2007) for the exposure to methyl mercury, its accumulation in the body and its corresponding toxic risk. In order to take differences between the kinetics of PCDD/F and dl-PCBs into account (Milbrath et al., 2009) and possible differences in historic evolution of exposure, three groups of congeners were considered separately: PCDD/Fs, PCB 126 and remaining dl-PCBs.

This strategy resulted in long-term simulations of the body burden of PCDD/Fs and dl-PCBs in the French population. The toxic risk, i.e. reproductive, hepatic and thyroid toxicity, associated with these body burdens was obtained by comparison with reference body burdens for these types of toxicity. In concordance with current dioxin risk assessment procedures (Van Leeuwen and Younes, 2000; JECFA, 2001), the latter were obtained by means of interspecies extrapolation of animal toxicity. In this work, animal toxicity was analyzed by means of Bench Mark Dose modeling and associated animal body burdens were extrapolated to man. Body burdens estimated from our model were then compared with reference body burdens through a probabilistic approach using distributions for BMDL and population body burdens.

2. Material and methods

2.1. Dietary exposure and blood concentrations of a fishermen French population

Biomonitoring data such as concentrations in blood provides relevant information for the estimation of kinetic parameters such as half-lives and exposure trends of dioxins (Van der Molen et al., 1996; Van der Molen, 1998; Aylward et al., 2005; Hsu et al., 2010; Fromme et al., 2009). Hence blood concentration of PCDD/Fs and dl-PCBs measured in a population of French fishermen families were used to characterize the long-term exposure of the current French population (Anses/InVS, 2011). The blood data were obtained from the French National study on Dioxin blood levels in French consumers of freshwater fish (ICAR study, Anses/InVS, 2011). In 2009, this study recorded the dioxin concentrations in blood (pg per g fat) of 606 adults between 18 and 75 years of age. All individuals were from fishermen families. The concentrations were converted into total body burdens assuming a balance between all lipid compartments in the body and that the fat content of the human body (BF) varies with age, sex, weight and height according to the equation proposed by Deurenberg et al. (1991) i.e. BF% = 1.20 BMI + 0.23 age − 10.8 sex − 5.4. Consumed quantities of 115 food items covering the main diet of each individual taking part in the study, were collected by means of a food frequency questionnaire and a photograph manual of portions size. Dietary exposure of the 606 individuals was assessed by combining these consumed quantities with concentration data from the ICAR study for freshwater fishes, the CALIPSO study (Leblanc, 2006) for marine fishes and other sea foods and the second French Total Diet Study (Sirot et al., 2012) for other foods known to contribute to the exposure to dioxins. Body burdens and levels of dietary exposure for different age classes of the ICAR population are summarized in Table 1.

2.2. Dietary exposure of the general French population

Exposure data of the French population are provided by the Second French Total Diet Study (Sirot et al., 2012). "Total Diet Studies (TDS)" follow a standardized international methodology and aim at assessing the dietary exposure of individuals for many contaminants taking into account residue levels in foods as consumed at home. In a TDS, exposure is assessed by combining individual food consumption data and residue levels data from food sample analysis. The individual food consumption data used in the Second French Total Diet Study (Sirot et al., 2009) were provided by the second "Individual and National Study on Food Consumption", INCA2 survey, carried out by the French Food Safety Agency between late 2005 and April 2007 (Dubuisson et al., 2010; Lioret et al., 2010). Two independent population groups were included in the survey: 2624 adults aged 18–79 years and 1455 children aged 3–17 years. Each participant was asked to complete a 7-day food diary as well as other questionnaires on anthropometric and socio-economic factors. Foods declared were subsequently categorized into 1305 "as consumed" food items. The mean of the quantities of the same food consumed by each individual during the week is used in order to assess chronic exposure. The consumption data were combined with contamination data measured for 212 core foods selected to cover about 90% of the whole diet in terms of quantity consumed through a methodology described in Sirot et al. (2009). Levels of dietary exposure for different age classes of the French population are summarized in Table 2.

2.3. The kinetic model

2.3.1. Model definition

Verger et al. (2007) proposed a Kinetic Dietary Exposure Model (KDEM) which describes the dynamic evolution of the dietary exposure over time. Between intakes, the change of human body burden x is described by a simple one-compartment, first order, pharmacokinetic model. For chemicals with long half-lives, absorption time is insignificant compared to the elimination time. In concordance with Emond et al. (2005); US EPA (2010) and EFSA (2011) the absorption of dioxins, furans and DL-PCBs from food was considered to be complete, i.e. 100%. Given one compartmental kinetics, the elimination rate between two eating occasions is given as:

$$\frac{dx}{dt}(t) = -k \times x(t) \quad (1)$$

The elimination rate k is described by the apparent elimination half-life HL as $k = \ln(2)/HL$.

Table 1
Distribution characteristics of the dietary exposure, the observed body burdens and the simulated body burdens of the ICAR population (Anses, 2001) for PCDD/F, PCB-126 and the remaining dl-PCBs and for the total TEQ.

	Classes of age	Dietary exposure per day (pg TEQ_{2005}/kg bw)			Observed body burden (ng TEQ_{2005}/kg bw)			Estimated body burden (ng TEQ_{2005}/kg bw)			Half-life considered (years)
		2.5th	50th	97.5th	2.5^{th}	50th	97.5th	2.5th	50th	97.5th	Value
Dioxins/furans	18–29	0.2	0.3	0.9	1.2	2.8	7.8	1	2.1	5.3	8.7 (Aylward et al., 2013)
	30–44	0.2	0.4	0.9	1.3	3.5	8.9	1.3	3.2	8	
	45–59	0.2	0.4	1.2	1.4	4.1	11	1.8	4	10.1	
	60–79	0.2	0.4	1	1.4	4.3	10	1.9	4	12	
PCB 126	18–29	0.3	0.6	4.5	0.2	0.9	7	0.3	1	8.5	2.7 (Milbrath et al., 2009)
	30–44	0.3	0.8	4.1	0.3	1.1	6.1	0.4	1.1	4.8	
	45–59	0.3	0.9	4.7	0.3	1.3	6.6	0.4	1.3	5.9	
	60–79	0.3	1	3.7	0.3	1.4	8.6	0.4	1.5	8.5	
PCB-DL without PCB 126	18–29	0.02	0.1	1.1	0.1	0.6	3	0.1	0.4	7	11.8 (Milbrath et al., 2009)
	30–44	0.03	0.1	1.1	0.2	0.9	2.5	0.2	0.5	3.5	
	45–59	0.03	0.1	0.9	0.2	1	3.5	0.2	0.8	4.7	
	60–79	0.03	0.1	0.8	0.2	1.1	3	0.2	1	6	
Dioxins/Furans/DL-PCB	18–29	0.5	1.1	6.5	1.6	5	18	1.6	3.8	21	
	30–44	0.5	1.3	5.9	1.9	5.5	15	2	4.8	15	
	45–59	0.5	1.5	7.1	2	6.3	21	2.5	6.1	21	
	60–79	0.5	1.6	5.6	2.1	6.9	20	2.5	6.6	25	

PCDD/Fs: half-life for the TEQ (Aylward et al., 2013), PCB126: median value (Milbrath et al., 2009), dl- PCBs: average of median values (Milbrath et al., 2009).

Table 2
Distribution characteristics of the dietary exposure and the simulated body burdens in 2009 and 2030 in the general French population for PCDD/F, PCB-126 and the remaining dl-PCBs and for the total TEQ.

	Classes of age	Dietary exposure per day (pg TEQ_{2005}/kg bw)			Estimated body burden in 2009 (ng TEQ_{2005}/kg bw)			Estimated body burden in 2030 (ng TEQ_{2005}/kg bw)			Half-life considered (years)
		2.5th	50th	97.5th	2.5th	50th	97.5th	2.5th	50th	97.5th	Value
Dioxins/Furans	18–29	0.2	0.3	0.9	1.2	2.8	7.8	0.3	0.8	1.8	8.7 (Aylward et al., 2013)
	30–44	0.2	0.4	0.9	1.3	3.5	8.9	0.4	1	1.9	
	45–59	0.2	0.4	1.2	1.4	4.1	11	0.4	1.2	2.1	
	60–79	0.2	0.4	1	1.4	4.3	10	0.8	1.3	2	
PCB 126	18–29	0.3	0.6	4.5	0.2	0.9	7	Idem 2009			2.7 (Milbrath et al., 2009)
	30–44	0.3	0.8	4.1	0.3	1.1	6.1				
	45–59	0.3	0.9	4.7	0.3	1.3	6.6				
	60–79	0.3	1	3.7	0.3	1.4	8.6				
PCB-DL without PCB 126	18–29	0.02	0.1	1.1	0.1	0.6	3	Idem 2009			11.8 (Milbrath et al., 2009)
	30–44	0.03	0.1	1.1	0.2	0.9	2.5				
	45–59	0.03	0.1	0.9	0.2	1	3.5				
	60–79	0.03	0.1	0.8	0.2	1.1	3				
Dioxins/Furans/DL-PCB	18–29	0.5	1.1	6.5	1.6	5	18	0.6	2.3	12	
	30–44	0.5	1.3	5.9	1.9	5.5	15	0.9	3	10	
	45–59	0.5	1.5	7.1	2	6.3	21	0.9	3.5	12	
	60–79	0.5	1.6	5.6	2.1	6.9	20	1.3	3.6	14	

PCDD/Fs: half-life for the TEQ (Aylward et al., 2013), PCB126: median value (Milbrath et al., 2009), dl- PCBs: average of median values (Milbrath et al., 2009).

The accumulation in the body is then calculated as follows. At each eating occasion $T_n \varepsilon, n \in N$, the dynamic exposure process jumps with size equal to the intake U_n which is related to this eating occasion. Between intakes the exposure process decreases exponentially according the Eq. (1). The value of the total body burden X_{n+1} at intake time T_{n+1} is thus defined as

$$\begin{aligned} X_{n+1} &= X_n e^{-kT_{n+1}} + U_{n+1} \qquad n \in N \\ X_0 &= x_0 \end{aligned} \tag{2}$$

with $\Delta T_{n+1} = T_{n+1} - T_n, n > 1$, the time between two intakes and X_0 the initial body burden.

In order to take into account the possible historic variation of the exposure through time, a function of time f is introduced in KDEM to correct the intake estimated from current data (in 2009 in this study):

$$\begin{aligned} X_{n+1} &= X_n e^{-kT_{n+1}} + f(y).U_{n+1}, \qquad n \in N \\ X_0 &= x_0 \end{aligned} \tag{3}$$

where y is the year corresponding to the food intake.

2.3.2. Model parameters and group of congeners

Milbrath et al. (2009) reviewed the half-lives for individual PCDD/F and dl-PCB and showed that, on average, the half-life of individual PCDD/F congeners approximates that of TCDD. Sequentially Aylward et al. (2013) provided a half-life for the PCDD/F TEQ. In contrast dl-PCBs showed a wider variety. Since PCB 126 is the dominant dl-PCBs in French food products (Tard et al., 2007), contributes to around 60% of the dietary exposure of the French population (Table 2) and has a relative high toxic potency (Toxic Equivalency Factor of 0.1, to be compared with the <0.03 for the other dl-PCBs and 1.0 for the most toxic reference, i.e. 2,3,7,8-TCDD), this congener was modeled separately from the PCDD/Fs and other dl-PCB. The apparent elimination half-lives used for PCDD/Fs, PCB-126 and the remaining dl-PCBs are given in Tables 1 and 2.

The initial body burden x_0 was considered to be equal to the burden of breastfed children of 6 months of age. In breast milk, Zeilmaker et al. (2002) estimated for 2009 a concentration of 8 pg TEQ/g milk fat (PCDD/Fs: 6.1 pg/g milk fat; PCB-126: 1.77 pg/

g milk fat; remaining dl-PCBs: 0.13 pg/g milk fat). Given a daily breast milk consumption of 800 ml/day and a fat percentage of 2.7% (Zeilmaker, personal communication; Frery and Deloraine, 2000), a daily breast milk fat consumption of 21.6 g fat/day could be calculated. Given a body weight 7.6 kg for a 6 months old child (Winter et al., 2003), the initial body burden of breastfed children at 6 months of age was then calculated at $6 \times 30 \times 21.6/7600 = 4.1$ ng TEQ/kg bw. It corresponds to 3.1 ng/kg bw for PCDD/Fs, 0.90 ng/kg bw for PCB-126 and 0.066 ng/kg bw for remaining dl-PCBs.

2.3.3. Simulation of individual body burden

In order to take into account the inter-individual variability, a lifetime exposure trajectory started at the year of birth is simulated for each individual. The consumption is known to be subject to large fluctuations during childhood (Lioret et al., 2010). The age dependency of the intake of the general French population calculated from the INCA2 population is shown in Fig. S1 in the Supplementary Material. Consequently, reference intakes U_n at age lower than 18 are simulated by combining the consumption distribution of children observed in the INCA2 study. To be more specific, for each age a between 3 and 17 occurring in the INCA2 database, a reference distribution of the dioxins intake at the age a is estimated by combining the consumption and the contamination data as explained in Section 2.2. Therefore, 2624 individual lifetime exposure trajectories are simulated as follow. Starting with x_0, then for each intake time which occurs when the individual is of age a, an intake value U_n is sampled according to the distribution of French children of age a. When older than 18 the age dependency of the variability in the intake with age was neglected and U_n equal the intake of the individual at his current age.

2.4. Calibration of the function f and validation of the model

2.4.1. PCDD/Fs

Regarding PCDD/Fs Van der Molen et al. (1996), Van der Molen (1998), Pinsky and Lorber (1998), Aylward and Hays (2002), highlighted a time-trend of the historic exposure that should be taken into account to predict verifiable body burdens. Unfortunately the time-trend in the historic exposure to PCDD/F and dl-PCBs, i.e. consumption pattern and occurrence data, is not available in France. However, cross-sectional biomonitoring data such blood concentrations may be of help. Indeed, when collected at a certain point in time age-dependent PCDD/F and dl-PCB concentrations reveal the accumulation of these compounds resulting from their historic exposure. This allows for defining a function $f(y)$ which scales the age-dependent intake as determined in a certain reference year y (2009 in this work) over time. Fitting a *known* kinetic model with input the *known* age-dependent intake coupled to the *unknown* function $f(y)$ to the *known* biomonitoring data then results in defining f. In this study Bayesian inferences were carried out to calibrate the function $f(y)$ on French blood data of the ICAR cohort. The chosen function $f(y)$ contains five shape parameters (for details see van der Molen et al., 1996; van der Molen, 1998). These parameters were calibrated on (cross-sectional) blood data in Germany and intake data from the Dutch population. As The Netherlands and France have implemented the same European rules with regard to dioxin emissions to the environment it was decided to hold the parameters which scale the function $f(y)$ across time (parameters B and G) at their default value, i.e. the values 60 and 95 as calibrated on the German/Dutch data which scale and shift $f(y)$ over the time axis. Similarly the parameters which determine the shape of $f(y)$ (parameter C, default value: 3.5) and the background concentration in blood (parameter A, default value: 0.1) were fixed at their default values. Only the parameter H affecting the height of the function was (re)calibrated on the French data. In this procedure, it was chosen not to give *a priori* information on the value of the parameter H (i.e. the prior is a gamma distribution with shape parameter $k = 0.001$ and scale parameter $\theta = 0.001$). The stability of the estimates was checked by estimating the parameter H of $f(y)$ from several samples of 130 individuals. The model was implemented with the software OpenBUGS version 3.1.2 (Lunn et al., 2009).

2.4.2. dl-PCBs

Regarding the dl-PCBs, and in particular PCB 126, Consonni et al. (2012) showed no age-dependency in the blood concentrations of these compounds. Therefore, defining $f(y)$ seems not necessary for dl-PCBs.

2.4.3. Validation

Before using it to predict body burdens of the general French population, the model is implemented on the ICAR population in order to be validated by comparison between observed and estimated body burdens. For each age, a 95% confidence interval is calculated as the interval which contains 95% of the body burdens values predicted for individuals of the corresponding age. Then, it is checked if the observed body burdens are included in this interval.

2.5. Risk assessment

2.5.1. Reference values for body burdens

To interpret the calculated body burdens in terms of toxic risk, reference body burdens were calculated by means of benchmark dose (BMD) analysis. In BMD modeling a dose–response model is fitted to the toxicity data. Subsequently, the dose is derived to correspond to a predetermined percentage change in response (compared to the background) in the case of continuous toxicity data or which results in an effect in the typical, average individual (i.e. ED_{50}) in the case of dichotomous data. The BMD is reported with its lower (5th percentile) and upper (95th percentile) confidence limits: BMDL and BMDU respectively.

Supplementary Material S2 shows the BMD analysis of reproductive, hepatic and thyroid toxicity in experimental animals. This analysis resulted in a BMDL of 171 ng TEQ/kg body fat for reproductive toxicity, corresponding with a 10% decrease in sperm production in offspring of dams which were exposed to TCDD. For thyroid toxicity, the BMDL is 186 ng TEQ/kg body fat, corresponding with a 10% reduction in serum TT4. For hepatic toxicity, the BMDL is 1520 ng TEQ/kg body fat, corresponding with an increase of fatty infiltration in the liver in the average animal.

To extrapolate them to human inter-species and intra-species variability so-called Extrapolation Factor (EF) are used. In the case of dioxins this extrapolation is limited to the extrapolation of inter-human variability of dioxin kinetics (WHO, 2005; SCF, 2001; JECFA, 2001). This variability was quantified by means of a log normally distributed EF with a geometric mean of 1 and a geometric standard-deviation of 2. Then, the Human BMDLs in ng TEQ/kg body fat were converted in ng TEQ/kg body weight considering 20% of fat in human body. Thus, the final BMDL distribution for

- reproductive toxicity was characterized by a mean of 34 ng TEQ/kg body weight, a 2.5th percentile of 8 ng TEQ/kg body weight and a 97.5th percentile of 125 ng TEQ/kg body weight. In pregnant women these body burdens may lead to reduced sperm production in male offspring.
- thyroid toxicity was characterized by a mean of 39 ng TEQ/kg body weight, a 2.5th percentile of 10 ng TEQ/kg body weight and a 97.5th percentile of 146 ng TEQ/kg body weight. During life, these body burdens may lead to reduce thyroid hormone levels.

- hepatic toxicity was characterized by a mean of 302 ng TEQ/kg body weight, a 2.5th percentile of 72 ng TEQ/kg body weight and a 97.5th percentile of 1160 ng TEQ/kg body weight. During life these body burdens may lead to fatty infiltration in the liver.

2.5.2. Risk indicators and probabilistic risk assessment

The induction of reproductive toxicity in animals has shown that this effect can be caused by a single exposure of the pregnant dams. Therefore, as a "worst case" approach, it was assumed that a body burden of women of reproductive age (18–44 years) which exceeds the reference BMDL for this effect for at least one day is indicative for reproductive toxicity. Regarding thyroid and hepatic toxicity the same "worst case" approach was applied (see Supplementary Material S3).

Since a distribution was used to take inter-individual variability, a distribution is available for the human BMDL. Since individual simulations were implemented with the model, the body burdens of the 2624 individuals from the INCA2 study were estimated. Consequently, in concordance with Van Der Voet and Slob (2007), a probabilistic risk assessment was conducted. Its implementation can be described as follow:

(1) A value of BMDL in ng TEQ/kg bw is randomly sampled from the BMDL distribution and written u.
(2) For each of the 2624 individuals, the estimated body burdens for PCDD/Fs, PCB 126 and remaining dl-PCBSs through the life are summed and it is determined whether the estimated individual body burden exceeded the threshold u at least one time i.e. for a time period of at least one day.
(3) The number of individuals highlighted in the step 2 provides the probability of exceeding the threshold u

These steps were repeated 1000 times in order to provide a mean probability (i.e. the mean of the 1000 probabilities provided by the step 3) and its 95% confidence interval (i.e. the 2.5th and the 97.5th percentile).

3. Results

3.1. Model validation and scaling of the dietary intake across time

As shown in Fig. 1A for PCB 126 and Fig. 1B for the remaining dl-PCBs a kinetic model without any specific correction for a historic time-trend for the exposure was able to simulate the accumulation of these compounds in the body of a French subpopulation, i.e. French fishermen. Indeed, regarding PCB126, for 551individuals (i.e. 91% of the population) the observed body burdens, i.e. blood fat concentrations, fell within the 95% confidence interval of the simulated blood concentrations of the corresponding age. Regarding the remaining dl-PCBs, for 569 individuals (i.e. 94% of the population) the observed body burdens are in this confidence interval. The quality of the model prediction can also be checked from Table 1 which summarizes the simulated and the observed body burdens of individuals of the ICAR study as stratified in 4 age classes.

Regarding PCDD/Fs (Fig. 1C), the kinetic model without historic correction underestimates the body burdens and the addition of the scaling function $f(h)$ was necessary to describe the data. Indeed, without the correction, the kinetic model predicts correctly the body burden for less than 40% of the population. This is low compared to the 91% and 94% for the PCB 126 and other dl-PCBs.

Consequently, a scaling function was calibrated to the data from the approach described in Section 2.4. It results in $H = 1.6$ with a 95% confidence interval of [1.45; 1.75]. Fig. 2 shows the function $f(y)$ which scales the dietary exposure in France across time for $H = 1.6$. From Fig. 2, it can be concluded that the biomonitoring data of the ICAR study are compatible with a historic time trend in the exposure which peaks in the 1960s at a level which is 6 times higher than the exposure in France in 2009. Fig. 1C indicates that the model correctly predicts the current age-dependency of the body burdens for families of French fishermen, taking their individual intake history into account. To check the quality of the model, a cross validation was implemented as follow: the data set is randomly divided into 3 subsets respecting the proportion of individuals in each age class. Then the model is fitted from two-thirds of the data set and the number of individual for which the observed body burdens are in the confidence interval of the estimated once is calculated. This procedure is repeated three times and in mean the model predicts well the body burden for 92% of the population. Compared to the 40% of the previous model, it shows that taking into account the historic variation of the exposure in the model improves significantly the prediction quality of the body burden. This finding is also illustrated in Fig. 3 which shows the comparison of the distribution of the simulated sum of the PCDD/F, PCB-126 and remaining dl-PCB TEQ in the blood of the ICAR population with the observed TEQ in blood.

3.2. Simulation of current body burdens of the general French population and risk assessment

From the dietary intakes of the general French population in 2009 and the models proposed in Section 2 current body burdens of the French population were calculated for 3 groups of congeners: (i) PCDD/Fs, (ii) PCB 126 and (iii) the remaining dl-PCBs. The simulated body burdens calculated for four different age groups are presented in Table 2.

Regarding the PCFF/Fs, as expected older individuals have a higher body burden than younger individuals, thereby reflecting the effect of bioaccumulation. Due to their relative high contribution to the dietary exposure and their long half-lives, PCDD/Fs contribution to total body burden is high (around 70%). Regarding PCB 126, Table 2 indicates the high contribution of this congener to the dietary exposure, i.e. higher than 50%, whereas the contribution of PCB 126 to the total body burden is less than 10%. This is caused by the relative low half-life of PCB 126 compared to PCDD/Fs and the other dl-PCB congeners. As expected, regarding other dl-PCBs, due to the mean half-life considered that is around 12 years, the contribution of these congeners to the body burden is higher than the contribution to the dietary exposure.

The distributions of the body burdens presented in Table 2 were compared with the distribution of the reference body burdens indicative for toxicity. As shown in Table 3, the probability that the simulated body burden exceeds the BMDL for reproductive toxicity is below 3.5% regarding the mean value and 5.2% regarding the 97.5th. The probabilities that the body burden in 2009 exceeds the human reference BMDL for thyroid toxicity range from 2.8% in 18–29 years old (mean value, 97.5th percentile: 4.9%) to 3.9% in 60–79 year olds (97.5th percentile: 7.1%). With respect to hepatic toxicity the probability of the current body burden to exceed the reference BMDL for this effect (even in 60–79 year olds) is negligible.

3.3. Simulation of future body burdens of the general French population and corresponding risk

Although Fig. 2 suggests that PCDD/Fs dietary exposure in France would continue to decline after 2009, the unknown uncertainty herein in fact makes such an extrapolation speculative. Consequently, in this work, a "worst case" scenario was chosen i.e. the PCDD/Fs dietary exposure would stay at its current level after 2009. Under this scenario, the body burdens of the French population in 2030 were simulated. Table 2 presents the comparison of the 2030 body burden with the simulated 2009 body

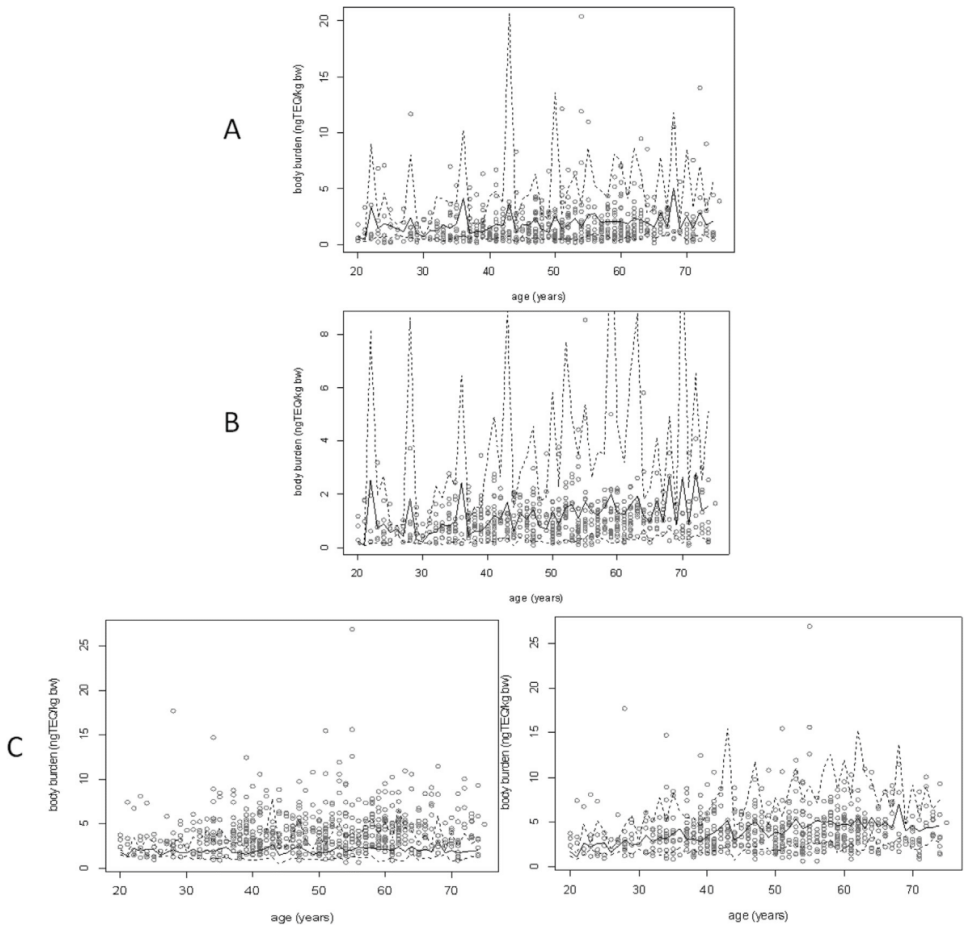


Fig. 1. Simulated body burden (ng TEQ/kg bw) as a function of age (the arithmetic mean: solid line; the 2.5th and the 97.5th percentile: dotted line) and observed body burden from ICAR study (dots) for (A) PCB 126, (B) remaining dl-PCBs and (C) PCDD/Fs without historic correction ($f(y) = 1$) (left) and with historic correction ($f(y)$ as in Fig. 2) (right).

burden. Again, elder individuals have a higher PCDD/F body burden than young ones. So even in 2030 an age-dependency of the body burden is expected. Furthermore, for a given age class (18–29, 30–44, 45–59 and 60–79 years) the body burden decreases between 2009 and 2030, illustrating the effect of the long-term decreasing time trend of the PCDD/F exposure. For example, the mean body burden of individuals of 30 years of age in 2030 would be lower than the mean body burden of individuals of 30 years of age in 2009. Consequently, the cumulative risk related to PCDD/Fs and dl-PCBs is lower in 2030. Indeed, the probabilities that the body burden in 2030 exceeds the human reference BMDL for thyroid toxicity range from 0.3% (mean value, 97.5th percentile: 1.3%) in 18–29 years old to 1.5% (mean value, 97.5th percentile: 2.5%) in 60–79 year olds. In 18–29 year olds the probability for reproductive toxicity is negligible, whereas for 30–44 olds it is 0.7% (mean value, 97.5th percentile: 1.6%).

4. Discussion

The objective of this work was to develop an integrative approach to assess the toxic risk through a body burden approach for PCDD/Fs and dl-PCBs. The implementation of this approach needed to combine several methods and data. Firstly, the Kinetic dietary exposure model (KDEM) as proposed by Verger et al. (2007) was extended with variable exposure, including inter-individual and intra-individual variability. Furthermore, kinetically, the exposure was divided into PCDD/Fs, PCB-126 and remaining dl-PCBs. In concordance with Aylward et al. (2013) and Milbrath et al. (2009) PCDD/Fs were considered a homogeneous group. PCB 126 was singled out because of its high contribution to the overall TEQ exposure and is relative short half-life. Finally, the remaining dl-PCBs were, as the PCDD/Fs, kinetically considered as a homogeneous group. In order to gain insight in the long-term time trend of the dietary exposure to PCDD/Fs and dl-PCBs in France, a scaling function was calibrated using biomonitoring data of a specific French sub population i.e. fishermen families. This sub-population (i.e. the ICAR population) differs in exposure from the general French population (see Tables 1 and 2). Consequently, the scaling function reflects the historic correction of the background dietary exposure in the ICAR population, which may be assumed similar with that in the general French population, and a correction for the specific historic ICAR intake. However, when

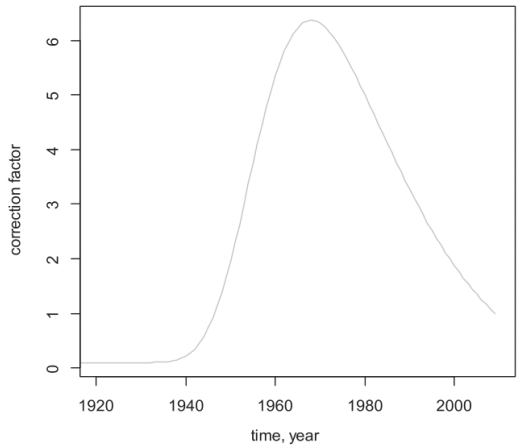


Fig. 2. Mathematical function ($f(y)$) scaling the PCDD/F dietary intake as determined in 2009 in France across time.

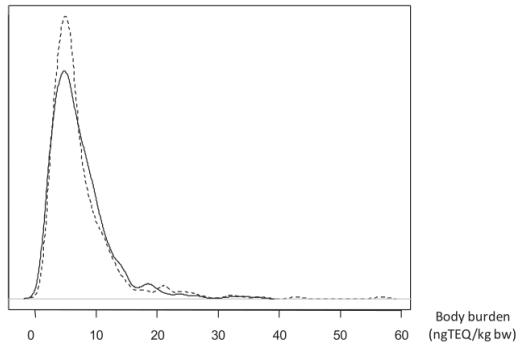


Fig. 3. Distributions of the measured (solid line) and predicted (dashed line) PCDD/Fs and dl-PCBs body burdens of the ICAR population (Anses/InVS, 2011) in ng TEQ/kg bw.

compared with other sources a value of 6 is quite acceptable. Indeed, Van der Molen (1998) mentions 6.7 for PCDD/F in 1966 in The Netherlands, whereas Lorber (2002) mentions 6 in this year for the TEQ in the US. From our modeling, $f(y) = 1$ for 1947 and 2009. This result can be compared to $f(y) = 1$ for 1940 and 1988 in Van der Molen (1998) and $f(y) = 1$ in 1930 and 1990 in Lorber (2002). In conclusion, applying a maximal correction of 6 as depicted in Fig. 2 on the general French population is in concordance with findings by others for the general Dutch and US population. Regarding dl-PCBs, Consonni et al. (2012) found no significant decrease in exposure notably for PCB 126. The KDEM model implemented without scaling function for exposure of PCB 126 and remaining dl-PCBs indeed fitted well with the observed body burdens of the fishermen population. These results, furthermore, are in concordance with the findings of Fernandes et al. (2004) who highlighted that contribution of dl-PCBs to total TEQ has increased between the 1980s and today. The model was validated using the biomonitoring data of the fishermen families. However, to strengthen the validation, the estimated body burdens can be compared with the levels measured in previous French biomonitoring studies. First, Frery and Deloraine (2000) report that breast milk from French women (18–44 years old) in 1998/1999 contains a mean level of 16.5 pg TEQ/g lipid for dioxins and furans only. Considering a default relative fat body of 20%, it corresponds to an approximate body burden of 3.3 ng TEQ/kg bw. Our modeling resulted in an estimated body burden of 1.4–1.8 (18–44 years) in 2009 for PCDD/Fs and a twofold factor between 1998 and 2009. Consequently our results seem consistent with the results of this biomonitoring study. Second, the French dioxin and incinerators study (InVS, 2009) reports a median level in blood of 18.7 pg TEQ/g MG for PCDD/Fs and dl-PCBs for the population of interest (individuals living around municipal solid waste incinerators and individuals living further away) in 2005. This corresponds to an approximate body burden of 3.7 ng TEQ/kg bw. In our work, the body burden of the French population in 2009 was estimated between 2.1 and 4.6, which is thus consistent with the results of this biomonitoring study.

Individual trajectories were simulated in order to take into account the intra- and inter-individual variability in exposure and accumulation kinetics. Some uncertainties remain over the estimation of the body burdens due to the way the KDEM trajectories were simulated and especially regarding childhood. Additional simulations based on different values of the body burden at six months of age show that this body burden can significantly influence the trajectories up to the age of around 20 years. Therefore, before 20 years, the estimations are dependent on the initial body burden. Consequently, the model as proposed in this work cannot be used to assess the risk focusing on the children.

The risk assessment presented here was based on a comparison of body burdens in the general French population with reference values for reproductive, hepatic and thyroid toxicity. Here it should be kept in mind that the assumptions made have added to the "worst case" character of the risk assessment:

- The use of a default relative fat body of 20% in deriving the human BMDLs for reproductive, thyroid and hepatic toxicity may have led to an overestimation of the human risk for these effects. For example, using 30% instead would have led to a 1.5 times lower risk.

Table 3
Distribution of the human BMDLs and probability of the general French population with a body burden exceeding the BMDLs (arithmetic mean and 95% confidence interval). The mean values are in bold.

Effects	BMDL	Classes of age (years)	2009 (%)	2030 (%)
Thyroïd toxicity	**39** [10 ; 146]	18–29	**2.8** [1.5 ; 4.9]	**0.3** [0 ; 1.3]
		30–44	**3.1** [1.9 ; 5.2]	**0.5** [0 ; 1.1]
		45–59	**3.5** [2.2 ; 7.3]	**0.7** [0 ; 1.4]
		60–79	**3.9** [2.7 ; 7.1]	**1.5** [0.3 ; 2.5]
Reproductive toxicity	**34** [8 ; 125]	18–29	**3.1** [1.4 ; 5.0]	**0** [0 ; 0]
		30–44	**3.5** [2.2 ; 5.2]	**0.7** [0 ; 1.6]
Hepatic toxicity	**302** [72 ; 1160]	18–29	**0** [0 ; 0]	**0** [0 ; 0]
		30–44	**0** [0 ; 0]	**0** [0 ; 0]
		45–59	**0** [0 ; 0]	**0** [0 ; 0]
		60–79	**0** [0 ; 0]	**0** [0 ; 0]

- In this study a generic kinetic half-life was attributed to PCDD/Fs and dl-PCBs other than PCB 126. It might be argued that congener specific half-lived for these compounds might have been used instead. However, doing so would have led to a 1.5 times lower risk (data not shown).
- Human toxic risk was defined as the state where a human body burden exceeds set human BMDL reference values for reproductive, thyroid or hepatic toxicity. With regard to reproductive toxicity in animals this effect has a critical window of one single dose at one specific day during pregnancy, i.e. GD 15. Assuming such a narrow window in humans too is "worst case", considering that the critical human window would comprise a period beyond one day. Furthermore, the risk assessment assumes all women of reproductive age to become pregnant. Similar arguments hold even more for thyroid and hepatic toxicity. Animals experiment suggests the critical time window for these effects to be well beyond one day. However, as even in animals the actual time-span of these windows are unknown one has to fall back on the "worst case" assumption made, i.e. a time-window of one single day.

Furthermore, the used risk indicators (human BMDL values) were obtained by extrapolating animal toxicity to man. Alternatively US EPA has recently identified preliminary reference values for exactly the same endpoints, however based on the interpretation of epidemiological data (US EPA, 2010, see Supplementary Material S4). In the case of thyroid toxicity this study used life-time exposure as the critical time window, the life-time body burden as the dose metric associated with toxicity and a (mean) body burden of 39 ng/kg bw as the reference value for the dose metric. US EPA used the maternal body burden during pregnancy as the dose metric/critical time window and 47 ng/kg bw as the reference value. Using the US EPA's criteria instead would have led to a lower percentage (0.7%) of the general French population with a body burden exceeding the reference value during reproductive age. For reproductive toxicity the corresponding benchmarks were: maternal body burden during pregnancy and a (mean) of 34 ng/kg bw (this study) vs. the body burden up to ten years of age and <23 ng/kg bw (US EPA). Unfortunately, the uncertainty (for <23 ng/kg bw) in US EPA's critical body burden for reproductive toxicity does not allow for a comparison with the results of this work. In a follow-up study we are currently studying the appropriateness of the epidemiological data for the methodology applied in this study, i.e. BMD analysis and intra-species extrapolation of toxicity to derive reference values for a human toxicity body burden. It should also be stressed that the effects used in this study are to be considered as subclinical indicators for overt toxicity such as hepatic failure, infertility or long term thyroid dysfunction.

In conclusion, this work combines various data sources (occurrence data, consumption data, biomonitoring data and toxicological data) with several methods (kinetic exposure model, Bayesian inferences, Bench Mark Dose modeling) to arrive at an integrative risk assessment for persistent chemicals. The advantages of the proposed approach are as follows. First, it provides an estimation of the body burden at the population level. Certainly such estimation may be obtained from biomonitoring data as well. However, these studies are costly and require complex protocol management and are usually confined to one point in time. Despite this, some biomonitoring programs could be implemented on long period of time as done in the US by the Centers for Disease Control and Prevention (CDC) and in this case, provide time trend in body burdens. Such programs remain rare and the modeling strategy allows the estimation of an individual's lifetime dioxin body burden for any year of birth, thereby enabling the development of targeted monitoring campaigns. Furthermore, this methodology can generically be applied to any population for which consumption habits are available, even for situations with absent historic dioxin exposure (situations corresponding with the function $f(y)$ set at 1). Moreover, such an approach also allows for the development of specific monitoring scenarios of external exposure and consequences on body burden levels and associated risk. Typical areas of application here are the evaluation of the effectiveness of risk management policy, such as the setting of food standards and its effect on the long term exposure. As an example, we predicted the body burden of the French population in 2030 assuming that the dietary exposure will remain stable at the 2009 level. It results that risk related to PCDD/Fs and dl-PCBs would be lower in 2030 but also that congener profiles would continue to change in France. Fernandes et al. (2004) notices for the UK that the relative contribution of the dl-PCBs to the total TEQ exposure increases over time and our results show that this also holds for France. Finally, the proposed modeling approach can be useful to determine the impact of a specific short-term, high level exposure excursing the background exposure on the body burden. Typical examples here are the consumption of accidental contaminated food items such as eggs or fish or specific consumption patterns such as the eating of contaminated freshwater fish.

Conflict of interest

RIVM-WHO Collaborating Centre Pathogens in Food and Water

This sponsor had no involvement in the study design; collection, analysis and interpretation of data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Acknowledgments

This work was supported by the WHO programme for the estimation the burden of foodborne diseases (FERG), ANSES and RIVM.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.yrtph.2014.07.004.

References

Anses/InVS, 2011. Etude nationale d'imprégnation aux polychlorobiphényles des consommateurs de poissons d'eau douce (ICAR-PCB). Rapport d'étude scientifique et synthèse. Maisons-Alfort, Saint-Maurice, France.

Aylward, L.L., Hays, S.M., 2002. Temporal trends in human TCDD body burden: decreases over three decades and implications for exposure levels. J. Expo. Anal. Environ. Epidemiol. 12 (5), 319–328.

Aylward, L., Brunet, R.C., Carrier, G., Hays, S.M., Cushing, C.A., Needham, L.L., Mocarelli, P., 2005. Concentration-dependent TCDD elimination kinetics in humans: toxicokinetic modeling for moderately to highly exposed adults from Seveso, Italy, and Vienna, Austria, and impact on dose estimates for the NIOSH cohort. J. Expo. Sci. Environ. Epidemiol. 15 (1), 51–65.

Aylward, L.L., Collins, J.J., Bodner, K., Wilken, M., Bodnar, C.M., 2013. Elimination rates of dioxin congeners in former chlorophenol workers from Midland, Michigan. Environ. Health Perspect. 121 (1), 39.

Consonni, D., Sindaco, R., Bertazzi, P.A., 2012. Blood levels of dioxins, furans, dioxin-like PCBs, and TEQs in general populations: a review, 1989–2010. Environ. Int. 44, 151–162.

De Mul, A., Bakker, M.I., Zeilmaker, M.J., Traag, W.A., Leeuwen, S.P.J., Hoogenboom, R.L.A.P., 2008. Dietary exposure to dioxins and dioxin-like PCBs in The Netherlands anno 2004. Regul. Toxicol. Pharmacol. 51, 278–287.

Deurenberg, P., Weststrate, J.A., Seidell, J.C., 1991. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br. J. Nutr. 65 (2), 105–114.

Dubuisson, C., Lioret, S., Touvier, M., Dufour, A., Calamassi-Tran, G., Volatier, J.L., 2010. Trends in food and nutritional intakes of French adults from 1999 to 2007: results from the INCA surveys. Br. J. Nutr. 103, 1035–1048.

EFSA, Scientific Opinion on Polybrominated Diphenyl Ethers (PBDEs) in Food, EFSA Panel on Contaminants in the Food Chain (CONTAM), EFSA J., 2011; 9(5):2156.

Frery, N., Deloraine, A., 2000. Étude sur les dioxins et les furanes dans le lait maternel en France, Institut de Vieille Sanitaire, 12, Rue du Val d'Osne, Saint-Maurice Cedex, France, ISBN 2-11-091490-4.

Emond, C., Michalek, J.E., Birnbaum, L.S., DeVito, M.J., 2005. Comparison of the use of a physiologically based pharmacokinetic model and a classical pharmacokinetic model for dioxin exposure assessments. Environ. Health Perspect. 113 (12), 1666.

Fernandes, A., Gallani, B., Gem, M., White, S., Rose, M., 2004. Trends in the dioxin and PCB content of the UK diet. Organohalogen Compd. 66, 2053–2060.

Fromme, H., Albrecht, M., Boehmer, S., Büchner, K., Mayer, R., Liebl, B., Bolte, G., 2009. Intake and body burden of dioxin-like compounds in Germany: the INES study. Chemosphere 76 (11), 1457–1463.

Gies, A., Neumeier, G., Rappolder, M., Konietzka, R., 2007. Risk assessment of dioxins and dioxin-like PCBs in food – comments by the German Federal Environmental Agency. Chemosphere 67 (9), S344–S349.

Hays, S.M., Aylward, L.L., 2003. Dioxin risks in perspective: past, present, and future. Regul. Toxicol. Pharmacol. 37 (2), 202–217.

Hsu, J.F., Chang, Y.C., Liao, P.C., 2010. Age-dependent congener profiles of polychlorinated dibenzo-p-dioxins and dibenzofurans in the general population of Taiwan. Chemosphere 81 (4), 469–477.

InVS, 2009. (French Institute for Public Health Surveillance). Etude d'imprégnation par les dioxines des populations vivant à proximité d'usines d'incinération d'ordures ménagères (The French dioxin and incinerators study). Available online at: <http://www.invs.sante.fr/publications/2009/impregnation_dioxines_uiom/>.

JECFA, 2001. Joint FAO/WHO Expert Committee on Food Additives (JECFA), Summary and Conclusions of the fifty-seventh. JECFA, Rome.

Leblanc, J.C., 2006. CALIPSO, fish and seafood consumption study and biomarker of exposure to trace elements, pollutants and omega3.

Lioret, S., Dubuisson, C., Dufour, A., Touvier, M., Calamassi-Tran, G., Maire, B., Lafay, L., 2010. Trends in food intake in French children from 1999 to 2007: results from the INCA (étude Individuelle Nationale des Consommations Alimentaires) dietary surveys. Br. J. Nutr. 103 (04), 585–601.

Lorber, M., 2002. A pharmacokinetic model for estimating exposure of Americans to dioxin-like compounds in the past, present, and future. Sci. Total Environ. 288 (1–2), 81–95.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: evolution, critique and future directions (with discussion). Stat. Med. 28, 3049–3082.

Milbrath, M.O., Wenger, Y., Chang, C.W., Emond, C., Garabrant, D., Gillespie, B.W., et al., 2009. Apparent half-lives of dioxins, furans, and polychlorinated biphenyls as a function of age, body fat, smoking status, and breast-feeding. Environ. Health Perspect. 117 (3), 417–425.

National Research Council (NRC), 2006. Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment. Washington, DC: The National Academies Press.

Pinsky, P.F., Lorber, M., 1998. A model to evaluate past exposure to 2,3,7,8-TCDD. J. Expo. Anal. Environ. Epidemiol. 8 (2), 187–206.

SCF, 2001. Opinion of the Scientific Commission on Food on the Risk Assessment of Dioxins and Dioxin-like PCBs in Food, Update based on new scientific information available since the adoption of the SCF opinion of the 22th November 2000. CS/CNTM/Dioxin/20 final.

Sirot, V., Volatier, J.L., Calamassi-Tran, G., Dubuisson, C., Ménard, C., Dufour, A., Leblanc, J.C., 2009. Core food of the French food supply: second Total Diet Study. Food Addit. Contam. 26 (5), 623–639.

Sirot, V., Tard, A., Venisseau, A., Brosseaud, A., Marchand, P., Le Bizec, B., et al., 2012. Dietary exposure to polychlorinated dibenzo-p-dioxins, polychlorinated dibenzofurans and polychlorinated biphenyls of the French population: results of the second French Total Diet Study. Chemosphere 88 (5), 492–500.

Tard, A., Gallotti, S., Leblanc, J.C., Volatier, J.L., 2007. Dioxins, furans and dioxin-like PCBs: occurrence in food and dietary intake in France. Food Addit. Contam. 24 (9), 1007–1017.

U.S. Environmental Protection Agency (U.S. EPA) (2010). EPA's Reanalysis of Key Issues Related to Dioxin Toxicity and Response to NAS Comments (external review draft). US Environmental Protection Agency, Washington, EPA/600/R-10/038A.

Van den Berg, M., Birnbaum, L.S., Denison, M., De Vito, M., Farland, W., Feeley, M., et al., 2006. The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicol. Sci. 93 (2), 223–241.

Van der Molen, G.W., Kooijman, S.A.L.M., Slob, W., 1996. A generic toxicokinetic model for persistent lipophilic compounds in humans: an application to TCDD. Fundam. Appl. Toxicol. 31 (1), 83–94.

Van der Molen, G.W., 1998. A physiologically-based mathematical model for the long-term kinetics of dioxins and furans in humans (Ph.D.), Free University, Amsterdam.

Van Leeuwen, F.X.R., Younes, M.M., 2000. Consultation on assessment of the health risk of dioxins: re-evaluation of the tolerable daily intake (TDI): executive summary. Food Addit. Contam. 17 (4), 223–240.

Van Der Voet, H., Slob, W., 2007. Integration of probabilistic exposure assessment and probabilistic hazard characterization. Risk Anal. 27 (2), 351–371.

Verger, P., Tressou, J., Clémençon, S., 2007. Integration of time as a description parameter in risk characterisation: application to methyl mercury. Regul. Toxicol. Pharmacol. 49 (1), 25–30.

WHO (World Health Organization), 2002. Polychlorinated dibenzodioxins, polychlorinated dibenzofurans, and coplanar polychlorinated biphenyls. Safety Evaluation of Certain Food Additives and Contaminants. WHO Food Additives Series 48. Joint FAO/WHO Expert Committee on Food Additives, Geneva, Switzerland.

WHO, 2005. Harmonization project document no. 2: Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration-response assessment. WHO, IPCS. Geneva, Switzerland.

Winter, R., Bakker, M.I., Baumann, R.A., Hoogerbrugge, R., Zeilmaker, M.J., 2003. Exposure assessment of Dutch nursing infants to brominated flame retardants. (RIVM report 310305002/2003).

Zeilmaker, M.J., Houweling, D.A., Cuijpers, C.E.J., Hoogerbrugge, R., Baumann, R.A., 2002. Contamination of breast milk with chlorinated hydrocarbons in the Netherlands: Levels in 1998 and time-trend. (RIVM report 529102012/2002).

2.3.1 Summary of the objectives, methodological developments, main results and limitations of the work

As for the two previous works, the main objective of this work is to link external exposure and biomonitoring data in order to interpret it and to improve the risk assessment for persistent chemicals. This work focuses on urine cadmium concentrations and different sources of exposure are considered: exposure from food and cigarette smoke.

A dynamic modeling of exposure is also used to take into account the evolution of the exposure through the life of the individuals. The ABC approach proposed in the first article is implemented using same summary statistics and distances but an ABC-SMC algorithm was used instead of the ABC-MCMC algorithm in order to speed the time of computation (code of the algorithm in R language is provided in SM). In contrast to previous work, a multi-compartments physiologically-based toxicokinetic model (PBTK) is used. Such model allows to take the two different pathways of exposure into account but also to describe more accurately the process of the ADME. A hierarchical model was proposed to properly model inter-individual variability in physiological parameters, in exposure and in smoking habits.

Such approach succeeded to explain current urine cadmium concentrations from past and current dietary and cigarette smoke exposure. A toxicodynamic part that links internal exposure measured through urine cadmium concentrations to kidney effects was added and related risk was assessed for French men and women according to their age. A comparison of risk assessment from external exposure and internal exposure is also proposed for current and future exposure.

Some limitations remain on the proposed modeling, and among them, there are still some sources of inter-individual variability that were not modeled. Finally, such results that partly aim at predicting changes in exposure that may have occurred could be strengthen or invalidated with biomonitoring data collected during a long period of time (if available).

2.3.2 Publication

Toxicology and Applied Pharmacology 279 (2014) 364–372

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

PBPK and population modelling to interpret urine cadmium concentrations of the French population

Camille Béchaux [a,*], Laurent Bodin [a], Stéphan Clémençon [b], Amélie Crépet [a]

[a] ANSES, French Agency for Food, Environmental and Occupational Health Safety, 27-31 Avenue du Général Leclerc, 94701 Maisons-Alfort, France
[b] Telecom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France

ARTICLE INFO

Article history:
Received 8 May 2014
Revised 18 June 2014
Accepted 22 June 2014
Available online 3 July 2014

Keywords:
Bayesian population model
Biomonitoring
Cadmium
PBPK
Risk assessment

ABSTRACT

As cadmium accumulates mainly in kidney, urinary concentrations are considered as relevant data to assess the risk related to cadmium. The French Nutrition and Health Survey (ENNS) recorded the concentration of cadmium in the urine of the French population. However, as with all biomonitoring data, it needs to be linked to external exposure for it to be interpreted in term of sources of exposure and for risk management purposes. The objective of this work is thus to interpret the cadmium biomonitoring data of the French population in terms of dietary and cigarette smoke exposures. Dietary and smoking habits recorded in the ENNS study were combined with contamination levels in food and cigarettes to assess individual exposures. A PBPK model was used in a Bayesian population model to link this external exposure with the measured urinary concentrations. In this model, the level of the past exposure was corrected thanks to a scaling function which account for a trend in the French dietary exposure. It resulted in a modelling which was able to explain the current urinary concentrations measured in the French population through current and past exposure levels. Risk related to cadmium exposure in the general French population was then assessed from external and internal critical values corresponding to kidney effects. The model was also applied to predict the possible urinary concentrations of the French population in 2030 assuming there will be no more changes in the exposures levels. This scenario leads to significantly lower concentrations and consequently lower related risk.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Cadmium is a toxic metal found widely distributed in the environment, through industrial and agricultural sources (Nordberg et al., 2007). Two major sources of cadmium exposure are identified for the general population: foodstuffs and cigarette smoke (Järup and Åkesson, 2009). Cadmium is now recognised as a risk factor for a broad spectrum of health conditions, including cardiovascular, kidney, and bone disease (Järup and Åkesson, 2009; Nordberg et al., 2007; Satarug et al., 2010). Since cadmium presents a very long biological half-life and accumulates mainly in kidney, urine concentration can be used as a marker of long-term cadmium exposure (Nordberg et al., 2007). Therefore, urinary cadmium concentrations are commonly used as a surrogate for the body burden in health risk assessment (Amzal et al., 2009). The French Nutrition and Health Survey (ENNS) which was carried out by the French Institute for Public Health Surveillance in 2006–2007 (Castetbon et al., 2009) recorded the concentration of cadmium in urine of 1900 adults. This data constitutes a relevant source of information to conduct the risk assessment for cadmium in France. However, as highlighted in Lyons et al. (2008) and Hays et al. (2007), to understand the effects of exposure to environmental chemicals on public health requires establishing relationships among events along an exposure-health evaluation-risk assessment continuum. Therefore, biomonitoring data which is a measure of internal exposure representing one event along the continuum has to be linked with external exposure and a biologically effective dose. The reverse dosimetry approach employs kinetic models to estimate the intake dose or external environmental concentration based on a measured tissue concentration. Indeed, most of the time biomonitoring data are reported without corresponding external exposure data, leaving the relationship between internal and external exposure as one to be determined (Lyons et al., 2008). In the ENNS survey no external exposure data were measured but dietary and smoking habits of the individuals were recorded. Focusing on cadmium, Amzal et al. (2009), Ruiz et al. (2010) highlighted that dietary intake of cadmium need to be related to the internal dose in order to perform a reliable and comprehensive assessment of the health risks associated with long-term exposure to cadmium in food. In these previous works, exposure by inhalation of cigarette smoke was not taken into account whereas it has been shown as an important source of exposure in the general population (EFSA, 2009a; Järup and Akesson, 2009). Moreover, few exposure studies conducted in France in the 1990s highlight a trend in dietary exposure to cadmium

* Corresponding author at: ANSES, 27-31 Avenue du Général Leclerc, 94701 Maisons-Alfort, France. Fax: +33 1 49 77 38 92.
E-mail address: Camille.bechaux@anses.fr (C. Béchaux).

http://dx.doi.org/10.1016/j.taap.2014.06.026
0041-008X/© 2014 Elsevier Inc. All rights reserved.

(Decloître, 1998; DGS, 1992; Leblanc et al., 2000). From the results of these studies, the dietary exposure to cadmium in France in 1992, 1998 and 2000 are higher than the current one. This trend in exposure is consistent with the data on the production and use of cadmium in France which also shows a decline since the late 1970s and can also be related to the successive regulations. Because of the long half-life of cadmium, urinary concentrations measured today also reflect the past exposure of the individuals. Therefore the interpretation of this data through exposure needs to take this information into account. The objective of this work is thus to interpret the biomonitoring data of the French population in terms of dietary and cigarette smoke exposures by performing a reverse-dosimetry approach which takes these trends in exposure into account.

A cadmium PBPK model is used to link dietary and cigarette smoke exposures to biomonitoring data, as made in Ruiz et al. (2010) to interpret the National Health and Nutrition Examination Surveys (NHANES) through dietary exposure. A Bayesian population-based model is implemented to account for the variability in the population, both in the modelling of the disposition of the chemical in the body, and in the description of the exposure conditions (Lyons et al., 2008). In contrast to previous work that reconstructed the external exposure from the internal dose without any quantitative information about external exposure, in this work we proposed to input the individual intakes of the ENNS population in the Bayesian model. Moreover, in order to take the exposure trends into account, we added a function to the model which corrects the dietary intakes according to the date they occurred. As proposed in Béchaux et al. (2014) for the assessment of the past exposure to dioxin using biomonitoring data, parameters of model are updated thanks to the information provided by the urinary cadmium concentrations. Inferences about the parameters are performed from an Approximate Bayesian Computation method (ABC, Beaumont et al., 2002) with an ABC SMC algorithm (Toni et al., 2009) to speed the time of computation. Finally, a risk assessment is conducted from internal and external critical values corresponding to a benchmark dose lower confidence limit for a 5% increase of the prevalence of elevated urinary beta-2-microglobulin (EFSA, 2009b).

Material and method

Health and nutrition examination surveys

The French Nutrition and Health Survey (ENNS) was carried out by the French Institute for Public Health Surveillance in 2006–2007 (Castetbon et al., 2009). The main objectives of this survey were to describe food consumption, nutritional status, and physical activity in the general population in France (adults and children) and to study nutritional and environmental biomarkers. ENNS measured the levels of various different metals (including lead, cadmium, mercury and arsenic) and pesticides (organophosphates, organochlorines and pyrethroids) in blood, urine and hair. It should be noted that this study aimed at investigating the exposure and nutritional status of the general population. Consequently exposure of certain population sub-groups that may have higher exposure to cadmium such as vegetarians (Berglund at al., 1994) or high consumers of crustaceans, molluscs and cephalopods (Sirot et al., 2008) were not specifically investigated in this study.

In this context, the concentration of cadmium in urine was measured for 1900 adults between 18 and 75 years. Apart from two values that were below detection limit and replaced by half the limit of detection, all the values were quantified, with the limits of detection and quantification respectively equal to 0.004 g/L and 0.013 g/L (Fréry et al., 2011). The concentrations of cadmium in urine of the 1900 individuals are summarised and described by gender, age and smoking status in Table 1. The long biological half-life of cadmium in kidneys (around 10–30 years) explains the observed differences in measured urinary concentrations between age classes. Tobacco leaves are known to accumulate cadmium and one cigarette may contain between 1 and 2 μg cadmium depending on the type and brand (EFSA, 2009b; Järup and Akesson, 2009). Consequently, tobacco smoking is one important source of cadmium exposure which explains the difference in urinary concentrations observed between smokers and non-smokers. Women have higher cadmium urine concentration levels than men and this significant difference between genders is taken into account in the kinetic modelling.

Dietary, smoking habits and contamination data

The food consumption of the 1900 individuals were estimated through three 24 h recalls, randomly distributed within a period of 2 weeks and one of which was on the weekend. Amounts consumed were described using a validated photography manual or according to standard measurements (home containers, grams indicated on the package, etc.) and data were entered on software developed for the study that included more than 3400 foods and beverages. Individuals were asked to fill a questionnaire about their smoking habits. The number of cigarettes smoked on average per day and per week and the age they started and stopped smoking were asked. Anthropometry was measured using identical devices and according to standardised procedures adapted from WHO recommendations.

Contamination data of 212 core foods was provided by the Second French Total Diet Study (TDS2, Arnich et al., 2012). The 212 core foods cover about 90% of the whole diet of adults and children estimated from Second French dietary Survey (INCA2, Dubuisson et al., 2010). To take into account the potential regional and seasonal variations in occurrence, the sampling was performed in eight great metropolitan regions and each food collected in a region was sampled during two different seasons. In this work, we consider the national contamination which is estimated as the mean of the contamination levels of each region for the two seasons.

Some food items recorded in the ENNS were grouped together in order to match with the 212 food items of the TDS2 and the sum of the consumption of the foods forming a group was considered.

External exposure assessment

Dietary exposure. Dietary exposure U_i of an individual i (pg kg^{-1} day^{-1}) is estimated by combining data on food consumption (g day^{-1}) with data on cadmium concentration (pg g^{-1}) in each food. The intake U_i is therefore the sum of several intakes relative to the P products containing cadmium. Each intake is calculated by multiplying the total cadmium concentration C_p of a product p by the quantity consumed $Q_{i,p}$ of the product p. The sum of the P intakes is then divided by the body weight of the individual w_i (kg)

$$U_i = \frac{\sum_{p=1}^{P} C_p Q_{i,p}}{w_i}. \quad (1)$$

Table 1
Distribution of the urinary cadmium concentration of the French population (μg/g creatinine) according to the gender, the age and the smoking status (data from ENNS study, Fréry et al., 2011).

	5th p	50th p	Mean	95th p
Women	0.13	0.33	0.41	0.96
Men	0.09	0.25	0.33	0.81
18–39 years old	0.09	0.21	0.25	0.55
40–59 years old	0.13	0.33	0.41	0.93
60–74 years old	0.16	0.41	0.50	1.13
Non smokers	0.10	0.29	0.35	0.79
smokers	0.12	0.31	0.41	1.00

Exposure from cigarette smoke. Exposure from cigarette smoke S_i of an individual i (pg kg^{-1} day^{-1}) is estimated by combining data on number of cigarettes smoked on average per day Q'_i (cigarette day^{-1}) with an assumed mean cadmium content per cigarette C' of 1.5 μg (EFSA, 2009b). This intake is then divided by the body weight of the individual w_i (kg)

$$S_i = \frac{C'Q'_i}{w_i} \tag{2}$$

Cadmium PBPK model

We used the model proposed by Kjellstrom and Norberg (1978) with parameter values already used by Ruiz et al. (2010) to interpret urinary concentrations of cadmium in the US biomonitoring study NHANES. This model considers two intake routes: inhalation and oral, amount and flow between 8 compartiments: lung, intestine, 3 blood compartments (plasma (B1), erythocytes (B2) and mettalothioneine (B3)), liver, kidney and other tissues and two excretion routes: feces and urine. This generic model was validated in Ruiz et al. (2010) by comparing simulations to other published model simulations and to human data sets (Choudhury et al., 2001; Diamond et al., 2003; Hays et al., 2008). Human physiological and chemical-specific parameters describing the absorption, distribution, and blood and tissue partitioning were taken from Kjellstrom and Norberg (1978) and Ruiz et al. (2010) and provided in Table 2.

As highlighted by Clewell and Clewell (2008), applied to a population, PBPK modelling should take into account inter-individual variability when known. Bois et al., 2010 detailed the many possible sources of inter-individual variability. To take inter-individual variability into account in our PBPK modelling, we used distributions instead of a mean value for the parameters of the PBPK. Information about the ranges of values for parameters of this model was provided by Kjellstrom and Norberg (1978) and Ruiz et al. (2010). To translate this information, we used triangular distributions where the lower and upper limits are the limits values of the range and mode is equal to the empirical fitting value. The variability related to the modification of the body weight and the renal function with the age is also included in the model as it was proposed in the model implemented in Berkeley Madonna (Ruiz et al., 2010).

Regarding inter-individual variability, the significant role of the gender has been highlighted in previous studies and in the data and need to be taken into account in the modelling. The main explanation retained in the literature for this observed difference is an increased intestinal absorption of dietary cadmium for women which may be caused by lower body iron store (Choudhury et al., 2001; Järup and Akesson, 2009; Kippler et al., 2007). Choudhury et al. (2001) found a 2-fold factor in urinary concentration between adult women and men in the US population and consequently proposed an absorption factor of 0.1 for female and 0.05 for male. However, in the French population this factor is less high. Indeed, the impact of the gender on the urinary concentration value corrected for the intake value, the age and the smoking status, is estimated to be +0.11 μg/g creatinine for women (from a general linear model) which corresponds to a factor of 1.35. This is consistent with Ruiz et al., 2010 who found a factor of 1.4 for all age groups and 1.6 for 40–59-year-old women. Therefore, regarding the data, an assumed absorption factor of 0.1 for female compared to 0.05 to male seems to be too high. This observation is strengthen by the results of Ruiz et al., 2010 who used an absorption factor of 0.05 for males and 0.1 for females and showed that their model predictions were better for males than for females. Therefore, due to the uncertainty around the absorption value for female, we propose to use a distribution for the absorption factor of female. As for the other parameters, using a distribution instead of a single value also allows to take the variability between women into account. To stay in line with the other parameters modelling, we propose to use a triangular distribution where the lower limit is 0.05, the upper limit is 0.1 and the mode is equal to 0.075.

Bayesian population model and trend of exposure

We apply a Bayesian population model to link external exposure of the French population from the internal dose provided by the urinary concentrations data. In our model, individual values for current dietary exposure U_i and exposure from smoke cigarette S_i estimated from individual consumption $Q_{i,p}$ and Q'_i and mean contamination values C_p and C' are considered. As made in Béchaux et al. (2014) for dioxins, we introduce to the model a function λ characterised by a vector of parameters β which corrects the dietary intakes according to the date they occurred. Therefore, the intake G_i of individual i occurring at the year y is given by $G_i = U_i \times \lambda(y)$.

The shape of the exposure evolution is unknown. Therefore, the function λ must be represented in a flexible manner. That is why we choose a linear combination of functions from a B-spline basis of degree m. The shape of λ can be described with varying degrees of accuracy m. In this work, we choose to test two different models: one very simple with 3 parameters and one more complicated with 4 parameters which allow for an inflexion in the shape of λ. Consequently, a model selection will be performed to choose the best form for the scaling function λ. Information from other studies are often used in Bayesian approaches to fix appropriate prior distribution for the parameters. However, to the best of our knowledge, very little information is available regarding the evolution of exposure to cadmium in France or in Europe. A techno-economic report on cadmium in France made by the French National Institute for Industrial Environment and Risks (Pilchard et al., 2005) shows that the production of cadmium in France was divided by 2 since the 1980s. From this information, we fixed informative but large prior distribution for the parameters of the function λ. Resulting a priori mean, 95th and 5th percentiles values for λ are shown in Fig. 1. Considering that the evolution of the dietary exposure has been the same for all individuals, parameters of the function lambda were only estimated at the population level. Individual values for each parameter of the PBPK model $\theta_{PBPK,i}$ are considered as arising independently from a population distribution parameterised by a mode c_θ, lower limit a_θ and upper limit b_θ. Individual data from the ENNS survey $\theta_{indiv,,i}$ were also including in the model such as the age of the individual, his weight, the year he started to smoke, the year he stopped. The hierarchical model showing the conditional dependencies among the population and the individual parameters and the link through the deterministic PBPK model to the urinary concentrations data is summarised by a graphical model (Fig. 2). To stay in line with the previous works, we used the visual model of Lyons et al. (2008) adapted from Bernillon and Bois (2000).

Approximate Bayesian Computation method and implementation

To determine the posterior distribution of the parameters β of the function lambda with the Bayesien model, we implemented an Approximate Bayesian Computation method (ABC, Beaumont et al., 2002). Such approach has recently become popular for statistical inference in complex models in population-genetics (Csillery et al., 2010) or in mathematical epidemiology (Blum and Chi Tran, 2010). It has also been shown suitable for reverse dosimetry problem. Indeed, Béchaux et al. (2014) proposed an ABC approach to reconstruct the past exposure of the French population to dioxin from biomonitoring data. This approach bypasses the exact evaluation of the likelihood function to estimate the posterior distributions by means of simulations and well-chosen summary statistics. The principle of standard ABC relies on a basic rejection algorithm: candidate values β^* are generated from their prior distribution, then from each β^*, urinary concentrations are generated with the PBPK model. Simulated and observed urinary

Table 2
Model parameters and physiologic parameters of the PBPK model for cadmium from Kjellström and Nordberg (1978) and Ruiz et al. (2010). * indicates that the value of the parameter was proposed in this work.

Model parameter	Values fitting to empirical data	Lower limit of assumed range	Upper limit of assumed range	Unit
C1 = fraction of inhalation intake deposited in NP and TB regions	0.1	0.1	0.2	
C2 = fraction of inhalation intake deposited in pulmonary region	0.4	0.4	0.6	
C3 = rate of transfer of inhalation intake to uptake pool	0.05	0.01	1	Day^{-1}
C4 = rate of transfer of air cadmium to the GI-tract	0.005	0.001	0.1	Day^{-1}
C5 = fraction of GI-tract cadmium retained in epithelium	0.05 for men	0.03	0.1	
	0.075 for women*	0.05	0.1	
C6 = rate of transfer of cadmium in the GI-tract to uptake pool	0.05			Day^{-1}
C7 = fraction of uptake pool transferred to blood pool B4	0.25	0.2	0.4	
C8 = maximum mass of cadmium in uptake pool	1	0.5	5	ug
C9 = rate of transfer from blood pool B1 to other tissues	0.44	0.4	0.8	
C10 = rate of transfer of cadmium in other tissues to blood pool B1	0.00014	0.00004	0.0002	Day^{-1}
C11 = rate of transfer of cadmium in blood pool B1 to feces	0.27	0.05	0.5	Day^{-1}
C12 = rate of transfer of cadmium in blood pool B1 to liver,	0.25	0.1	0.4	Day^{-1}
C13 = rate of transfer of cadmium in liver to blood pool B1	0.00003	0	0.0001	Day^{-1}
C14 = rate of transfer of cadmium in liver to blood pool B3	0.00016	0.0001	0.0003	Day^{-1}
C15 = rate of transfer of cadmium in liver to feces	0.00005	0	0.0001	Day^{-1}
C16 = rate of transfer of cadmium in blood pool B2 to blood pool B3	0.012	0.004	0.015	Day^{-1}
C17 = rate of transfer of cadmium in blood pool B3 to kidney	0.95	0.8	0.98	Day^{-1}
C18 = rate of transfer of cadmium in kidney to blood pool B1	0.00001	0	0.0001	Day^{-1}
C19 = age-specific rate constant for transfer of cadmium in kidney to urine	0.00014	0.00002	0.0002	
C20 = fraction of cadmium in blood pools B1 + B2 that contribute to cadmium in blood pool B4	0.1	0.05	0.5	
C21 = increased transfer of cadmium from kidney to urine after age 30	0.0000011	0	0.000002	Day^{-1}
CX = rate of transfer of cadmium in B1 to blood B2,	0.04	0.01	0.05	
Average liver weight	1500			g
Average blood volume	70			mL/kg
Average blood specific gravity	1.06			
Average daily urine excretion (adult)	1			L
Average daily urine excretion (aged)	0.9			L
Average daily urine excretion (child)	0.5			L

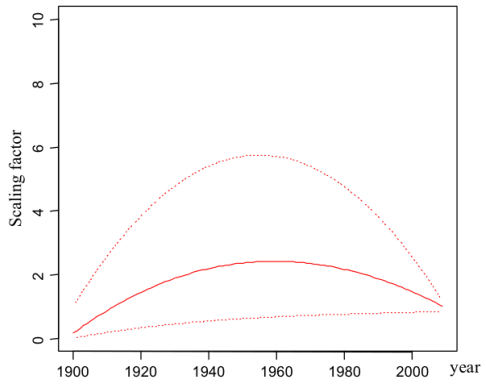


Fig. 1. Scaling function λ obtained from the prior distributions of the parameters (p5, Mean and p95).

concentrations are then compared according to a distance. If this distance is considered as small enough i.e. smaller than a previously fixed tolerance level, the accepted values for theta form the posterior distributions.

The choice of the distance between the simulated and observed urinary concentration is paramount in this method. This aspect has been widely studied in the case of population genetic data (Beaumont et al., 2002; Blum and Chi Tran, 2010) but regarding biomonitoring data, specific summary statistics must be defined. We proposed to build

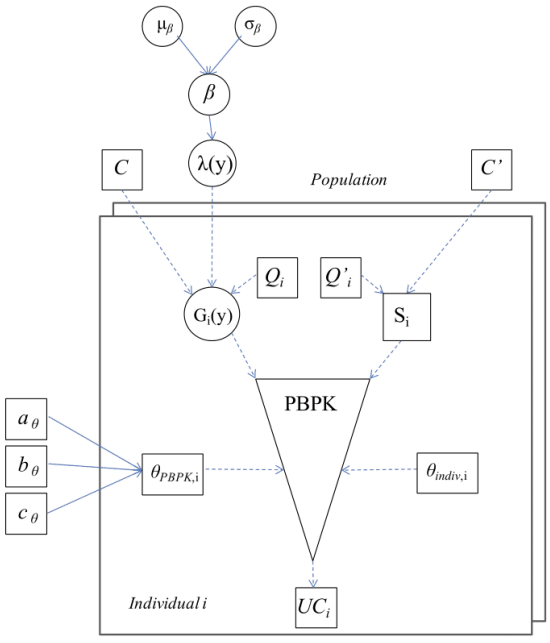


Fig. 2. Graphic model (adapted from Bernillon and Bois, 2000; Lyons et al., 2008). Circles represent unknown quantities to be updated via ABC inferences: means (μ) and variance (σ) of the parameters β that define the scaling function λ and therefore the individual dietary intake at year y $G_i(y)$. Squares represent the known quantities: PBPK model parameters (θ_{PBPK}) sampled from a triangular distribution (which parameters a_θ, b_θ and c_θ) individual parameters from ENNS study ($\theta_{indiv,i}$), exposure parameters (C,C′,Q and Q′) and measured urine concentrations (UC). The triangle represents the deterministic PBPK model. The solid arrows represent a stochastic link, and the dashed arrow represents a deterministic link.

summary statistics from distances between the observed and the simulated distributions of body burdens of different age classes of individuals. Various quantities can be considered to evaluate the discrepancy between two distributions and we choose the two-sample Cramer–Von Mises statistic which is a very popular measure of global discrepancy widely used in nonparametric hypothesis testing. In practice, the 1900 individuals are divided into 5 age classes with similar size. In the ABC algorithm, at each run and for each age class, the two-sample Cramer–Von Mises statistic between the distribution of the predicted urinary concentration and the observed ones is calculated. Then the sum of these 5 distances is used as the distance between simulations and observations (Béchaux et al., 2014).

In order to reduce the number of simulations required to estimate the posterior distribution, an ABC-SMC (Sequential Monte Carlo) algorithm was implemented is the software R. Indeed, Toni et al. (2009) showed that the sequential Monte Carlo method derived from a sequential importance sampling methods shows very good computational performance. The implementation this algorithm requires some choices regarding the algorithm parameters such as the number of particles *N*, the number of times the particles are perturbed *T* and the transition kernel *K*. Based on Toni et al. (2009) and Filippi et al. (2013) we choose $N = 1000$, $T = 5$ and a Gaussian transition kernel.

Since two different models are tested, a model selection is performed by calculating the Root Mean Square Error of Prediction (RMSEP) of each model as follow: the data set is randomly divided into 3 subsets respecting the proportion of individuals in each age class. Then the model is fitted from two-thirds of the data set and the RMSEP is calculated on the remaining third. This procedure is repeated three times and the mean of the RMSEP for both models is compared. Comparison with the few available data on past exposures in 1992, 1998 and 2000 (Decloître, 1998; DGS, 1992; Leblanc et al., 2000) will also participate to the validation of the model.

Results

External exposures assessed from individual habits and contamination data

By combining contamination data with the food consumption of each individual (Eq. (1)), the dietary exposure of the 1900 individuals is assessed. It results in a mean exposure of 0.15 μg/kg b.w/day, a 5th percentile of 0.07 and a 95th percentile of 0.27. For comparison, the mean exposure of the French population estimated in Arnich et al. (2012) from the same contamination data (i.e. TDS2) and consumption data from the Second French dietary Survey (INCA2, Dubuisson et al., 2010) was 0.16 μg/kg b.w/day with a 95th percentile of 0.27. Järup and Akesson (2009) reported that the average cadmium intake from food in world generally varies between 8 and 25 μg/day which correspond to 0.1 and 0.4 μg/kg b.w/day considering a mean body weight of 70 kg.

By combining data on number of cigarettes smoked on average per day by each individual with assumed mean cadmium content per cigarette of 1.5 μg (Eq. (2)), the exposure from cigarette smoke of the 1900 individuals is assessed. Regarding the smokers, the mean of exposure is 0.28 μg/kg b.w/day, the 5th percentile is 0.02 and the 95th percentile is 0.71. Consequently, for smokers, the intake from cigarette smoke can be higher than the one from food. This external exposure assessment does not take into account the higher pulmonary absorption compared to the gastrointestinal absorption and therefore the contribution from smoking would increase if the internal dose is considered.

Linking external exposures and urinary concentration

As a first attempt to link external exposures to measured urinary concentration, the PBPK model described in section 2.4 is used to predict urinary concentration from individual dietary and cigarette smoke intakes for each of the 1900 individuals. The distribution of the thus

predicted concentrations for the population is compared to the distribution of the observed concentrations in Fig. 3. The results show an underestimation of the urinary concentration in cadmium. As mentioned before, the current observed urinary concentration results from past exposure which are expected to have been higher than the current one.

In a second attempt, the model was implemented in order to take changes in exposure into account. Convergence of the algorithm was checked by comparing the difference of the successive *a posteriori* distributions of each parameter. The posterior shape of the scaling function λ resulting from the posterior distribution of the parameters β for the two models is shown in Fig. 4. Both models provide relatively closed scaling function which shows that the dietary exposure may have been 3 times higher in the past compared to the current exposure. Through Bayesian method, prior information (i.e. prior distribution on parameters) is combined with data to provide posterior estimate of the parameters. Consequently, the modelling provides a distribution for each estimated parameter values (i.e. the posterior distribution) instead of single value. This makes uncertainty about the way exposure have evolved available. This uncertainty is represented in Fig. 4 by a confidence interval which was calculated from the 5th and 95th of each posterior distribution. Uncertainty related to the model with 4 parameters is higher than for the model with 3 parameters.

To compare and validate the models, values of the scaling factor corresponding to the exposure in 1992, 1998 and 2000 from DGS (1992), Decloître (1998) and Leblanc et al. (2000) are shown in Fig. 4. For both models the function λ predicts these values well whereas the model with 4 parameters appears a little better. The capacity of prediction of the current urinary concentration of the two models is also measured by the calculation of the RMSEP. The first model presents a RMSEP of 14602 and the second model a RMSEP of 14163. RMSEP of the second model is lower than RMSEP of the first model which means that the second model is better for prediction. However the values obtained remain really comparable and the gain of prediction (only 3% of the RMSEP) provided by an increasing number of parameters does not seem significant enough, thus we selected the model with 3 parameters.

Since the oldest individuals in the ENNS study are 74 years old (30 individuals), no data was available to determine the shape of this function before the 1933. Consequently, the uncertainty around the value of the corrected factor for this period (grey-colored in Fig. 4) is huge and the shape of the function for this period is mainly related to the prior information. As it seems unlikely that exposure were zero in 1900, given the natural presence of cadmium in the environment, the prior value for exposure in 1900 was centred on 10% of the current exposure and ranges from 0 to 100% of the current exposure.

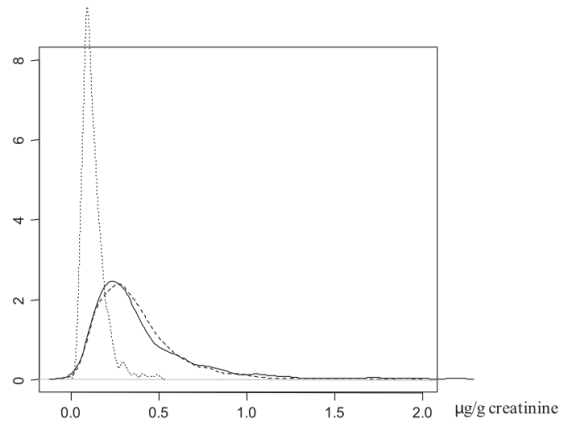


Fig. 3. Distributions of the measured urinary concentration of the French population in µg/g creatinine from the ENNS study (solid line), the urinary concentration predicted directly from current external exposure by the PBPK model (dotted line) and urinary concentration predicted by the population model including current external exposure and correcting function for past exposure (dashed line).

The distribution of the predicted concentrations for the population is compared to the distribution of the observed concentrations in Fig. 3. This figure shows how the addition of the scaling function to take the evolution in the exposure into account improves the prediction of the urinary concentration. The distribution of the observed concentrations is also compared to the distribution of the predicted ones for men and women in 2009 in Table 3. This comparison shows that for both genders the mean is well predicted whereas the 5th percentile is slightly overestimated and the 95th percentile is slightly underestimated.

Risk assessment from external and internal critical values in 2009 and 2030

Using the PBPK model and the scaling function λ for past exposure through a Bayesian population model, we succeed to link current individual exposure from food and tobacco with measured urinary concentration in cadmium. Consequently risk can be assessed from different reference values i.e. external or internal values.

EFSA (2009b) proposed an internal value to evaluate risk regarding cadmium from a meta-analysis on a selected set of studies to evaluate the dose–response relationship between urinary cadmium and urinary beta-2-microglobulin (B2M). From their model, a benchmark dose lower confidence limit for a 5% increase of the prevalence of elevated B2M (BMDL5) of 4 µg Cd/g creatinine was derived. They then applied a chemical-specific adjustment factor of 3.9, to account for inter-individual variation of urinary cadmium within the study populations, leading to a value of 1.0 µg Cd/g creatinine.

Considering the measured urinary concentration in 2009 (Fréry et al., 2011), 4.8% of the women present a urinary concentration higher than 1.0 µg Cd/g creatinine and 2.9% of the men. From the model including the scaling function to correct past exposure, we estimated what the urinary concentration of the French population in 2030 would be under the assumption that dietary exposure will remain stable from 2009 to 2030. Under this assumption, no individual would exceed this value in 2030. The distributions of the urinary cadmium concentrations of the French population and the corresponding risk assessed from internal critical value in 2009 and 2030 are summarised in Table 3. The differences in the urinary concentrations between these two dates are explained by the long half-life of cadmium and the trend in dietary exposure. Indeed, the urinary concentration of a 50 years old individual measured in 2009 is explained by his exposure between 1959 and 2009. In the same way, the urinary concentration of a 50 years old individual in 2030 is explained by his exposure between 1980 and 2030, which, according to our modelling, is lower. Moreover, in this modelling we didn't take any trend in exposure from tobacco smoke into account whereas Tellez-Plaza et al., 2012 highlighted a decreasing trend in urinary concentration in the US population that is partly explained by the decreasing exposure from cigarette smoke.

In order to remain below 1 µg Cd/g creatinine in urine in 95% of the population by age 50, the average daily dietary cadmium intake should not exceed 0.36 µg Cd/kg b.w., corresponding to a weekly dietary intake of 2.52 µg Cd/kg b.w. (EFSA, 2009b). Consequently, considering the estimated dietary exposure in 2009, 1.8% of the men exceed this TWI (tolerable weekly intake) and 1.5% of the women. Assuming that dietary exposure would remain stable from 2009, risk assessment from dietary in 2030 should be the same. The distributions of the dietary exposure of the French population and the corresponding risk assessed from external critical value are summarised in Table 3.

Discussion

The objective of this work was to interpret the measured urinary concentrations of cadmium in the French population through exposure from food and cigarette smoke. In a first attempt to link external exposure estimated from food and cigarette smoke with urinary concentration, a

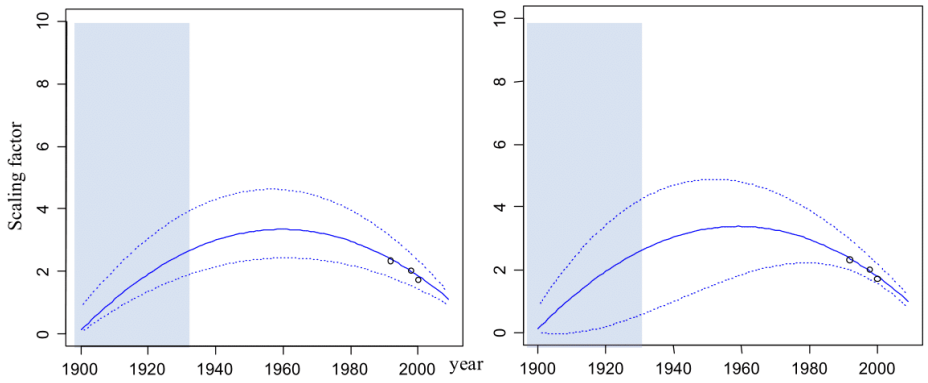


Fig. 4. Scaling function λ obtained from the posterior distributions of the parameters (p5, Mean and p95) from the two models. Points represent values of the scaling factor corresponding to the exposure assessed in 1992, 1998 and 2000 (Decloître, 1998; DGS, 1992; Leblanc et al., 2000). Grey-colored area represents the period of time which was not informed by the data and consequently for which there is a high uncertainty around the factors.

PBPK model for cadmium was used to predict urinary concentrations and then the predicted concentrations were compared with the measured ones. This method was proposed for the interpretation of the National Health and Nutrition Examination Surveys (NHANES) data (Ruiz et al., 2010). The results showed a significant underestimation of cadmium concentrations in urine predicted by the model. As proposed in the discussion of the work by Ruiz et al. (2010), due to the long half-life of cadmium in the human body and its accumulation in kidneys, a potential trend in exposure can explain an underestimation of current urinary concentration. Indeed, techno-economic data on cadmium in France shows a significant decrease in the production of cadmium in France since the 1980s (Pilchard et al., 2005) and some exposure studies conducted in the past highlighted that the dietary exposure to cadmium in France should have been higher in the past (Decloître, 1998; DGS, 1992; Leblanc et al., 2000). The so modified model was able to link external exposures and current urinary concentration in a fairly interpretable way. Indeed, the trend proposed by the modelling is consistent with the past exposure assessments which were not used to inform the model in any way. It was chosen not to include the first Total Diet Study (TDS1) conducted in France in 2000–2004 (Leblanc et al., 2005) in this work. The reason is the incoherence in concentrations which was highlighted in Arnich et al. (2012) but not explained. Indeed Millour et al. (2011) highlighted that the concentration measured in food products known to accumulate cadmium i.e. animal products was actually higher in the TDS1 than in the second Total Diet Study (TDS2) which is coherent with the assumption of a decreasing trend in environment. But due to due to a strangely low contamination level measured in cereals compared to the current one measured in the TDS2, a lower exposure was assessed in this study.

Confirmed by a cross validation, the model was shown to predict well the distribution of the French population urinary concentrations from the individual dietary intakes and intakes from cigarette smoke which were assessed from cadmium levels in foods and cigarettes and individual habits recorded in ENNS.

From the measured urinary concentration recorded in ENNS, a risk assessment can be conducted using the internal value proposed by EFSA (2009b). The comparison does not rule out the risk for women and for men. With this model we are also able to test scenarios of external exposure and consequences on urinary concentration levels and associated risk. As an example, we predicted the urinary concentration of the French population in 2030 assuming that the dietary exposure and the exposure from cigarette smoke will remain stable at the 2009 level. Under this assumption, it results that no individuals would be over the critical value of 1 μg/g creatinine. These results are consistent with the trend in urinary concentration highlighted in the US by Tellez-Plaza et al. (2012) using data from the National Health and Nutrition Examination Surveys (NHANES) from 1988 to 2008. This showed that successive urine cadmium concentrations decreased by 34.3% from 1988–1994 to 2003–2008 in the US adult population after adjustment for age, sex, and race/ethnicity.

The risk assessment conducted in this work was based on kidney effects as proposed by EFSA (2009b). However, other effects of cadmium have been highlighted such as bone effects or cancer (Järup and Akesson, 2009; Nordberg et al., 2007; Satarug et al., 2010). Consequently, other

Table 3
Distributions of the urinary cadmium concentration of the adult French population (W: women, M: men) from ENNS study (*Fréry et al., 2011), predicted by the model in 2009 and 2030, dietary exposure and related risk assessment from external and internal critical values.

			Population distribution			Risk assessment	
			5th *p*	*mean*	95th *p*	*Critical value*	*risk*
Urinary concentration *(μg/g creatinine)*	2009*	W	0.13	0.41	0.96	1	4.8%
		M	0.09	0.33	0.81	(μg/g creatinine)	2.9%
	2009	W	0.16	0.43	0.92		4.4%
		M	0.13	0.35	0.76		2.1%
	2030	W	0.08	0.18	0.35		0%
		M	0.07	0.15	0.29		0%
Dietary exposure *(μg/kg b.w/week)*	2009/2030	W	0.45	1.06	1.8	2.5	1.5%
		M	0.49	1.04	1.9	(μg/kg b.w/week)	1.8%

possible endpoints associated with cadmium exposure could be considered to strengthen this risk assessment. Järup and Akesson (2009) highlighted that some adverse effects on bone have been detected at urinary cadmium levels from 0.5 μg/g creatinine. They also suggest that increased risk of cancer has been indicated at similar levels. For illustration purposes, compare the urinary concentrations measured in the general French population to this critical value instead of 1 μg/g creatinine would have lead to a risk of 22%. In this work, the risk was assessed for the general French population from the ENNS population. Some population sub-groups are known to have higher exposure to cadmium and their exposure and corresponding risk need to be investigated from other cohorts. For example, Sirot et al. (2008) investigated the cadmium dietary intake and urinary concentration levels in French high seafood consumers. It resulted in a mean dietary intake of 2.4 μg/kg bw/week (compared to 1 μg/kg bw/week for the general population in this work) and a mean urinary concentration of 0.65 μg/g creatinine (compared to 0.39 μg/g creatinine for the general population).

Some limitations remain also in the modelling. First, regarding the PBPK modelling, Bois et al. (2010) reported many sources of inter-individual variability that should be taken into account in a PBPK modelling. To refine the model, it would be relevant to investigate more the inter-individual variability in the parameters of the model using some of the many tools now available and described in Bois et al., 2010 and the consequences on the outputs of the model. Regarding, the age-dependent variability, only the modification of the body weight and the renal function with the age were considered in the model. However, it was demonstrated that many parameters of the model can vary with age (Clewell et al., 2004) and this could also be investigated to refine the model. The difference between women and men levels are often observed in the biomonitoring data. We also observed a significant difference in the French population. This difference is explained in Järup and Akesson (2009) by the close relation between cadmium absorption and the expression of the divalent metal transporter 1 (DMT-1), which transports cadmium and iron into the mucosa cell in a competitive manner. The PBPK model we used in this work models this difference by using a different absorption factor for male and female in the general population. However, this situation seems to be exacerbated during pregnancy when enterocytes have an increased DMT-1 density at the apical surface to optimise micronutrients absorption (Åkesson et al., 2002; Vahter et al., 2007). This is also an identified source of inter-individual variability that would be interesting to add to the model to precisely investigate the risk for pregnant women.

In our modelling, we considered exposure from two sources: food and cigarette smoke whereas often only dietary exposure is considered. Nevertheless, other sources of exposure to cadmium were highlighted. EFSA (2009b), estimated that exposure from house dust could raised 20% of the total exposure for the children whereas it remains very low for adults. Regarding, exposure from cigarette smoke, we did not consider any change in the exposure in this work, apart from the individuals who stopped smoking. We assumed that the average number of cigarettes smoked by day declared by the individuals didn't change during their life from the date they declared started to the date they declared stopped. This assumption is justified when we are linking the external exposure assessed from the individual habits recorded in ENNS with the measured individual urinary concentration for this population in 2009. However, when our results are extrapolated to a generational context and use to predict the risk in France in 2030, a potential trend could have existed. Tellez-Plaza et al. (2012) showed that the reduction in adjusted urine cadmium concentrations observed in the NHANES survey from 1988 to 2008 was stronger in smokers which highlighted a population trend in the smoking habits. Taking this potential decreasing trend in exposure from cigarette smoke into account would have lead to an ever lower predicted urinary concentration in 2030.

This work demonstrates that it is possible to deal with the relation of current exposure assessment and internal doses measurements when considering persistent chemicals. The objective of developing models to link external and internal exposure was the interpretation of the biomonitoring data. Our model can be used to test scenarios of external exposure and consequences on urinary concentration levels and associated risk. From the measured urinary concentration recorded in the ENNS and the critical urinary concentration proposed by Efsa in 2009 the risk for the French population cannot be ruled out. From our model and assuming that the dietary exposure and the exposure from cigarette smoke will remain stable at the 2009 level, it results in no more risk in 2030. This result should be interpreted with caution, taking into account the limitations discussed above but it still of interest for risk managers.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank Usen and InVS for the access to ENNS data and their assistance.

References

Åkesson, A., Berglund, M., Schütz, A., Bjellerup, P., Bremme, K., Vahter, M., 2002. Cadmium exposure in pregnancy and lactation in relation to iron status. Am. J. Public Health 92, 284–287.
Amzal, B., Julin, B., Vahter, M., Wolk, A., Johanson, G., Åkesson, A., 2009. Population toxicokinetic modeling of cadmium for health risk assessment. Environ. Health Perspect. 117 (8), 1293.
Arnich, N., Sirot, V., Rivière, G., Jean, J., Noël, L., Guérin, T., Leblanc, J.C., 2012. Dietary exposure to trace elements and health risk assessment in the 2nd French Total Diet Study. Food Chem. Toxicol. 50 (7), 2432–2449.
Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162 (4), 2025–2035.
Béchaux, C., Crépet, A., Clémençon, S., 2014. Improving dietary exposure models by imputing biomonitoring data through ABC methods. Int. J. Biostat. 1557–4679. http://dx.doi.org/10.1515/ijb-2013-0062 (ISSN (Online), ISSN (Print) 2194–573X, July).
Berglund, M., Åkesson, A., Nermell, B., Vahter, M., 1994. Intestinal absorption of dietary cadmium in women depends on body iron stores and fiber intake. Environ. Health Perspect. 102, 1058–1066.
Bernillon, P., Bois, F.Y., 2000. Statistical issues in toxicokinetic modeling: a Bayesian perspective. Environ. Health Perspect. 108 (Suppl. 5), 883–893.
Blum, M., Chi Tran, V., 2010. Hiv with contact tracing: a case study in approximate bayesian computation. Biostatistics 11, 644–660.
Bois, F.Y., Jamei, M., Clewell, H.J., 2010. PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology 278 (3), 256–267.
Castetbon, K., Vernay, M., Malon, A., Salanave, B., Deschamps, V., Roudier, C., Oleko, A., Szego, E., Hercberg, S., 2009. Dietary intake, physical activity and nutritional status in adults: the French nutrition and health survey (ENNS, 2006–2007). Br. J. Nutr. 102 (05), 733–743.
Choudhury, H., Harvey, T., Thayer, W.C., Lockwood, T.F., 2001. Urinary cadmium elimination as a biomarker of exposure for evaluating a cadmium dietary exposure-biokinetics model. J. Toxicol. Environ. Health A 63, 321–350.
Clewell, R.A., Clewell III, H.J., 2008. Development and specification of physiologically based pharmacokinetic models for use in risk assessment. Regul. Toxicol. Pharmacol. 50 (1), 129–143.
Clewell, H.J., Gentry, P.R., Covington, T.R., Sarangapani, R., Teeguarden, J.G., 2004. Evaluation of the potential impact of age- and gender-specific pharmacokinetic differences on tissue dosimetry. Toxicol. Sci. 79, 381–393.
Csillry, K., Blum, M., Gaggiotti, O., Franois, O., 2010. Approximate Bayesian Computation (ABC) in practice. Trends Ecol. Evol. 25, 490–491.
Decloître, F., 1998. La part des différents aliments dans l'exposition au plomb, au cadmium et au mercure, en France. Cah. Nutr. Diet. 33 (3), 167–175.
DGS (Direction Générale de la Santé)., 1992. La diagonale des métaux, Etude sur la teneur en métaux dans l'alimentation (Paris).
Diamond, G.L., Thayer, W.C., Choudhury, H., 2003. Pharmacokinetics/pharmacodynamics (PK/PD) modeling of risks of kidney toxicity from exposure to cadmium: estimates of dietary risks in the U.S. population. J. Toxicol. Environ. Health A 66, 2141–2164.
Dubuisson, C., Lioret, S., Touvier, M., Dufour, A., Calamassi-Tran, G., Volatier, J.L., Lafay, L., 2010. Trends in food and nutritional intakes of French adults from 1999 to 2007: results from the INCA surveys. Br. J. Nutr. 103, 1035–1048.
EFSA (European Food Safety Authority), 2009a. Scientific opinion on the reevaluation of Allura Red AC (E 129) as a food additive on request from the European Commission. The EFSA Journal 7 (11), 1327.
EFSA (European Food Safety Authority), 2009b. Scientific opinion on the Cadmium in Food. The EFSA Journal 980, 1–139.
Filippi, S., Barnes, C.P., Cornebise, J., Stumpf, M.P., 2013. On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. Stat. Appl. Genet. Mol. Biol. 12 (1), 87–107.

Fréry, N., Saoudi, A., Garnier, R., Zeghnoun, A., Falq, G., 2011. Exposition de la population française aux substances chimiques de l'environnement. Institut de veille sanitaire, Saint-Maurice (151 pp.).
Hays, S.M., Becker, R.A., Leung, H.W., Aylward, L.L., Pyatt, D.W., 2007. Biomonitoring equivalents: a screening approach for interpreting biomonitoring results from a public health risk perspective. Regul. Toxicol. Pharmacol. 47 (1), 96–109.
Hays, S.M., Aylward, L.L., LaKind, J.S., Bartels, M.J., Barton, H.A., Boogaard, P.J., Brunk, C., DiZio, S., Dourson, M., Goldstein, D.A., Lipscomb, J., Kilpatrick, M.E., Krewski, D., Krishnan, K., Nordberg, M., Okino, M., Tan, Y.M., Viau, C., Yager, J.W., 2008. Guidelines for the derivation of biomonitoring equivalents, report from the biomonitoring equivalents expert workshop. Regul. Toxicol. Pharmacol. 51, S4–S15.
Järup, L., Åkesson, A., 2009. Current status of cadmium as an environmental health problem. Toxicol. Appl. Pharmacol. 238 (3), 201–208.
Kippler, M., Ekström, E.C., Lönnerdal, B., Goessler, W., Åkesson, A., El Arifeen, S., Persson, L. A., Vahter, M., 2007. Influence of iron and zinc status on cadmium accumulation in Bangladeshi women. Toxicol. Appl. Pharmacol. 222, 221–226.
Kjellström, T., Nordberg, G.F., 1978. A kinetic model of cadmium metabolism in the human being. Environ. Res. 16, 248–269.
Leblanc, J.C., Malmauret, L., Guérin, T., Bordet, F., Boursier, B., Verger, P., 2000. Estimation of the dietary intake of pesticide residues, lead, cadmium, arsenic and radionuclides in France. Food Addit. Contam. 17 (11), 925–932.
Leblanc, J.C., Guérin, T., Noël, L., Calamassi-Tran, G., Volatier, J.L., Verger, P., 2005. Dietary exposure estimates of 18 elements from the 1st French total diet study. Food Addit. Contam. 22, 624–641.
Lyons, M.A., Yang, R.S., Mayeno, A.N., 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 (8), 1040–1046.
Millour, S., Noël, L., Kadar, A., Chekri, R., Vastel, C., Sirot, V., Leblanc, J.C., Guérin, T., 2011. Pb, Hg, Cd, As, Sb and Al levels in foodstuffs from the 2nd French total diet study. Food Chem. 126 (4), 1787–1799.
Nordberg, G.F., Nogawa, K., Nordberg, M., Friberg, L., 2007. Cadmium. Handbook on the Toxicology of Metals. Elsevier, Amsterdam, pp. 445–486.
Pichard, A., Bisson, M., Diderich, R., Houeix, N., Hulot, C., Lacroix, G., Pépin, G., 2005. Cadmium et ses dérivés. Fiche de données toxicologiques et environnementales des substances chimiques. INERIS, Verneuil-en-Halatte.
Ruiz, P., Mumtaz, M., Osterloh, J., Fisher, J., Fowler, B.A., 2010. Interpreting NHANES biomonitoring data, cadmium. Toxicol. Lett. 198 (1), 44–48.
Satarug, S., Garrett, S.H., Sens, M.A., Sens, D.A., 2010. Cadmium, environmental exposure, and health outcomes. Environ. Health Perspect. 118, 182–190.
Sirot, V., Samieri, C., Volatier, J.L., Leblanc, J.C., 2008. Cadmium dietary intake and biomarker data in French high seafood consumers. J. Expo. Sci. Environ. Epidemiol. 18, 400–409.
Tellez-Plaza, M., Navas-Acien, A., Caldwell, K.L., Menke, A., Muntner, P., Guallar, E., 2012. Reduction in cadmium exposure in the United States population, 1988–2008: the contribution of declining smoking rates. Environ. Health Perspect. 120 (2), 204.
Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6 (31), 187–202.
Vahter, M., Åkesson, A., Lidén, C., Ceccatelli, C., Berglund, M., 2007. Gender differences in the disposition and toxicity of metals. Environ. Res. 104, 85–95.

3 Human variability in polymorphic metabolism and related uncertainty factors: focus on inter-ethnic and inter-phenotype differences in CYP2D6 metabolism.

Béchaux C, Amzal B, Quinot N, Crépet A and Dorne JLM.

Abstract : Conventionally, a 3.16-fold default factor has been used in non-cancer risk assessment to cover human variability in toxicokinetics. The objective of this work is to investigate the human variability in kinetic in phenotyped healthy adults in order to test the adequacy of such factor in case of polymorphic metabolism. Published pharmacokinetic studies in Caucasian and Asian extensive (EMs) intermediate (IMs) and poor metabolizers (PMs) for CYP2D6 have been analyzed using data primarily related to chronic oral exposure (oral clearance and area under the plasma concentration time curve). A multi-level hierarchical Bayesian model has been proposed to meta-analyze these data integrating all quantifiable sources of variability, including inter-compound and inter-study variability. This work highlights large individual differences in kinetic within the Caucasian EMs (coefficients of variation above 50%). A slightly weaker inter-individual variability was found within Caucasian PM (coefficients of variation around 35%) and within Asian population (coefficients of variation below 40% for EMs and IMs). Comparisons between Caucasian EMs and PMs and Asian EMs and IMs revealed an increase internal dose for PMs and IMs with a ratio compared to EMs that ranges from 1 to 34 and from 1 to 5. Uncertainty factors were calculated for each subgroup and for the two populations taking the prevalence of PMs and IMs into account through Monte-carlo simulations. It results that none of the subgroup would be cover by the 3.16-fold default factor and a 11-fold CYP2D6-related factors would be necessary to cover 95% of Caucasian and Asian populations. An exponential relationship was found between the percentage of CYP2D6 metabolism in EMs and the uncertainty factors for PMs. Such relation combined with in vitro method could be used to establish chemical-specific uncertainty factors or as input to for kinetic models used in risk assessment.

Key words: Human variability, pharmacokinetic, Bayesian inference, Meta-analysis, Risk assessment, Uncertainty factor

1.Introduction

A 100-fold default factor has been usually used in non-cancer risk assessment to account for interspecies differences (10-fold) and human variability (10-fold) (Truhaut 1991). Then, the uncertainty factor that cover the human variability have been subdivided into two equal default uncertainty factors ($10^{0.5}$ = 3.16) to allow for differences in toxicokinetics and toxicodynamics (Renwick 1993) WHO 1999). Such subdivision was proposed in order to allow for the replacement of this default uncertainty factor with chemical-specific factors. As an intermediate approach, Renwick and Lazarus (1998) proposed using pathway-related default uncertainty factors. They first applied this approach to human variability in metabolism by CYP1A2 and glucuronidation (Renwick and Lazarus 1998; Dorne, Walton et al. 2001; Dorne, Walton et al. 2001; Walton, Dorne et al. 2001; Walton, Dorne et al. 2001). Following this approach, a number of reviews have been produced using therapeutic drug databases and analyzing human variability for both the kinetic and the dynamic aspects that can be used to propose pathway-related default uncertainty factors (Dorne, Walton et al. 2001; Dorne, Walton et al. 2001; Naumann, Silverman et al. 2001; Walton, Dorne et al. 2001; Walton, Dorne et al. 2001; Dorne, Walton et al. 2002; Ginsberg, Hattis et al. 2002; Dorne, Walton et al. 2003; Dorne, Walton et al. 2003; Dorne, Walton et al. 2004; Dorne, Walton et al. 2005).

From such a review, Dorne, Walton et al. (2002) highlighted some potentially sensitive subgroups of the human population to which a particular attention should be paid for risk assessment. Among them, they identified the individuals with a specific genetic polymorphism that leads to a slower metabolism referred as poor metabolisers. Earlier, Calabrese (1985) when attempted to characterize human inter-individual variability in kinetic, already suggested that genetically determined biochemical differences could result in variability spanning several orders of magnitude. Consequently investigating the impact of genetic polymorphism in metabolism on the inter-individual variability and thus the corresponding uncertainty factor that would allow to protect potential sensitive subgroups is of great interest for risk assessment (Martin, Martin et al. 2013). Metabolic pathways are usually described as separated phases: phase I that includes CYP1A2, CYP2A6, CYP2E1, CYP3A4, ADH and hydrolysis, phase II that includes glucuronidation, glycine and sulphate conjugation, and renal excretion. Polymorphic pathways have been highlighted from in vivo kinetic data in phenotyped individuals regarding phase I (CYP2C9, CYP2C19 and CYP2D6) and phase II (NAT2) (Dorne 2010). However, when these reviews of therapeutic drug databases were conducted, only few studies included the phenotypes of the individuals. Since the mid-2000s a lot of new data has been published on the kinetic of therapeutic drug in phenotyped individuals. Especially, many studies investigated the interethnic differences in the pharmacokinetics.

The objective of this work is thus to update the knowledge on human variability in kinetic of polymorphic pathway using phenotyped data from therapeutic drug studies and focusing on phenotype and interethnic differences. This work focuses on the effect of CYP2D6 polymorphism on the kinetic of CYP2D6 substrates.

Indeed, CYP2D6 is an important enzyme that metabolizes about 25% of prescribed drugs since it has been shown to be clinically important in the oxidation of many drugs including anti-arrhythmic drugs, neuroleptics, antidepressants, b-blockers and opiates. Eichelbaum (2003) also demonstrated several endogenous substrates of CYP2D6. Especially, it has been shown that CYP2D6 is responsible for the O-demethyletion of the 5-methoxytriptamide to serotonin. Consequently, many studies investigated the relationship between suicidal behavior, personality and CYP2D6 phenotypes and found relationships between CYP2D6 metabolic capacity and personality. For example, González, Peñas-Lledó et al. (2008) demonstrated on a population of healthy Cuban volunteers that the CYP2D6 hydroxylation capacity was related to the degree of anxiety and socialization. These results support the postulated reduction of serotonin in CYP2D6 poor metabolisers, which may be associated with a cluster of behavioral traits (e.g., anxiety, impulsivity). Bertilsson (2010) investigated more deeply a potential relationship between CYP2D6, serotonin level and suicide. They conclude that even if no full explanation why CYP2D6 is related to suicidality can be provide yet, the relation still seems real and related to serotonin.

Therefore, CYP2D6 polymorphism has been investigated intensively during the last decade and more than 80 polymorphisms have been described for CYP2D6. Many different alleles have been identified which can be associated with a non-activity, a decreased activity or a normal activity and the number of identified alleles is still increasing (Zhou 2009). The combinations of these allele form different genotypes and leads to different metabolism capacities of CYP2D6 that are grouped into three main phenotypes: extensive metabolisers (EMs), intermediate metabolisers (IMs) and poor metabolisers (PMs). An additional phenotype is related to multiple copies of the CYP2D6 gene that leads to a higher activity of the enzyme: the ultra-rapid metabolisers (UMs). Frequencies of these phenotypes in different ethnic groups have been widely studied. The PM phenotype was thus shown to have a prevalence ranging from 5 to 15% in the Caucasian population whereas the prevalence of this phenotype is much lower in African population (ranging from 1 to 2%) and in the Asian population (between 0 and 1%). Conversely, the prevalence of IM in the Asian population is high (between 35 and 55%) whereas it is very slow in Caucasian population (less than 2%). In this work, a focus is made on inter-variability related to phenotype and ethnicity. Since most of the new data are provided for Caucasian or Asian, the inter-individual variability is studied in this work within these two groups. A

meta-analysis is realized for each group from a data base built from a review of literature of studies providing pharmacokinetic data of CYP2D6 substrates on phenotyped healthy adults.

The approaches usually used to meta-analyze human TK data provide weighted averages of the variability for normal and lognormal data in each subgroup of the population assuming fixed effect models (weighted geometric means according to study sample sizes) (Dorne, Walton et al. 2005). Such studies have been proven useful to risk assessors to include TK and mechanistic data in the risk assessment process in a more quantitative fashion. However, refined methods to include a full account of the variability structure and in particular, the quantification of inter-study variability together with the choice of weights would be optimal. A drawback of TK studies is the small sample size (less than 10) so that alternative methods such as inverse of estimate variances are difficult to implement. Additionally, these meta-analyses were conducted for each substrate independently and then combined to derive variability for the metabolic route for each identified population sub-group (Dorne, Walton et al. 2001; Dorne, Walton et al. 2002). Another improvement would thus be a more explicit quantification of inter-substrate variability for optimal use of the information. Refined approaches to meta-analysis for health-care and risk assessment have been investigated including from a Bayesian perspectives (Sutton and Higgins 2008; Rigaux, Denis et al. 2013). Indeed, Bayesian inference is particularly adequate associated with hierarchical models to account for inter-study variability, or to discount information from various types of studies. In this work, a full-Bayesian approach is proposed for the meta-analysis of TK data using a multi-level hierarchical model to integrate all quantifiable sources of variability, including inter-compound variability. In this way, inter-individual variability and related uncertainty factors will be provided for each group, each pharmacokinetic parameter, each phenotype and each substrate. Then, an uncertainty factor for each pharmacokinetic parameter and each substrate will be estimated for each group by simulating a whole population taking the prevalence of the different phenotypes in each population into account. Finally, the predictability of the CYP2D6-related default uncertainty factors will be investigated by modeling the relationship between the percentage of CYP2D6 metabolism of the substrates and the proposed uncertainty factors.

2.Material and Method

2.1.Data collection and standardization

A literature review has been performed from year 2006 to 2013 (December, the 17th 2013) using online databases PUBMED (NML) and Web of Science. Kinetic parameters related primarily to chronic

exposure i.e. clearance (CL) and area under the plasma-concentration-time curve (AUC) after an oral dose were extracted. Only the phenotyped data on healthy individuals were considered. This data base was supplemented with phenotyped data on healthy adult from (Dorne, Walton et al. 2002) and control data from the NOMIRACLE project (Løkke 2010).

Data extracted from the selected studies were first standardized in same units for each kinetic parameter: ng.h.ml^{-1} for AUC and mL.h^{-1} for Clearance. Clearance that were not already reported in kg^{-1} were standardized by the individual weight using the reported weight when available or 70 kg for males, 60 kg for females and 65 kg for mixed males and females. AUC values were normalized with the dose expressed in mg by kg of body weight in order to reduce variability related to the dose and body weight.

Data in the kinetic studies were most of the time either reported as arithmetic means (X) and standard deviations (SD) or by geometric mean (GM) and geometric standard deviation (GSD). Since kinetic data are generally recognized to be lognormally distributed (Naumann, Weideman et al. 1997; Renwick and Lazarus 1998; Dorne, Walton et al. 2001) and the geometric mean (GM) and geometric standard deviation (GSD) are better appropriate to summarize a lognormal distribution, all the kinetic data in this work are given in GM and GSD. When these measures were not reported, they were estimated for each individual study using the following equations:

$$GM = \frac{X}{\sqrt{(1 + CV_N^2)}}$$

$$GSD = \exp(\sqrt{\ln(1 + CV_N^2)})$$

Where CV_N is the coefficient of variation for normally distributed data given by: $CV_N = \frac{SD}{X}$.

2.2.Phenotypes groups and ethnicity

There was also a need of standardization regarding the reported phenotypes. Indeed, sometimes phenotypes were already reported as EM (extensive metabolizer), IM (intermediate metabolizer), PM (poor metabolizer), UM (ultrar apid metabolizer) whereas sometimes the genotypes (i.e. the combination of alleles) were given. In order to assign phenotype groups to reported genotypes, most common associations between phenotype groups and alleles were investigated from 20 studies (Lee, Sohn et al. 2006; Pedersen, Damkier et al. 2006; Cui, Teng et al. 2007; Curry, Watts et al. 2007; Gan, Ismail et al. 2007; Garcia-Quetglas, Azanza et al. 2007; Kubo, Koue et al. 2007; Kuhn, Kirsch et al.

2007; Slanar, Nobilis et al. 2007; Kirchheiner, Keulen et al. 2008; Malhotra, Guan et al. 2008; Borobia, Novalbos et al. 2009; Malhotra, Wood et al. 2009; Samer, Daali et al. 2010; Sharma, Pibarot et al. 2010; Nichols, Focht et al. 2011; Azuma, Hasunuma et al. 2012; Yoo, Cho et al. 2012; Blake, Kharasch et al. 2013; Gasso, Papagianni et al. 2013). From this analysis, it was chosen to associate PM phenotype to homozygosity for alleles with no activity, IM with homozygosity with decreased activity alleles or heterozygosity with one decreased activity alleles and one non activity alleles, and EM with homozygosity with normal activity alleles or heterozygosity with one decreased activity alleles and one non activity alleles. Table 1 summarizes these associations. UM gathers individuals with at least one active gene duplication.

Two different ethnic groups are studied in this work: Caucasian and Asian. Specific allele's prevalence is associated with these groups. Regarding Caucasian, the PM phenotype was shown to have a prevalence ranging from 5 to 15% according to different studies (Cho and Lee 2006; Lee, Sohn et al. 2006; Pedersen, Damkier et al. 2006; Li, Wang et al. 2010; Malhotra, Darsey et al. 2011; Nichols, Focht et al. 2011; Chiba, Kato et al. 2012; Barter, Tucker et al. 2013) and the IM phenotype is rare with a prevalence lower than 2% (Gan, Ismail et al. 2007; Barter, Tucker et al. 2013). The main non-functional alleles responsible for the PM phenotype in Caucasian is CYP2D6*4. In Asian populations, the phenotype IM is very frequent due to the allele CYP2D6*10 with a prevalence between 35 and 55% (Cho and Lee 2006; Lee, Sohn et al. 2006; Gan, Ismail et al. 2007; Chiba, Kato et al. 2012; Barter, Tucker et al. 2013) and the PM very rare (between 0 and 1%).

Table 1 : Combinations of alleles associated with phenotype groups PM, IM and EM. The numbers in the columns represent the number of alleles belonging to the corresponding group.

	Non activity : (*3,*4,*5,*6,*7,*8,*11,*14, *15, *18,*19, *20, *21,*40)	**Decreased activity :** (*9,*10,*17,*21, *36,*29,*41,*45,*46)	**Normal activity :** (*1,*2)
PM	2		
IM		2	
	1	1	
EM			2
		1	1

2.3.Bayesian hierarchical model for meta-analysis

The objective of the meta-analysis is to provide accurate information on the means (μ_j) and the inter-individual variability (τ_j) of the PK parameters for a substrate *j* based on the combination of results from multiple independent studies *k*. It is thus necessary to properly separate and identify the variability related to differences between studies (τ_{st}), the variability related to differences between

substrates (τ_{sub}) and the variability related to differences between individuals (τ_j) by decomposing the variance of the PK parameter (CL or AUC). Consequently, a hierarchical model is used and the structure of the model showing the conditional dependencies among the population and the individual parameters is summarized by a graphical model (Figure 1).

On the logarithmic scale, individual value for a chosen PK parameter (i.e. CL or AUC) X_{ijk} for *i=1, …,n* are assumed to be independently and identically distributed according to normal distribution of mean μ and variance σ^2 for a given substrate *j* in a given study *k.* Therefore, according to the central limit theorem, the means and the variances $\overline{X_{jk}} = \frac{1}{n}\sum_{i=1}^{n} X_{ijk}$ and $S_{jk} = \frac{1}{n_{jk}-1}\sum_{i=1}^{n}(X_{ijk} - \overline{X_{jk}})^2$ are independent conditionally to the study and the substrate and distributed according to:

$$\overline{X_{jk}} \sim Normal\ (\mu\ , \frac{\sigma^2}{n})$$

$$V_{jk} \sim \frac{\sigma^2}{n} Chi^2(n\ -1)$$

From the literature review, the individual PK parameter X_{ijk} are not provided. Only the geometric means (gm_{jk}) and the variance (v_{jk}) are available for a substrate *j* in a given study *k*. Consequently, the log of the geometric means (lgm_{jk}) and the variance (lv_{jk}) are used and modeled by:

$$lgm_{jk} \sim Normal\ (\mu_{jk}, \frac{n_{jk}}{\tau_j})$$

$$lv_{jk} \sim \frac{n_{jk}}{\tau_j} Chi^2(n_{jk} - 1)$$

where τ_j is the precision (inverse of the variance) that describes the inter-individual variability regarding the substrate *j*. This model accounts for all the information recorded from the study under the assumption of lognormality of data, and allows for the inference on the inter-individual variability τ_j that is the key parameter in this work.

In order to properly describe inter-study and inter-substrate variability, two second layers in the model are required. A first step was build assuming that μ_{jk} is normally distributed around the substrate-specific mean μ_j with variance the inter-study variance τ_{st}:

$$\mu_{jk} \sim Normal\ (\mu_j, \frac{1}{\tau_{st}})$$

For the sake of simplicity and to avoid identifiability issues, the inter-study variability τ_{st} was assumed to be identical for all substrates.

At the last layer of the model, the means (μ_{sub} and μ_{ind}) and the precision (τ_{sub} and τ_{ind}) of the mean and inter-individual variability of the PK parameter considering all substrate together are provided, assuming that μ_j and τ_j are normally distributed:

$$\mu_j \sim Normal\ (\mu_{sub}, \frac{1}{\tau_{sub}})$$

$$\tau_j \sim Normal\ (\mu_{ind}, \frac{1}{\tau_{ind}})$$

Bayesian inferences are used to infer on parameters of the model as it was seen as the most convenient approach to handle such a multi-level model. Since the purpose of this model is the meta-analysis of data from an extensive literature search, it was not consistent to look at independent information to fixe proper prior distributions. For the same reason, expert knowledge exists but it may be related to data from the literature used to run the model. Consequently, non-informative priors were chosen for all parameters. The JAGS software (Plummer 2003) is used to implement the model, the chi-square distribution being described using a Gamma distribution of parameters:

$$lv_{jk} \sim Gamma\ (\frac{n_{jk}-1}{2}, \frac{\tau_j\, n_{jk}}{2}).$$

For each meta-analysis, 2 different Markov chains are run and convergence of the chains are assessed via Gelman-Rubin tests implemented in the Coda package of the R software (Plummer, Best et al. 2006).

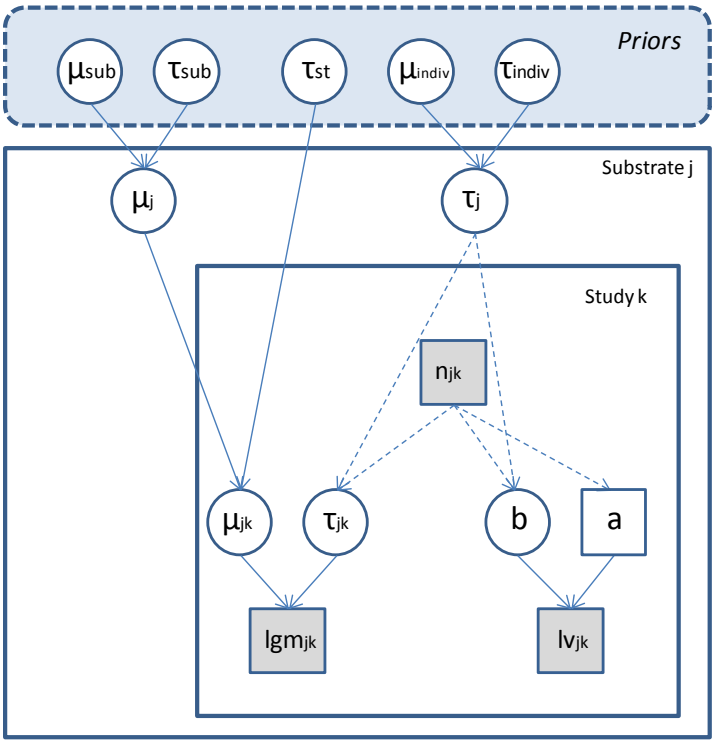


Figure 3: Graphical representation of the hierarchical model for meta-analysis. Squares represent the known quantities: the logarithm of the geometric mean (lgm) and variance (lv) of the study k for the substrate j, the number of individual of this study (n) and a=(n-1)/2. Circles represent unknown quantities to be updated via Bayesian inferences: the mean (μ_{jk}) and the precision τ_{jk} of lgm, the mean (μ_j) and the precision τ_j of the PK parameter for the substrate j and b=n.τ_j/2. Solid arrows represent a stochastic link and dashed arrow represents a deterministic link.

2.4. Model output

The model for meta-analysis is implemented for each PK parameter (i.e. CL, AUC) and each phenotype group (i.e. PM and EM for Caucasian, IM and EM for Asian). Therefore, the highest level of the model provides the variability inter-study (CV_{st}) and the variability inter-subsrate (CV_{sub}), calculated as:

$$CV_{st} = 100 \times \sqrt{e^{1/\tau_{st}-1}}$$

$$CV_{sub} = 100 \times \sqrt{e^{1/\tau_{sub}-1}}$$

At the substrate level, the model provides, for each phenotype group, the meta-analyzed geometric mean (GM_j) and coefficient of variation (CV_j) of the considered PK parameter, calculated as follow:

$$GM_j = e^{\mu_j}$$

$$CV_j = 100 \times \sqrt{e^{1/\tau_j-1}}$$

Thanks to Bayesian modeling, a distribution is available for each parameter estimated in the model. This makes it possible to take the uncertainty around the parameter into account. Therefore, in this work, the median value and a 90% confidence interval are given for each parameter estimation.

For each substrate, the ratio of the GM for a phenotype group on the GM for the reference group (i.e. EM group) is calculated. A confidence interval around this ratio is also provided by sampling 10000 values in the distribution of each GM and calculating 10000 corresponding ratio. Thus, the median is reported with a confidence interval formed by the 5^{th} and 95^{th} percentile of this distribution of ratio.

2.5. Uncertainty factors

Several, CYP2D6-related uncertainty factors (UF) are calculated in this work. Indeed, as highlighted in Martin, Martin et al. (2013), there is several way to calculate UF depending on the population of interest. Therefore, when considering an unimodal distribution e.g. the distribution of the EM population, the UF are calculated as the ratio between the chosen high percentile (i.e. 95^{th} or 99^{th}) and the median of the distribution. In order to especially protect a specific group (e.g. PM or IM), the UF is calculated as the ratio between the chosen high percentile of the group and the median of the reference group (i.e. EM) as made in Dorne, Walton et al. (2002). As previously, a confidence interval around the UF is given by calculating 10000 UFs and providing median, 5th and 95^{th} percentile.

Finally, the distribution of the whole general population can be seen as a bimodal distribution (Martin, Martin et al. 2013) formed by the combination of the EM distribution and the PM distribution (regarding Caucasian population). In this case, the UF is also calculated as the ratio between the chosen high percentile of the bimodal distribution and the median of the bimodal distribution. In order to calculate such UF factor, a distribution for the general population is simulated combining the distribution of the EM, the distribution of the PM and the percentage of PM in the population. The percentage of PM in the population α is determined from the literature. There is some uncertainty around this value, which ranges for the Caucasian population from 5 to 15% according to the selected study with a highest probability around 10% (Cho and Lee 2006; Lee, Sohn et al. 2006; Pedersen, Damkier et al. 2006; Li, Wang et al. 2010; Malhotra, Darsey et al. 2011; Nichols, Focht et al. 2011; Chiba, Kato et al. 2012; Barter, Tucker et al. 2013). In order to take this uncertainty into account a triangular distribution $F\alpha$ is used where the lower limit equals to 5, upper limit equals to 15 and mode equals to 10. The UFs are calculated from the following procedure:

1) Sample a value for α in Fα
2) Sample (1-α) x 10000 values in the EM distribution and α x 10000 values in the PM distribution
3) Calculate the 50th and the 95th (or 99th) percentile of the 10000 values
4) UF = 95thp / 50th (or 99thp / 50th)

Repeat the 4 steps 1000 times to provide uncertainty related to α

Same calculations are done for Asian population by replacing PM by IM. Fα is also a triangular distribution where the lower limit equals to 35, upper limit equals to 55 and mode equals to 45% (Cho and Lee 2006; Lee, Sohn et al. 2006; Gan, Ismail et al. 2007; Chiba, Kato et al. 2012; Barter, Tucker et al. 2013). The different ways for the calculation of UFs are illustrated in figure 2 from the example of the Caucasian population.

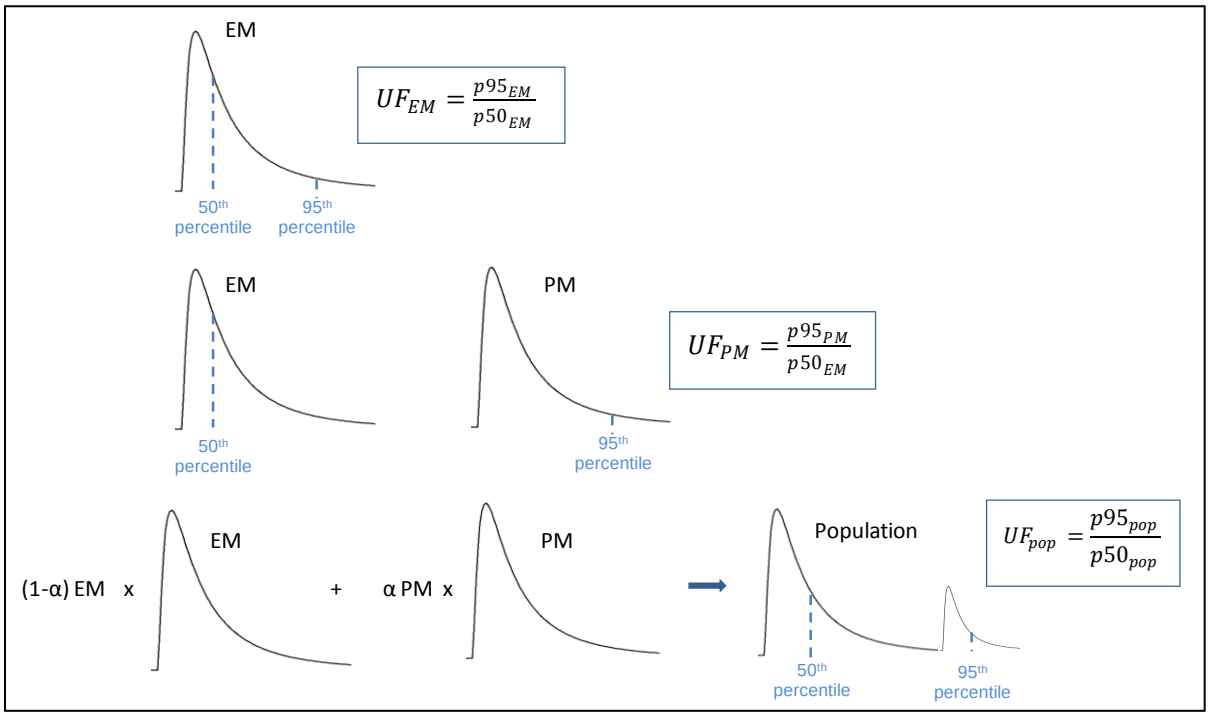


Figure 4 : Different ways of calculating the uncertainty factors (UF) from extensive metabolizer (EM) distribution, poor metabolizer (PM) distribution and combination of both distributions according to the rate of poor metaboliser in the population (α).

3.Results

3.1.Collected data and inter-study variability

The number of studies extracted from the literature and the corresponding number of substrates and individuals included in each meta-analysis are reported table 2. Only few studies were available for IM in Caucasians (Lessard, Yessine et al. 1999; Lee, Kim et al. 2006; Borobia, Novalbos et al. 2009; Sehrt, Meineke et al. 2011) and none for PM in Asians which is consistent with the low prevalence of these phenotypes in the respective populations. Due to insufficient number of individuals, no meta-analysis was implemented for these groups.

For the remaining groups, meta-analyses were implemented and inter-study and inter-substrate coefficients of variation provided by the model are reported in Table 2. The CV corresponding to inter-study variability (CV_{st}) remains around 50-60% whatever the PK parameter considered. This is consistent with Dorne, Walton et al. (2002) who found inter-study differences within a 30% range of the overall mean. Inter-substrate variability (CV_{sub}) is necessary in the model but is not really of interest for this work since it only describes how kinetic can differ from a substrate to another and thus only depends on the choice of the substrates included in the analysis.

It is chosen in this work to investigate inter-individual variability by substrate and these results are reported in next session. However, meta-analyzed values of CV for inter-individual variability considering all substrates (CV_{ind}) are also reported in table 2 and still highlight a lower inter-individual variability within PM than within EM. This is consistent with the results of Dorne, Walton et al. (2002) who also show a higher variability when considering EM than when considering PM. These outputs also highlight a lower inter-individual variability within the Asians than within the Caucasians.

Table 2 : Inter-study, inter-substrate and inter-individual variability in pharmacokinetic parameters obtained by Meta-analysis in phenotyped healthy Caucasian and Asian adults.

		Data			Meta-analyse	
PK	**Group**	Nstudy	Nsubstrate	Nindiv	CV inter-study (%) (CV_{st})	CV inter-indiv (%) (CV_{ind})
AUC	**Caucasian EM**	63 (ref1)	26	865	**51** [48;55]	**48** [27;71]
	Caucasian IM	3 (ref2)	3	40		
	Caucasian PM	33 (ref3)	19	225	**53** [47;59]	**35** [28;41]
	Asian EM	31 (ref4)	13	321	**63** [60;67]	**39** [32;45]
	Asian IM	16 (ref5)	12	128	**61** [42;85]	**37** [32;49]
CL	**Caucasian EM**	69 (ref6)	25	706	**58** [54;63]	**47** [29;72]
	Caucasian IM	4 (ref7)	3	38		
	Caucasian PM	31 (ref8)	12	178	**62** [60;63]	**34** [27;39]
	Asian EM	26(ref9)	14	256	**65** [62;67]	**39** [33;46]
	Asian IM	11(ref10)	8	110	**62** [61;64]	**36** [31;41]

Nstudy: number of study, **Nsubstrate**: number of substrates, **Nindiv:** number of individuals.

Ref1 (Kazierad, Lalonde et al. 1989; PRESKORN, ALDERMAN et al. 1994; Salazar, Dockens et al. 1995; Hamelin, Turgeon et al. 1996; Alderman, Preskorn et al. 1997; Dilger, Greiner et al. 1999; Somer, Kallio et al. 2000; Bondolfi, Eap et al. 2002; Hemeryck and Belpaire 2002; Damy, Pousset et al. 2004; Gupta, Banfield et al. 2004; Laine, De Bruyn et al. 2004; Pope, Khalil et al. 2004; Lee, Kim et al. 2006; Pedersen, Damkier et al. 2006; Werner, Lamprecht et al. 2006; Garcia-Quetglas, Azanza et al. 2007; Slanar, Nobilis et al. 2007; Kadokura, den Adel et al. 2008; Malhotra, Guan et al. 2008; Seeringer, Brockmoller et al. 2008; Borobia, Novalbos et al. 2009; Malhotra, Sachse et al. 2009; Preskorn, Patroneva et al. 2009; Simon and Malhotra 2009; Fudio, Borobia et al. 2010; Samer, Daali et al. 2010; Sharma, Pibarot et al. 2010; Nichols, Focht et al. 2011; Sehrt, Meineke et al. 2011; Pardo-Lozano, Farre et al. 2012; Gasso, Papagianni et al. 2013; Lee, Moy et al. 2013)

Ref2 (Lee, Kim et al. 2006; Borobia, Novalbos et al. 2009; Sehrt, Meineke et al. 2011)

Ref3 (Sandberg, Abrahamsson et al. 1993; Hamelin, Turgeon et al. 1996; Dilger, Greiner et al. 1999; Bondolfi, Eap et al. 2002; Giessmann, Modess et al. 2004; Pedersen, Damkier et al. 2006; Werner, Lamprecht et al. 2006; Garcia-Quetglas, Azanza et al. 2007; Slanar, Nobilis et al. 2007; Kadokura, den Adel et al. 2008; Malhotra, Guan et al. 2008; Seeringer, Brockmoller et al. 2008; Borobia, Novalbos et al. 2009; Preskorn, Patroneva et al. 2009; Simon and Malhotra 2009; Malhotra, Crownover et al. 2010; Samer, Daali et al. 2010; Sharma, Pibarot et al. 2010; Nichols, Focht et al. 2011; Sehrt, Meineke et al. 2011; Gasso, Papagianni et al. 2013; Lee, Moy et al. 2013)

Ref4 (Huang, Van Peer et al. 1993; Xu, Xie et al. 1996; Cho and Lee 2006; Kim, Yu et al. 2006; Lee, Kim et al. 2006; Park, Shon et al. 2006; Cui, Teng et al. 2007; Kubo, Koue et al. 2007; Yasui-Furukori, Saito et al. 2007; Jin, Chung et al. 2008; Lim, Cho et al. 2008; Li, Wang et al. 2010; Oishi, Chiba et al. 2010; Yoo, Lee et al. 2011; Azuma, Hasunuma et al. 2012; Yin, Mak et al. 2012; Chen, Xu et al. 2013)

Ref5 (Lee, Kim et al. 2006; Park, Shon et al. 2006; Cui, Teng et al. 2007; Kubo, Koue et al. 2007; Yasui-Furukori, Saito et al. 2007; Jin, Chung et al. 2008; Lim, Cho et al. 2008; Li, Wang et al. 2010; Oishi, Chiba et al. 2010; Yoo, Lee et al. 2011; Azuma, Hasunuma et al. 2012; Yin, Mak et al. 2012; Chen, Xu et al. 2013)

Ref6 (Brøsen, Klysner et al. 1986; Siddoway, Thompson et al. 1987; Boriarti, Strocchi et al. 1990; Jonkers, Koopmans et al. 1991; Turgeon, Uprichard et al. 1991; Bergstrom, Peyton et al. 1992; Brøsen, Hansen et al. 1993; Laurent-Kenesi, Funck-Brentano et al. 1993; Spina, Pollicino et al. 1993; Ujhelyi, O'Rangers et al. 1993; Kurtz, Bergstrom et al. 1997; Amchin, Zarycranski et al. 1999; Brynne, Forslund et al. 1999; Brynne, Svanstrom et al. 1999; Cai, Chen et al. 1999; Lessard, Hamelin et al. 1999; Lessard, Yessine et al. 1999; Lessard, Yessine et al. 1999; Albers, Reist et al. 2000; Dilger, Meisel et al. 2000; Hamelin, Bouayad et al. 2000; Labbe, O'Hara et al. 2000; Graff, Williamson et al. 2001; Lessard, Yessine et al. 2001; Belle, Ernest et al. 2002; Madani, Barilla et al. 2002; Lindh, Annas et al. 2003; Skinner, Kuan et al. 2003; Werner, Werner et al. 2003; Gorski, Huang et al. 2004; Sharma, Pibarot et al. 2005; Werner, Lamprecht et al. 2006; Curry, Watts et al. 2007; Kirchheiner, Keulen et al. 2008; Borobia, Novalbos et al. 2009; Malhotra, Sachse et al. 2009; Preskorn, Patroneva et al. 2009; Fudio, Borobia et al. 2010; Sharma, Pibarot et al. 2010; Nichols, Focht et al. 2011)

Ref7 (Lessard, Yessine et al. 1999; Borobia, Novalbos et al. 2009)

Ref8 (Siddoway, Thompson et al. 1987; Jonkers, Koopmans et al. 1991; Brosen, Hansen et al. 1993; Laurent-Kenesi, Funck-Brentano et al. 1993; Brynne, Forslund et al. 1999; Brynne, Svanstrom et al. 1999; Lessard, Hamelin et al. 1999; Lessard, Yessine et al. 1999; Hamelin, Bouayad et al. 2000; Labbe, O'Hara et al. 2000; Lessard, Yessine et al. 2001; Lindh, Annas et al. 2003; Sharma, Pibarot et al. 2005; Werner, Lamprecht et al. 2006; Kirchheiner, Keulen et al. 2008; Borobia, Novalbos et al. 2009; Malhotra, Sachse et al. 2009; Preskorn, Patroneva et al. 2009; Sharma, Pibarot et al. 2010; Nichols, Focht et al. 2011)

Ref9 (Zhou, Anthony et al. 1990; Xu, Xie et al. 1996; Cho and Lee 2006; Kim, Yu et al. 2006; Lee, Sohn et al. 2006; Cui, Teng et al. 2007; Curry, Watts et al. 2007; Gan, Ismail et al. 2007; Kubo, Koue et al. 2007; Lim, Cho et al. 2008; Azuma, Hasunuma et al. 2012; Yin, Mak et al. 2012; Wu, Yuan et al. 2013)

Ref10 (Lee, Sohn et al. 2006; Cui, Teng et al. 2007; Gan, Ismail et al. 2007; Kubo, Koue et al. 2007; Lim, Cho et al. 2008; Azuma, Hasunuma et al. 2012; Yin, Mak et al. 2012; Wu, Yuan et al. 2013)

3.2.Inter-individual variability and related uncertainty factors

3.2.1 Caucasian population

Meta-analyzed GM_j and CV_j for each substrate in EM and PM Caucasians are reported table 3 regarding the AUC and table 4 regarding the CL. As highlighted by Chiba, Kato et al. (2012), inter-individual variability estimated through the CV varies from one substrate to another. In EM, it ranges from 15% to 74% regarding AUC and from 14% to 71% regarding CL. A significant negative link between AUC and CV was also found which would therefore confirm the findings of Chiba, Kato et al. (2012), namely that the variability of AUC (normalized by the dose) decreases when AUC values increases. Regarding both CL and AUC, CV is lower in PM for many substrates (24 among 31) and the range of CV is much more restraint when considering PM compared to EM. Indeed, in PM, it ranges from 32% to 38% regarding AUC and from 31% to 36% regarding CL.

Table 3 : Meta-analyzed geometric means (GM) and coefficients of variation (CV) of AUC in Caucasian extensive metabolizer (EM) and poor metabolizers (PM) and ratio of GM for each CYP2D6 substrate. Underlined substrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

AUC ng.h.ml^{-1}/dose (mg.kg^{-1})	**EM**		**PM**		**PM/EM**
substrates	**GM**	**CV (%)**	**GM**	**CV (%)**	
Carvedilol	**435** [339 ; 570]	**47** [76 ; 209]	**924** [726 ; 1178]	**35** [32;40]	**2.1** [1.5;3.1]
Citalopram	**3215** [2942 ; 3553]	**13** [11 ; 15]			
Desipramine	**681** [625 ; 740]	**66** [59 ; 74]			
Dextromethorphan	**41** [31 ; 54]	**69** [54 ; 75]			
Encainide	**196** [159 ; 251]	**41** [31 ; 57]			
Fesoterodine	**423** [395 ; 451]	**42** [38 ; 46]	**804** [741 ;875]	**35** [32;39]	**1.9** [1.7;2.1]
Fluoxetine	**2506** [2127 ; 2949]	**64** [53 ; 80]	**3562** [3105 ;4102]	**35** [32;40]	**1.4** [1.1;1.8]
Haloperidol	**257** [207 ; 351]	**38** [30 ; 52]	**348** [278 ;435]	**38** [33;43]	**1.4** [1;1.9]
MDMA	**1366** [1203; 1536]	**29** [24 ; 35]			
Metoprolol	**496** [443 ; 559	**74** [64 ; 87]	**2274** [1716 ;2287]	**33** [29;37]	**4.4** [3.7;5.2]
Mirabegron	**504** [422 ; 607]	**22** [17 ; 28]	**660** [536 ;1259]	**34** [31;38]	**1.3** [1.0;1.7]
Mirtazapine	**1472** [1317 ; 1656]	**11** [9 ; 13]	**1876** [1572 ;2287]	**32** [28;37]	**1.3** [1.0;1.6]
Nefazodone	**1372** [1197 ; 1574]	**47** [40 ; 57]			
Nortriptyline	**2946** [2548 ; 3412]	**17** [14 ; 21]			
Oxycodone	**1005** [882 ; 1144]	**26** [22 ; 31]	**1048** [864 ;1259]	**35** [31;40]	**1.0** [0.8;1.3]
Propafenone	**503** [381 ; 659]	**50** [36 ; 76]	**3861** [2644 ;4263]	**33** [28;37]	**7.6** [4.9;9.7]
R,R-Tramadol	**555** [472 ; 656]	**35** [28 ; 44]	**1177** [901 ;1496]	**34** [30;38	**2.2** [1.6;2.9]
Ramosetron	**2151** [1843 ; 2503]	**40** [33 ; 49]	**2235** [1813 ;2850]	**33** [28;37]	**1.0** [0.8;1.3]
R-carvedilol	**432** [392 ; 472]	**41** [37 ; 46]	**1371** [1154 ;1636]	**35** [31;40]	**3.2** [2.6;3.9]
Risperidone	**1105** [986 ; 1331]	**50** [39 ; 66]	**6415** [5793 ;6893]	**36** [32;41]	**5.8** [4.2;6.7]
R-Metoprolol	**153** [133 ; 180]	**74** [60 ;93]	**1125** [877 ;1453]	**34** [30;38]	**7.2** [5.3;9.7]
R-S-Tramadol	**898** [821 ; 980]	**21** [18 ;25]	**1334** [1183 ;1524]	**32** [29;35]	**1.5** [1.3;1.7]
S,S-Tramadol	**456** [392 ; 533]	**29** [24 ;36]	**938** [736 ;1184]	**35** [31;38]	**2.1** [1.5;2.7]
S-carvedilol	**179** [164 ;196]	**38** [34 ;43]	**338** [284 ;410]	**34** [31;38]	**1.9** [1.5;2.3]
S-Metoprolol	**251** [223 ; 287]	**64** [53;76]	**1586** [1306 ;1975]	**34** [31;38]	**6.3** [4.9;8.0]
Venlaxatine	**412** [334 ; 504]	**59** [45 ;79]	**1818** [1518 ;2156]	**38** [34;44]	**4.3** [3.4;5.8]

Table 4 : Meta-analyzed geometric means (GM) and coefficients of variation (CV) of CL in Caucasian extensive metabolizer (EM) and poor metabolizers (PM) and ratio of GM for each CYP2D6 substrate. Underlined subsrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

CL $mL.h^{-1}.kg^{-1}$	EM		PM		EM/PM
substrates	**GM**	**CV (%)**	**GM**	**CV (%)**	
Atomoxetine	**4** [1 ; 6]	**71** [54 ; 102]			
Citalopram	**354** [247 ; 536]	**15** [13 ; 18]			
Desipramine	**1427** [1294 ; 1515]	**48** [43 ; 54]	**234** [203 ; 269]	**33** [29;37]	**6.1** [5.2;7.2]
Dextromethorphan	**7714** [1090 ; 5833]	**35** [28 ; 45]			
Duloxetine	**703** [364 ; 1324]	**32** [36 ; 42]			
Fesoterodine	**1490** [1128 ; 1974]	**48** [38 ; 60]			
Imipramine	**45** [32 ; 64]	**47** [37 ; 57]			
Metoprolol	**1992** [1539 ; 2312]	**71** [64 ; 80]	**451** [747 ;1092]	**34** [29;37]	**4.4** [3.5;5.2]
Mexiletine	**447** [253; 909]	**43** [33 ; 58]			
Mirtazapine	**657** [411 ; 1127]	**14** [11 ; 17]	**413** [372 ;457]	**34** [29;37]	**1.6** [0.8;2.8]
Oxycodone	**2102** [1479 ; 2742]	**31** [23 ; 42]			
Propafenone	**1261** [894 ; 1635]	**67** [58 ; 78]	**312** [271;409]	**31** [27;35]	**4.0** [2.8;5.3]
R,R-Tramadol	**765** [534 ; 1353]	**37** [30 ; 49]	**264** [211 ; 325]	**32** [27;36]	**3.1** [1.9;4.1]
R-carvedilol	**5** [2 ;9]	**46** [36 ; 63]			
Risperidone	**634** [490 ; 1099]	**71** [58 ; 91]			
R-metoprolol	**750** [447 ; 1237]	**52** [42 ; 68]	**269** [207 ; 421]	**34** [29;37]	**2.8** [1.4;4.6]
R-mexiletine	**451** [272 ; 805]	**32** [25 ; 42]			
R-Propafenone	**1645** [958 ; 2059]	**41** [34 ; 51]			
S,S-Tramadol	**948** [561 ; 1372]	**36** [28 ; 46]	**268** [198 ;379]	**33** [29;37]	**3.3** [1.6;4.8]
S-carvedilol	**18** [9 ; 31]	**51** [38 ;63]			
S-metoprolol	**579** [269 ; 1096]	**49** [39 ;63]	**332** [255 ; 417]	**34** [29;37]	**1.8** [1.2;3.8]
S-mexiletine	**497** [244 ; 533]	**38** [30 ;51]			
S-Propafenone	**512** [364 ;985]	**47** [39 ; 63]	**52** [41 ; 65]	**33** [82;89]	**9.8** [8;25]
Tolterodine	**4272** [1708 ; 2844]	**69** [55 ; 90]	**123** [106 ;147]	**35** [29;37]	**34** [22;46]
Venlafaxine	**1529** [1103 ; 1937]	**71** [62 ; 81]	**363** [333 ; 398]	**36** [23;41]	**4.2** [2.5;4.9]

Ratio between PM and EM (i.e. PM/EM for AUC and EM/PM for CL) are presented in tables 3 and 4. The value of the ratio varies a lot from one substrate to another (from 1 to 34) and it is interesting to notice that major CYP2D6 substrates (i.e. more than 65% of CYP2D6 metabolism) are associated with much higher ratio. This can be explained by the percentage of CYP2D6 metabolism. Indeed, Ito, Hallifax et al. (2005) proposed the following link between this percentage called fm_{CYP2D6} and the ratio:

$$fm_{CYP2D6} = 1 - \frac{CL_{PM}}{CL_{EM}} = 1 - \frac{AUC_{EM}}{AUC_{PM}} \qquad (1)$$

In order to investigate the validity of such link from our data, fm_{CYP2D6} was calculated from this equation and the ratio of tables 3 and 4 and compared to percentage of CYP2D6 metabolism in extensive metabolisers found in the literature. Results of this comparison are reported in table 5. It shows that fm_{CYP2D6} assessed from our results are consistent with fm_{CYP2D6} provided in literature since fm_{CYP2D6} are included in the confidence interval of the prediction for most of the substrates. However, meta-analysed data of clearance seems to allow for a better estimation of the ratio.

Table 5 : Comparison of percentage of CYP2D6 metabolism found in literature ($fm_{literature}$) with percentage of CYP2D6 metabolism calculated from AUC ratio of AUC (fm_{AUC}) and ratio of CL (fm_{CL}).

Subsrate	$fm_{literature}$	reference	fm_{AUC}	fm_{CL}
Desipramine	68	(Dorne, Walton et al. 2002)		**84** [79 ; 89]
Metoprolol	73	(Dorne, Walton et al. 2002)	**77** [72 ;81]	**77** [70 ; 81]
Mirtazapine	35	(Zhou 2009)	**23** [19 ; 29]	**38** [34 ; 42]
Propafenone	>70	(Dorne, Walton et al. 2002)	**87** [79 ; 91]	**75** [71 ; 79]
R,R-Tramadol	60	(Allegaert, Rochette et al. 2011)	**55** [51 ; 60]	**68** [64 ; 72]
R-carvedilol	74		**69** [62 ; 75]	
Risperidone	>60	(Zhou 2009)	**83** [76 ; 86]	
S,S-Tramadol	60	(Allegaert, Rochette et al. 2011)		**65** [61 ; 69]
S-carvedilol	50	(Sehrt, Meineke et al. 2011)	**47** [44 : 51]	
Tolterodine	85	(Dorne, Walton et al. 2002)		**97** [95 ; 98]
Venlafaxine	72	(Dorne, Walton et al. 2002)	**76** [74 ; 78]	**76** [74 ; 78]

Uncertainty factors (UF) calculated to protect 95% or 99% of EM and PM as explained in figure 2 are reported for each substrate in tables 6 and 7. UFs values vary according to the substrate, from 2 to 7 respectively 3 to 14 regarding 95% respectively 99% UFs for EM and from 7 to 40 respectively 14 to 62 for PM. Such inter-substrate differences observed in values of UFs related to EM are only explained by the differences in inter-individual variability. When considering UFs related to PM, such differences are explained by the combination between the ratio PM/EM and inter-individual variability. Consequently, major CYP2D6 substrates present higher UFs values for PM since there is more differences between PM and EM for these substrates. UF calculated to protect 95% and 99% of a simulated whole Caucasian population that is a combination of EM and PM according to the prevalence of each phenotype in the population (see figure 2) are also reported in tables 6 and 7. These UFs range from 3 to 10 when considering 95% UFs and from 4 to 33 when considering 99% UFs. According to these results, the default UF (equal to 3.16) would be sufficient to protect 95% of the population for only 4 substrates among the 50 studied substrates and this default UF would not sufficient to protect 99% of the population for any substrate.

Dorne, Walton et al. (2002) investigated the relationship between the percentage of CYP2D6 metabolism and CYP2D6 related uncertainty factors for poor metabolizers and found an exponential relation. To investigate such relationship from our data and for all the substrates, we propose to calculate $\mathrm{fm}_{\mathrm{CYP2D6}}$ from equation 1 and the ratio of table 2 and 3 taking the uncertainty around the ration into account. The relationship between these $\mathrm{fm}_{\mathrm{CYP2D6}}$ and UFs for poor metabolizers is shown in figure 3. Since UFs are strongly related to the ratio PM/EM which is also strongly related to $\mathrm{fm}_{\mathrm{CYP2D6}}$, this is not surprising to highlight, in our case, a relationship between $\mathrm{fm}_{\mathrm{CYP2D6}}$ and UFs. Indeed a significant exponential relationship (R^2=0.86) is found between $\mathrm{fm}_{\mathrm{CYP2D6}}$ and UFs for PMs. Uncertainty around the estimation of UFs also varies with the percentage of CYP2D6 metabolism. The higher $\mathrm{fm}_{\mathrm{CYP2D6}}$ is, the less there is uncertainty around the estimate which can be explained by the collected data. Indeed, there are more studies conducted on major CYP2D6 substrates in the literature and therefore more data available for these substrates. It thus results in less uncertainty in the model estimates for these substrates.

Table 6 : Uncertainty factors calculated to protect 95% (UF 95) and 99% (UF 99) of the Caucasian EM, PM and simulated Caucasian population from AUC. Calculs and simulations of the whole population are presented in figure 2. Underlined subsrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

AUC	EM		PM		Simulated population	
substrates	UF 95	UF 99	UF 95	UF 99	UF 95	UF 99
Carvedilol	**6** [4 ; 7]	**10** [7 ; 14]	**8** [6 ; 12]	**15** [10;22]	**7** [5 ; 8]	**16** [11 ; 17]
Citalopram	**2** [2 ; 3]	**3** [2 ; 3]				
Desipramine	**6** [5 ; 6]	**13** [11 ; 14]				
Dextromethorphan	**8** [5 ; 11]	**14** [10 ; 18]				
Encainide	**4** [4; 5]	**8** [6 ; 11]				
Fesoterodine	**4** [4 ; 5]	**8** [7 ; 9]	**8** [6 ;8]	**13** [11;15]	**5** [4 ;5]	**9** [8 ;10]
Fluoxetine	**6** [5 ; 7]	**12** [10 ; 16]	**7** [4 ;8]	**11** [7;15]	**6** [5 ;6]	**13** [10 ;15]
Haloperidol	**4** [3 ; 5]	**7** [6 ; 10]	**6** [4 ;8]	**10** [7;15]	**5** [4 ;5]	**10** [7 ;11]
MDMA	**3** [3; 4]	**6** [5; 7]				
Metoprolol	**7** [6 ; 7]	**14** [12 ; 17]	**14** [12 ;18]	**25** [21;32]	**8** [7 ;9]	**18** [15 ;20]
Mirabegron	**3** [2 ; 3]	**5** [4 ; 6]	**5** [3 ;7]	**9** [6;12]	**3** [3 ;3]	**6** [5 ;7]
Mirtazapine	**2** [2 ; 3]	**3** [2 ; 3]	**5** [4 ;6]	**8** [6;10]	**3** [2 ;3]	**4** [3 ;5]
Nefazodone	**5** [4 ; 6]	**9** [8 ; 11]				
Nortriptyline	**2** [2 ; 3]	**4** [3 ; 5]				
Oxycodone	**3** [2 ; 3]	**5** [4 ; 6]	**4** [3 ;5]	**7** [5;9]	**3** [3 ;3]	**6** [5 ;7]
Propafenone	**5** [4 ; 7]	**10** [7 ; 15]	**25** [18 ;37]	**45** [31;67]	**11** [7 ;12]	**33** [17 ;36]
R,R-Tramadol	**4** [3 ; 4]	**7** [6 ; 8]	**8** [6 ;11]	**14** [10;21]	**4** [4;5]	**10** [7;12]
Ramosetron	**4** [3 ; 4]	**8** [7 ; 9]	**4** [3 ;5]	**7** [5;10]	**5** [4 ;5]	**9** [7 ;10]
R-carvedilol	**4** [4 ; 5]	**8** [7 ; 9]	**12** [10 ;15]	**20** [16;26]	**6** [5 ;6]	**11** [9 ;13]
Risperidone	**5** [4 ; 6]	**10** [8 ; 13]	**16** [12 ;21]	**27** [20;37]	**8** [7 ;9]	**17** [12 ;19]
R-Metoprolol	**7** [6 ; 8]	**14** [11 ; 19]	**28** [20 ;38]	**49** [37;67]	**11** [9 ;13]	**27** [20 ;32]
R-S-Tramadol	**3** [2 ; 3]	**4** [4 ; 5]	**5** [4 ;6]	**9** [8;11]	**3** [3 ;3]	**6** [5 ;6]
S,S-Tramadol	**3** [3 ; 4]	**6** [5 ; 7]	**8** [6 ;10]	**14** [10;19]	**4** [4 ;4]	**8** [7 ;10]
S-carvedilol	**4** [4 ;4]	**7** [6 ;8]	**7** [6 ;9]	**13** [10;16]	**5** [4 ;5]	**9** [8;10]
S-Metoprolol	**6** [5 ; 7]	**12** [10 ; 15]	**24** [18 ;31]	**42** [31;55]	**9** [8 ;11]	**22** [17 ;26]
Venlaxatine	**6** [5; 7]	**11** [6 ; 16]	**19** [14 ;25]	**34** [28;46]	**8** [7 ;9]	**20** [14 ;23]

Table 7 : Uncertainty factors calculated to protect 95% (UF 95) and 99% (UF 99) of the Caucasian EM, PM and simulated Caucasian population from CL. Calculs and simulations of the whole population are presented in figure 2.

CL	EM		PM		Simulated population	
substrates	UF 95	UF 99	UF 95	UF 99	UF 95	UF 99
Atomoxetine	**7** [5 ; 8]	**14** [10 ; 20]				
Citalopram	**3** [2; 3]	**4** [3 ; 4]				
Desipramine	**5** [4 ; 5]	**9** [8 ; 10]	**16** [12 ; 23]	**25** [19; 31]	**8** [6;10]	**13** [8;17]
Dextromethorphan	**4** [3 ; 5]	**7** [6 ; 9]				
Duloxetine	**4** [3; 4]	**6** [5 ; 8]				
Fesoterodine	**5** [4 ; 6]	**9** [7 ; 11]				
Imipramine	**5** [4 ; 5]	**9** [7 ; 11]				
Metoprolol	**6** [6 ; 7]	**14** [13 ; 16]	**10** [7 ;12]	**16** [14;18]	**7** [6;8]	**11** [9;15]
Mexiletine	**5** [4; 6]	**8** [6 ; 11]				
Mirtazapine	**2** [2 ; 3]	**3** [3 ; 4]	**3** [2 ;6]	**4** [2;6]	**3** [2;3]	**4** [3;5]
Oxycodone	**4** [3 ; 4]	**6** [5 ; 8]				
Propafenone	**6** [5 ; 7]	**13** [11 ; 15]	**13** [7;20]	**22** [14;32]	**9** [5;12]	**12** [9;16]
R,R-Tramadol	**5** [4 ; 6]	**8** [7 ; 11]	**13** [8 ; 20]	**26** [18;34]	**7** [5;9]	**11** [8;20]
R-carvedilol	**5** [4; 6]	**9** [7 ; 12]				
Risperidone	**6** [5 ; 8]	**13** [11 ; 18]				
R-metoprolol	**5** [4 ; 6]	**10** [8 ; 13]	**10** [6 ;15]	**16** [8;28]	**6** [4;8]	**11** [7;16]
R-mexiletine	**4** [3 ; 5]	**6** [5 ; 8]				
R-Propafenone	**5** [4 ; 6]	**8** [7 ; 10]				
S,S-Tramadol	**4** [3 ; 5]	**7** [6 ; 9]	**8** [5 ;12]	**14** [7;20]	**6** [5;8]	**9** [6;15]
S-carvedilol	**5** [4 ; 6]	**9** [7 ; 12]	**11** [7 ;15]	**18** [11;28]	**5** [4;7]	**10** [7;16]
S-metoprolol	**5** [4 ; 6]	**9** [8 ; 12]	**8** [5 ;16]	**11** [5;16]	**5** [4;6]	**9** [7;12]
S-mexiletine	**4** [3 ; 5]	**7** [6 ; 10]				
S-Propafenone	**5** [4 ;6]	**9** [7 ; 11]	**31** [16 ;43]	**53** [31;70]	**10** [7;16]	**19** [14;29]
Tolterodine	**6** [5 ; 8]	**14** [10 ; 18]	**40** [26 ;62]	**62** [39;76]	**10** [7;14]	**26** [17;36]
Venlafaxine	**6** [6 ; 7]	**14** [12 ; 16]	**15** [10 ;21]	**27** [13;35]	**8** [6;10]	**17** [13;23]

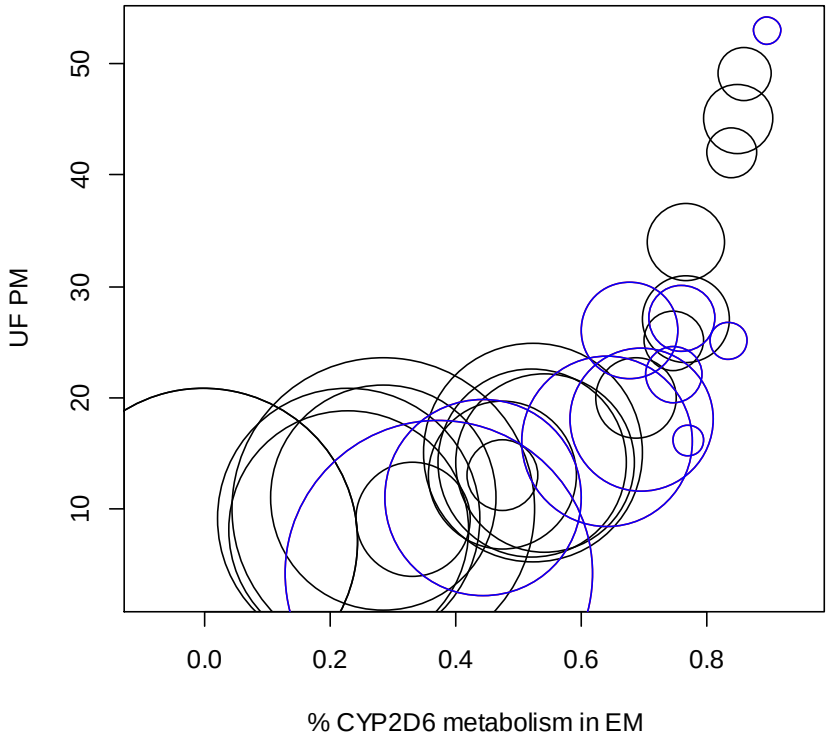


Figure 5 : Relationship between the percentage of CYP2D6 metabolism (fm_{CYP2D6}) in Caucasian extensive metabolisers and the uncertainty factors calculated to protect 99% of the Caucasian PM from the clearance (in blue) or the AUC (in black). Circles represent the uncertainty around the uncertainty factors.

3.2.2 Asian population

Meta-analyzed GM_j and CV_j for each substrate in EM and IM Asians are reported table 8 for AUC and table 9 for CL. As highlighted for Caucasians, inter-individual variability varies from one substrate to another and inter-individual variability is lower within IM than within EM. However, overall inter-individual variability is lower in Asian population than in Caucasian population which could highlight a greater homogeneity in this population.

Ratio between IM and EM (i.e. IM/EM for AUC and EM/IM for CL) are also presented table 8 and 9. The value of the ratio varies a lot from 1 to 4.5. As for PM, major CYP2D6 substrates (i.e. more that 65% of CYP2D6 metabolism) are associated with much higher ratio. However, even for these substrates, the ratio remains much lower than compared to the ratio with PM. This is explained by the activity of allele CYP2D6*10 that is responsible for the phenotype IM in Asian population. Indeed, this allele is associated with a decreased activity of the enzyme but still allows metabolism of the substrates by CYP2D6.

UFs calculated to protect 95% or 99% of EM and IM as explained in figure 2 are reported for each substrate in tables 10 and 11. UFs values vary according to the substrate, from 3 to 5 respectively 6 to 10 regarding 95% respectively 99% UFs for EM (except for Tolderodine which presents a huge CV

and thus UFs of 9 and 22). UFs for PM range from 4 to 13 (95%) and from 7 to 21 (99%). Even if inter-individual variability within IM and EM are lower than in Caucasian population and differences in metabolism between IM and EM are also lower, UFs calculated to protect 95% or 99% of th Asians are still higher than the defaults UF.

Table 8 : Meta-analyzed geometric means (GM) and coefficients of variation (CV) of AUC in Asian extensive metabolizer (EM) and intermediate metabolizers (IM) and ratio of GM for each CYP2D6 substrate. Underlined subsrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

AUC ng.h.ml^{-1}/dose (mg.kg^{-1})	**EM**		**IM**		**IM/EM**
substrates	**GM**	**CV (%)**	**GM**	**CV (%)**	
Aripiprazole	**12871** [11348 ; 14773]	**29** [25 ; 34]	**28489** [24037 ; 30713]	**22** [18 ; 26]	**2.2** [1.9 ; 2.5]
Atomoxetine	**3265** [2411 ; 4394]	**48** [39 ; 63]	**8063** [5344 ; 9151]	**62** [54 ; 71]	**2.5** [2.1 ; 2.9]
Bromperidol	**733** [527 ; 994]	**40** [32 ; 50]	**806** [673 ; 1102]	**33** [28 ; 37]	**1.1** [0.9 ; 1.3]
Citalopram	**4070** [3180 ; 5196]	**33** [28 ; 42]	**6941** [4740 ; 8438]	**25** [21 ; 30]	**1.7** [1.5 ; 1.9]
Desipramine	**734** [533 ; 979]	**41** [33 ; 52]			
Flecanaide	**1990** [1603 ; 2406]	**35** [30 ; 43]	**2485** [1915 ; 3109]	**27** [22 ; 33]	**1.2** [1.0 ; 1.5]
Haloperidol	**284** [207 ; 374]	**45** [37 ; 56]	**327** [273 ; 535]	**87** [82 ; 92]	**1.2** [0.9 ; 1.4]
Lovastatin lactone	**36** [29 ; 46]	**40** [33 ; 49]	**54** [49 ; 65]	**35** [27 ; 39]	**1.5** [1.2 ; 1.8]
Metoprolol	**462** [369; 571]	**40** [33 ; 51]	**1704** [1274 ; 2104]	**24** [21 ; 28]	**3.7** [3.3 ; 4.1]
Paroxetine	**356** [274 ; 480]	**41** [34 ; 51]	**716** [623 ; 902]	**14** [11 ; 19]	**2.0** [1.8 ; 2.2]
Risperidone	**1184** [1062 ; 1337]	**37** [33 ; 40]	**660** [598 ; 747]	**6** [5 ; 7]	**0.6** [0.5 ; 0.8]
R-S-Tramadol	**2364** [2023 ; 2716]	**34** [29 ; 39]	**4140** [3676 ; 4394]	**16** [12 ; 19]	**1.7** [1.5 ; 2.0]
Tolterodine	**773** [666 ; 905]	**108** [94 ; 127]	**3459** [2512 ; 4004]	**41** [36 ; 45]	**4.5** [4.2; 4.8]

Table 9 : Meta-analyzed geometric means (GM) and coefficients of variation (CV) of CL in Asian extensive metabolizer (EM) and intermediate metabolizers (IM) and ratio of GM for each CYP2D6 substrate. Underlined subsrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

CL mL.h^{-1}.kg^{-1}	**EM**		**IM**		**EM/IM**
substrates	**GM**	**CV (%)**	**GM**	**CV (%)**	
Aripiprazole	**60** [54 ; 67]	**35** [30 ; 39]	**35** [30 ; 41]	**22** [18 ; 26]	**1.7** [1.5 ; 2.0]
Atomoxetine	**254** [191 ; 327]	**44** [37 ; 52]	**109** [80 ; 142]	**62** [54 ; 71]	**2.3** [1.6 ; 3.2]
Bromperidol	**1.6** [1.2 ; 2.0]	**40** [34 ; 47]	**1.4** [1 ; 1.8]	**33** [28 ; 37]	**1.2** [0.8 ; 1.7]
Codein	**2.7** [2.3 ; 3.2]	**35** [30 ; 40]	**2.7** [2.1 ; 3.1]	**25** [21 ; 30]	**1.0** [0.7 ; 1.3]
Flecanaide	**486** [399 ; 534]	**38** [33 ; 43]	**383** [283 ; 500]	**27** [22 ; 33]	**1.3** [0.9 ; 1.9]
Imipramine	**9.7** [7.5 ; 13]	**42** [35 ; 49]			
Lovastatin lactone	**10173** [8354 ; 12675]	**41** [35 ; 48]	**10357** [8768 ; 13031]	**87** [82 ; 92]	**0.9** [0.7 ; 1.2]
Nortriptyline	**605** [498 ; 747]	**38** [33 ; 43]	**218** [145 ; 311]	**35** [27 ; 39]	**2.8** [1.8 ; 4.1]
Oxycodone	**1980** [1531; 2503]	**38** [33 ; 42]			
Propranolol	**1248** [952 ; 1646]	**39** [33 ; 45]			
Propranolol(-)-	**1130** [854 ; 1480]	**39** [33 ; 45]			
Propranolol(+)-	**1291** [972 ; 1765]	**39** [33 ; 45]			
risperidone	**797** [665 ; 940]	**54** [45 ; 66]			
Tramadol	**292** [263 ; 328]	**37** [34 ; 40]	**330** [192 ; 272]	**41** [36 ; 45]	**1.4** [1; 1.8]

Table 10 : Uncertainty factors calculated to protect 95% (UF 95) and 99% (UF 99) of the Asian EM, IM and simulated Asian population from AUC. Calculs and simulations are presented in figure 2. Underlined subsrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

AUC	**EM**		**IM**		**Simulated population**	
substrates	**UF 95**	**UF 99**	**UF 95**	**UF 99**	**UF 95**	**UF 99**
Aripiprazole	**4** [3 ; 4]	**6** [5 ; 7]	**6** [4 ; 8]	**10** [7;14]	**6** [5 ; 6]	**9** [7 ; 11]
Atomoxetine	**5** [4 ; 6]	**9** [7 ; 12]	**10** [8 ; 13]	**20** [17;24]	**9** [6 ;12]	**16** [12 ;18]
Bromperidol	**4** [4 ; 5]	**8** [6 ; 9]	**4** [3 ; 6]	**7** [6 ; 9]	**4** [4 ;5]	**6** [4 ;8]
Citalopram	**4** [3 ; 4]	**7** [6 ; 8]	**5** [4 ; 8]	**10** [7;15]	**5** [4 ;6]	**9** [7 ;11]
Desipramine	**4** [3 ; 4]	**8** [6 ; 11]				
Flecanaide	**4** [3 ; 5]	**8** [7 ; 9]	**5** [3 ; 8]	**7** [4;11]	**5** [4 ;7]	**8** [8 ;10]
Haloperidol	**3** [3 ; 4]	**7** [6 ; 8]	**8** [5 ; 11]	**17** [11 ; 22]	**6** [5 ;6]	**13** [10 ;15]
Lovastatin lactone	**4** [3 ; 5]	**9** [7 ; 11]	**6** [4 ; 8]	**11** [7;15]	**6** [4 ;7]	**10** [7 ;11]
Metoprolol	**4** [3 ; 5]	**8** [7; 9]	**13** [9 ; 18]	**21** [14; 31]	**12** [9 ;14]	**20** [15 ;24]
Paroxetine	**4** [3 ; 5]	**8** [7 ; 10]	**7** [5 ; 9]	**11** [8;17]	**7** [5 ;9]	**13** [11 ;15]
Risperidone	**4** [3 ; 4]	**7** [6 ; 8]	**6** [3; 8]	**9** [6; 11]	**6** [5 ;6]	**8** [7 ;10]
R-S-Tramadol	**4** [3 ; 4]	**7** [6 ; 8]	**5** [3 ; 8]	**8** [6; 11]	**5** [4 ;6]	**8** [5 ;11]
Tolterodine	**9** [8 ; 10]	**22** [19 ; 26]	**18** [13 ;22]	**36** [26;54]	**16** [11 ;18]	**33** [24 ;38]

Table 11 : Uncertainty factors calculated to protect 95% (UF 95) and 99% (UF 99) of the Asian EM, IM and simulated Asian population from CL. Calculs and simulations are presented in figure 2. Underlined subsrates are major CYP2D6 substrates (fm_{CYP2D6}>60%).

CL	**EM**		**IM**		**Simulated population**	
substrates	**UF 95**	**UF 99**	**UF 95**	**UF 99**	**UF 95**	**UF 99**
Aripiprazole	**4** [3 ; 5]	**6** [5 ; 7]	**7** [5 ; 9]	**12** [10 ; 17]	**5** [4 ; 6]	**9** [8 ; 11]
Atomoxetine	**5** [3 ; 6]	**8** [6 ; 9]	**9** [6 ; 14]	**17** [11 ; 26]	**8** [6 ; 11]	**15** [11 ; 21]
Bromperidol	**4** [3 ; 5]	**8** [7 ; 10]	**5** [3 ; 7]	**8** [5 ; 13]	**5** [4 ; 6]	**8** [7 ; 12]
Codein	**4** [3 ; 5]	**7** [6 ; 8]	**4** [3 ; 6]	**7** [5 ; 10]	**4** [3 ; 5]	**7** [6 ; 9]
Flecanaide	**4** [3 ; 5]	**7** [6 ; 8]	**5** [4 ; 7]	**8** [6 ; 13]	**5** [4 ; 6]	**8** [7 ; 11]
Imipramine	**4** [3 ; 5]	**8** [6 ; 9]				
Lovastatin lactone	**4** [3 ; 5]	**8** [7 ; 9]	**4** [3 ; 5]	**7** [6 ; 9]	**4** [3 ; 5]	**8** [7 ; 9]
Nortriptyline	**4** [3 ; 5]	**7** [5 ; 7]	**11** [7 ; 17]	**19** [12 ; 31]	**8** [6 ; 12]	**15** [10 ; 24]
Oxycodone	**4** [3 ; 5]	**7** [5 ; 9]				
Propranolol	**4** [3 ; 5]	**7** [6 ; 8]				
Propranolol(-)-	**4** [3 ; 5]	**7** [6 ; 8]				
Propranolol(+)-	**4** [3 ; 5]	**7** [7 ; 9]				
Risperidone	**5** [4 ; 7]	**10** [9 ; 13]				
Tramadol	**4** [3 ; 5]	**7** [6 ; 9]	**5** [4 ; 7]	**9** [6 ; 13]	**5** [4 ; 6]	**8** [7 ; 9]

4. Discussion

This work provided an analysis of inter-individual differences in CYP2D6 metabolism from a large database related to chronic oral exposure (AUC and clearance). A hierarchical Bayesian model was proposed to meta-analyze these pharmacokinetic data. Such approach made it possible to account for the different sample sizes of studies and their heterogeneity. Moreover, inter-study and inter-substrate variability were quantified so that strength can be borrowed from a study to another and from a compound to another.

In this work, a special focus was made on Caucasian EM and PM and Asian EM and IM. Inter-individual variability after oral exposure differs from one group to another. A high variability was highlighted for Caucasian EM (CV around 50%) which is in concordance with the previous study of Dorne, Walton et al. (2002). Such variability in metabolism can be explained by different factors. First, in this work, were considered as EM individuals, either homozygote with normal activity alleles, either heterozygote with one allele associated with normal activity and one allele with decreased activity. Consequently, there is some heterogeneity in enzymatic activity within the EM group related to this gathering. Second, Daneshmend, Jackson et al. (1981) highlighted some inter-individual and intra-individual variability in liver blood flow. As suggested in Dorne, Walton et al. (2002), such differences can explain some of the inter-individual variability observed after oral exposure, especially for compounds with high clearance (or low AUC). Finally, part of the inter-individual variability observed in this work may also be related to inter-individual differences in plasma protein binding which was not investigated in this work. Significant differences in inter-individual variability between substrates were also observed especially regarding EM as it was already highlighted in Chiba, Kato et al. (2012). A significant relation was found between the variability and the value of the AUC. However, in order to subsequently be able to predict chemical specific variability and associated uncertainty factor, a deeper investigation of such link would be of great interest. Especially, protein binding may also be a factor that can explain observed differences in variability between substrates. It could therefore be interesting, after collecting information on protein binding, to model the meta-analyzed CV as function of AUC (or CL) values and protein binding rate. A more powerful way to model such relation could also be to directly add protein binding as a variable in the hierarchical model for meta-analysis and to model CV as a function of Geometric mean and protein binding for example. A slightly weaker inter-individual variability was found within Caucasian PM (CV around 35%) which is also in concordance with previous study (Dorne, Walton et al. 2002). This difference with EM can be partly explained by a greater homogeneity in enzyme activity within the PM group. Indeed, in this work was considered as PM, individuals with 2 alleles associated with non activity which may induce no activity or very little CYP2D6 function in every case. Moreover, in

Caucasian PM phenotype is mainly caused by the same allele (i.e. CYP2D6*4) which may increase the homogeneity in enzyme activity. Additionally, since clearances are lower, liver blood flow is less driving the inter-individual variability for PM than for EM. Due to the weak or even no activity of CYP2D6 in PM, metabolism of CYP2D6 substrates can be carried by other enzymes. If this pathway of substitution is not polymorphic, there would be less variability. On contrary, if the other pathway is also polymorphic, then some variability would remain. A slightly weaker inter-individual variability was also found within Asian population than within Caucasian population (CV around 40% for EM and IM) which may reflects a greater genetic homogeneity within this ethnic group.

Uncertainty factors were calculated for each of these phenotypic and ethnic groups in order to investigate the adequacy of the 3.16 toxicokinetic default uncertainty factor used in risk assessment for these groups. The first result is that for none of these groups, the default uncertainty factor is sufficient to protect 95% of the population. From our analysis and to cover all substrates, a factor of 11 (respectively 33) would be necessary to protect 95% (respectively 99%) of the Caucasian population taking the prevalence of PM into account. A factor of 11 (respectively 33) would be necessary to protect 95% (respectively 99%) of the Caucasian group taking the prevalence of PM into account. A factor of 11 (respectively 20) would be necessary to protect 95% (respectively 99%) of the Asian group taking the prevalence of IM into account. Therefore, a CYP2D6 related uncertainty factors of 11 could be proposed to cover 95% of the population whatever the compound. Such uncertainty factor would better apply to chemicals eliminated primarily by CYP2D6. Indeed, as it was observed in this work, when substrates present a lower percentage of CYP2D6 metabolism, related UF are also lower. It may thus be useful for risk assessment to move from a CYP2D6-related factor to a chemical-specific factor. As a first step toward such factors, we investigated the relationship between the percentage of CYP2D6 metabolism in EM and the UFs and found a significant exponential relationship shown figure 3. Since, in vitro techniques are now able to identify the CYP isoform(s) responsible for the metabolism (Abass 2013), such relationship could be used to establish chemical-specific factors for risk assessment. Zhu, Sedykh et al. (2013) also show that protein binding can be predicted by in vitro methods which would allow to integrate this factor in the modeling of the variability.

Currently a number of toxicokinetic models are available which describe the relationship between external exposure level and chemical concentration in biological matrices over time. Such models and especially physiologically-based toxicokinetic (PBTK) models make it possible to better understand and predict the determinants of inter-individual variability in TK when the required information is adequately incorporated. Bois et al. (2010) reviewed the prediction and sources of inter-individual variability in ADME (Absorption, Distribution, Metabolism, Excretion) parameters

that can be input in a PBTK model. Among them, they highlight the variation resulting from the existence of subpopulations that differ in some way from the 'normal' population due to genetic polymorphisms. Therefore, the inter-individual variability provided by this work could be useful to inform PBTK model and thus better take the variability related to polymorphism into account in the modeling. However, before extrapolating the result of this work to exposure to chemicals, the question of the extrapolation to different doses should be raised. Indeed, all the data used in this work come from pharmaceutical studies and thus correspond to pharmaceutical doses (e.g. in mg) whereas environmental and dietary exposure to chemicals may occur at lower doses (e.g. in ng, µg). Some studies were recently conducted on micro doses. An analysis of such studies could reveal whether the observed variability in pharmaceutical dose is the same as that observed at lower doses.

The presented analysis on uncertainty factor was made assuming that PMs are the most susceptible group. This is true when the parent compound is the toxicant. Indeed, in this case, an increase in internal dose would increase the risk of adverse effect. On the reverse way, if the toxicant is a CYP2D6-mediated metabolite, PMs would have lower internal dose of the toxicant and thus be less at risk. In this case, the EMs would be at higher risk and the most susceptible group would be the ultra-rapid metabolizers (UMs). Some pharmacokinetic data for UM phenotype were also collected and around 2 fold factor was found between UM and EM (Gan, Ismail et al. 2007; Kirchheiner, Keulen et al. 2008; Seeringer, Brockmoller et al. 2008; Samer, Daali et al. 2010; Gasso, Papagianni et al. 2013).

Finally, in order to provide a comprehensive analysis of the human variability other analyses should be combined. As highlighted by Martin, Martin et al. (2013) about uncertainty factors in general and by Dorne, Walton et al. (2002) for CYP2D6, some other subgroups of population can also be more sensible e.g. neonates, infant, elderly and specific factors may thus be necessary to protect them. Moreover, CYP2D6 is one polymorphic enzyme among others that participates to the metabolic pathway. Therefore, to have complete information, human variability related to the other pathways need to be investigated and combined as suggested in Dorne (2010). Finally, all the approach proposed in this work and the hierarchical model for meta-analysis could be also useful to investigate inter-individual variability in toxicodynamic in order to better take human variability into account in the whole risk assessment process.

References

Abass, K. M. (2013). "From< i> in vitro</i> hepatic metabolic studies towards human health risk assessment: Two case studies of diuron and carbosulfan." Pesticide Biochemistry and Physiology **107**(2): 258-265.

Albers, L. J., C. Reist, et al. (2000). "Effect of venlafaxine on imipramine metabolism." Psychiatry Res **96**(3): 235-243.

Alderman, J., S. H. Preskorn, et al. (1997). "Desipramine Pharmacokinetics When Coadministered With Paroxetine or Sertraline in Extensive Metabolizers." J Clin Psychopharmacol **17**(4): 284-291.

Allegaert, K., A. Rochette, et al. (2011). "Developmental pharmacology of tramadol during infancy: ontogeny, pharmacogenetics and elimination clearance." Paediatr Anaesth **21**(3): 266-273.

Amchin, J., W. Zarycranski, et al. (1999). "Effect of venlafaxine on the pharmacokinetics of risperidone." J Clin Pharmacol **39**(3): 297-309.

Azuma, J., T. Hasunuma, et al. (2012). "The relationship between clinical pharmacokinetics of aripiprazole and CYP2D6 genetic polymorphism: effects of CYP enzyme inhibition by coadministration of paroxetine or fluvoxamine." Eur J Clin Pharmacol **68**(1): 29-37.

Barter, Z. E., G. T. Tucker, et al. (2013). "Differences in Cytochrome P450-Mediated Pharmacokinetics Between Chinese and Caucasian Populations Predicted by Mechanistic Physiologically Based Pharmacokinetic Modelling." Clin Pharmacokinet.

Belle, D. J., C. S. Ernest, et al. (2002). "Effect of potent CYP2D6 inhibition by paroxetine on atomoxetine pharmacokinetics." J Clin Pharmacol **42**(11): 1219-1227.

Bergstrom, R. F., A. L. Peyton, et al. (1992). "Quantification and mechanism of the fluoxetine and tricyclic antidepressant interaction." Clin. Pharm. Ther. **51**(3): 239-248.

Bertilsson, L. (2010). "CYP2D6, serotonin, and suicide—a relationship?" Clinical Pharmacology & Therapeutics **88**(3): 304-305.

Blake, C. M., E. D. Kharasch, et al. (2013). "A meta-analysis of CYP2D6 metabolizer phenotype and metoprolol pharmacokinetics." Clin Pharmacol Ther **94**(3): 394-399.

Bondolfi, G., C. B. Eap, et al. (2002). "The effect of fluoxetine on the pharmacokinetics and safety of risperidone in psychotic patients." Pharmacopsychiatry **35**(2): 50-56.

Boriarti, G., E. Strocchi, et al. (1990). "Relationships Between Debrisoquine Hydroxylation and Propafenone Pharmacokinetics." Drug Investigation **2**(2): 114-119.

Borobia, A. M., J. Novalbos, et al. (2009). "Influence of sex and CYP2D6 genotype on mirtazapine disposition, evaluated in Spanish healthy volunteers." Pharmacol Res **59**(6): 393-398.

Brosen, K., J. G. Hansen, et al. (1993). "Inhibition by paroxetine of desipramine metabolism in extensive but not in poor metabolizers of sparteine." Eur J Clin Pharmacol **44**(4): 349-355.

Brøsen, K., J. G. Hansen, et al. (1993). "Inhibition by paroxetine of desipramine metabolism in extensive but not in poor metabolizers of sparteine." European Journal of Clinical Pharmacology **44**(4): 349-355.

Brøsen, K., R. Klysner, et al. (1986). "Steady-state concentrations of imipramine and its metabolites in relation to the sparteine/debrisoquine polymorphism." European Journal of Clinical Pharmacology **30**(6): 679-684.

Brynne, N., C. Forslund, et al. (1999). "Ketoconazole inhibits the metabolism of tolterodine in subjects with deficient CYP2D6 activity." Br J Clin Pharmacol **48**(4): 564-572.

Brynne, N., C. Svanstrom, et al. (1999). "Fluoxetine inhibits the metabolism of tolterodine-pharmacokinetic implications and proposed clinical relevance." Br J Clin Pharmacol **48**(4): 553-563.

Cai, W. M., B. Chen, et al. (1999). "The influence of CYP2D6 activity on the kinetics of propafenone enantiomers in Chinese subjects." Br J Clin Pharmacol **47**(5): 553-556.

Calabrese, E. J. (1985). "Uncertainty factors and interindividual variation." Regulatory Toxicology and Pharmacology **5**(2): 190-196.
Chen, B., Y. Xu, et al. (2013). "Estimation of CYP2D6*10 genotypes on citalopram disposition in Chinese subjects by population pharmacokinetic assay." J Clin Pharm Ther.
Chiba, K., M. Kato, et al. (2012). "Inter-individual variability of in vivo CYP2D6 activity in different genotypes." Drug Metab Pharmacokinet **27**(4): 405-413.
Cho, H. Y. and Y. B. Lee (2006). "Pharmacokinetics and bioequivalence evaluation of risperidone in healthy male subjects with different CYP2D6 genotypes." Arch Pharm Res **29**(6): 525-533.
Cui, Y. M., C. H. Teng, et al. (2007). "Atomoxetine pharmacokinetics in healthy Chinese subjects and effect of the CYP2D6*10 allele." Br J Clin Pharmacol **64**(4): 445-449.
Curry, S. C., D. J. Watts, et al. (2007). "The effect of single-dose tramadol on oxycodone clearance." J Emerg Med **33**(4): 407-411.
Damy, T., F. Pousset, et al. (2004). "Pharmacokinetic and pharmacodynamic interactions between metoprolol and dronedarone in extensive and poor CYP2D6 metabolizers healthy subjects." Fundam Clin Pharmacol **18**(1): 113-123.
Daneshmend, T. K., L. Jackson, et al. (1981). "Physiological and pharmacological variability in estimated hepatic blood flow in man." British Journal of Clinical Pharmacology **11**(5): 491-496.
Dilger, K., B. Greiner, et al. (1999). "Consequences of rifampicin treatment on propafenone disposition in extensive and poor metabolizers of CYP2D6." Pharmacogenetics **9**(5): 551-559.
Dilger, K., P. Meisel, et al. (2000). "Disposition of propafenone in a poor metabolizer of CYP2D6 with Gilbert's syndrome." Ther Drug Monit **22**(3): 366-368.
Dorne, J. (2010). "Metabolism, variability and risk assessment." Toxicology **268**(3): 156-164.
Dorne, J., K. Walton, et al. (2001). "Uncertainty factors for chemical risk assessment: human variability in the pharmacokinetics of CYP1A2 probe substrates." Food and Chemical Toxicology **39**(7): 681-696.
Dorne, J., K. Walton, et al. (2003). "Human variability in CYP3A4 metabolism and CYP3A4-related uncertainty factors for risk assessment." Food and Chemical Toxicology **41**(2): 201-224.
Dorne, J. L., K. Walton, et al. (2005). "Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review." Food Chem Toxicol **43**(2): 203-216.
Dorne, J. L., K. Walton, et al. (2002). "Human variability in polymorphic CYP2D6 metabolism: is the kinetic default uncertainty factor adequate?" Food Chem Toxicol **40**(11): 1633-1656.
Dorne, J. L. C. M., K. Walton, et al. (2001). "Human variability in glucuronidation in relation to uncertainty factors for risk assessment." Food and Chemical Toxicology **39**(12): 1153-1173.
Dorne, J. L. C. M., K. Walton, et al. (2003). "Polymorphic CYP2C19 and N-acetylation: human variability in kinetics and pathway-related uncertainty factors." Food and Chemical Toxicology **41**(2): 225-245.
Dorne, J. L. C. M., K. Walton, et al. (2004). "Human variability for metabolic pathways with limited data (CYP2A6, CYP2C9, CYP2E1, ADH, esterases, glycine and sulphate conjugation)." Food and Chemical Toxicology **42**(3): 397-421.
Eichelbaum, M. (2003). "In search of endogenous CYP2D6 substrates." Pharmacogenet Genomics **13**(6): 305-306.
Fudio, S., A. M. Borobia, et al. (2010). "Evaluation of the influence of sex and CYP2C19 and CYP2D6 polymorphisms in the disposition of citalopram." Eur J Pharmacol **626**(2-3): 200-204.
Gan, S. H., R. Ismail, et al. (2007). "Impact of CYP2D6 genetic polymorphism on tramadol pharmacokinetics and pharmacodynamics." Mol Diagn Ther **11**(3): 171-181.
Garcia-Quetglas, E., J. R. Azanza, et al. (2007). "Pharmacokinetics of tramadol enantiomers and their respective phase I metabolites in relation to CYP2D6 phenotype." Pharmacol Res **55**(2): 122-130.

Gasso, P., K. Papagianni, et al. (2013). "Relationship between CYP2D6 genotype and haloperidol pharmacokinetics and extrapyramidal symptoms in healthy volunteers." Pharmacogenomics **14**(13): 1551-1563.
Giessmann, T., C. Modess, et al. (2004). "CYP2D6 genotype and induction of intestinal drug transporters by rifampin predict presystemic clearance of carvedilol in healthy subjects." Clin Pharmacol Ther **75**(3): 213-222.
Ginsberg, G., D. Hattis, et al. (2002). "Evaluation of Child/Adult Pharmacokinetic Differences from a Database Derived from the Therapeutic Drug Literature." Toxicological Sciences **66**(2): 185-200.
González, I., E. M. Peñas-Lledó, et al. (2008). "Relation between CYP2D6 phenotype and genotype and personality in healthy volunteers."
Gorski, J. C., S. M. Huang, et al. (2004). "The effect of echinacea (Echinacea purpurea root) on cytochrome P450 activity in vivo." Clin Pharmacol Ther **75**(1): 89-100.
Graff, D. W., K. M. Williamson, et al. (2001). "Effect of fluoxetine on carvedilol pharmacokinetics, CYP2D6 activity, and autonomic balance in heart failure patients." J Clin Pharmacol **41**(1): 97-106.
Gupta, S., C. Banfield, et al. (2004). "Pharmacokinetics/pharmacodynamics of desloratadine and fluoxetine in healthy volunteers." J Clin Pharmacol **44**(11): 1252-1259.
Hamelin, B. A., A. Bouayad, et al. (2000). "Significant interaction between the nonprescription antihistamine diphenhydramine and the CYP2D6 substrate metoprolol in healthy men with high or low CYP2D6 activity." Clin Pharmacol Ther **67**(5): 466-477.
Hamelin, B. A., J. Turgeon, et al. (1996). "The disposition of fluoxetine but not sertraline is altered in poor metabolizers of debrisoquin." Clin Pharmacol Ther **60**(5): 512-521.
Hemeryck, A. and F. M. Belpaire (2002). "Selective serotonin reuptake inhibitors and cytochrome P-450 mediated drug-drug interactions: an update." Curr Drug Metab **3**(1): 13-37.
Huang, M. L., A. Van Peer, et al. (1993). "Pharmacokinetics of the novel antipsychotic agent risperidone and the prolactin response in healthy subjects." Clin Pharmacol Ther **54**(3): 257-268.
Ito, K., D. Hallifax, et al. (2005). "Impact of parallel pathways of drug elimination and multiple cytochrome P450 involvement on drug-drug interactions: CYP2D6 paradigm." Drug metabolism and disposition **33**(6): 837-844.
Jin, S. K., H. J. Chung, et al. (2008). "Influence of CYP2D6*10 on the pharmacokinetics of metoprolol in healthy Korean volunteers." J Clin Pharm Ther **33**(5): 567-573.
Jonkers, R. E., R. P. Koopmans, et al. (1991). "Debrisoquine phenotype and the pharmacokinetics and beta-2 receptor pharmacodynamics of metoprolol and its enantiomers." Journal of Pharmacology and Experimental Therapeutics **256**(3): 959-966.
Kadokura, T., M. den Adel, et al. (2008). "The effect of paroxetine on the pharmacokinetics, safety, and tolerability of ramosetron in healthy subjects." Eur J Clin Pharmacol **64**(6): 605-609.
Kazierad, D. J., R. L. Lalonde, et al. (1989). "The effect of diltiazem on the disposition of encainide and its active metabolites." Clin. Pharm. Ther. **46**(6): 668-673.
Kim, E., K. S. Yu, et al. (2006). "Effects of DRD2 and CYP2D6 genotypes on delta EEG power response to aripiprazole in healthy male volunteers: a preliminary study." Hum Psychopharmacol **21**(8): 519-528.
Kirchheiner, J., J. T. Keulen, et al. (2008). "Effects of the CYP2D6 gene duplication on the pharmacokinetics and pharmacodynamics of tramadol." J Clin Psychopharmacol **28**(1): 78-83.
Kubo, M., T. Koue, et al. (2007). "Pharmacokinetics of aripiprazole, a new antipsychotic, following oral dosing in healthy adult Japanese volunteers: influence of CYP2D6 polymorphism." Drug Metab Pharmacokinet **22**(5): 358-366.
Kuhn, U. D., M. Kirsch, et al. (2007). "Reboxetine and cytochrome P450--comparison with paroxetine treatment in humans." Int J Clin Pharmacol Ther **45**(1): 36-46.
Kurtz, D. L., R. F. Bergstrom, et al. (1997). "The effect of sertraline on the pharmacokinetics of desipramine and imipramine." Clin Pharmacol Ther **62**(2): 145-156.

Labbe, L., G. O'Hara, et al. (2000). "Pharmacokinetic and pharmacodynamic interaction between mexiletine and propafenone in human beings." Clin Pharmacol Ther **68**(1): 44-57.
Laine, K., S. De Bruyn, et al. (2004). "Effect of the novel anxiolytic drug deramciclane on cytochrome P 450 2D6 activity as measured by desipramine pharmacokinetics." European Journal of Clinical Pharmacology **59**(12): 893-898.
Laurent-Kenesi, M. A., C. Funck-Brentano, et al. (1993). "Influence of CYP2D6-dependent metabolism on the steady-state pharmacokinetics and pharmacodynamics of metoprolol and nicardipine, alone and in combination." Br J Clin Pharmacol **36**(6): 531-538.
Lee, J., S. Moy, et al. (2013). "Role of cytochrome p450 isoenzymes 3A and 2D6 in the in vivo metabolism of mirabegron, a beta3-adrenoceptor agonist." Clin Drug Investig **33**(6): 429-440.
Lee, S. Y., Y. G. Kim, et al. (2006). "Pharmacokinetic parameters of bromperidol in Korean subjects." Hum Psychopharmacol **21**(6): 409-412.
Lee, S. Y., K. M. Sohn, et al. (2006). "Sequence-based CYP2D6 genotyping in the Korean population." Ther Drug Monit **28**(3): 382-387.
Lessard, E., B. A. Hamelin, et al. (1999). "Involvement of CYP2D6 activity in the N-oxidation of procainamide in man." Pharmacogenetics **9**(6): 683-696.
Lessard, É., M.-A. Yessine, et al. (1999). "Influence of CYP2D6 activity on the disposition and cardiovascular toxicity of the antidepressant agent venlafaxine in humans." Pharmacogenet Genomics **9**(4): 435-443.
Lessard, E., M. A. Yessine, et al. (2001). "Diphenhydramine alters the disposition of venlafaxine through inhibition of CYP2D6 activity in humans." J Clin Psychopharmacol **21**(2): 175-184.
Lessard, E., M. A. Yessine, et al. (1999). "Influence of CYP2D6 activity on the disposition and cardiovascular toxicity of the antidepressant agent venlafaxine in humans." Pharmacogenetics **9**(4): 435-443.
Li, Q., R. Wang, et al. (2010). "Relationship of CYP2D6 genetic polymorphisms and the pharmacokinetics of tramadol in Chinese volunteers." J Clin Pharm Ther **35**(2): 239-247.
Lim, K. S., J. Y. Cho, et al. (2008). "Pharmacokinetic interaction of flecainide and paroxetine in relation to the CYP2D6*10 allele in healthy Korean subjects." Br J Clin Pharmacol **66**(5): 660-666.
Lindh, J. D., A. Annas, et al. (2003). "Effect of ketoconazole on venlafaxine plasma concentrations in extensive and poor metabolisers of debrisoquine." Eur J Clin Pharmacol **59**(5-6): 401-406.
Løkke, H. (2010). "Novel methods for integrated risk assessment of cumulative stressors—results from the NoMiracle project." Science of the Total Environment **408**(18): 3719-3724.
Madani, S., D. Barilla, et al. (2002). "Effect of terbinafine on the pharmacokinetics and pharmacodynamics of desipramine in healthy volunteers identified as cytochrome P450 2D6 (CYP2D6) extensive metabolizers." J Clin Pharmacol **42**(11): 1211-1218.
Malhotra, B., E. Darsey, et al. (2011). "Comparison of pharmacokinetic variability of fesoterodine vs. tolterodine extended release in cytochrome P450 2D6 extensive and poor metabolizers." Br J Clin Pharmacol **72**(2): 226-234.
Malhotra, B., Z. Guan, et al. (2008). "Pharmacokinetic profile of fesoterodine." Int J Clin Pharmacol Ther **46**(11): 556-563.
Malhotra, B., R. Sachse, et al. (2009). "Evaluation of drug-drug interactions with fesoterodine." Eur J Clin Pharmacol **65**(6): 551-560.
Malhotra, B. K., P. H. Crownover, et al. (2010). "The pharmacokinetic profile of fesoterodine 8 mg with daytime or nighttime dosing." Eur J Clin Pharmacol **66**(2): 171-176.
Malhotra, B. K., N. Wood, et al. (2009). "Influence of age, gender, and race on pharmacokinetics, pharmacodynamics, and safety of fesoterodine." Int J Clin Pharmacol Ther **47**(9): 570-578.
Martin, O. V., S. Martin, et al. (2013). "Dispelling urban myths about default uncertainty factors in chemical risk assessment–sufficient protection against mixture effects?" Environmental Health **12**(1): 53.
Naumann, B. D., K. C. Silverman, et al. (2001). "Case studies of categorical data-derived adjustment factors." Human and Ecological Risk Assessment: An International Journal **7**(1): 61-105.

Naumann, B. D., P. A. Weideman, et al. (1997). "Use of toxicokinetic and toxicodynamic data to reduce uncertainties when setting occupational exposure limits for pharmaceuticals." Human and Ecological Risk Assessment **3**(4): 555-565.
Nichols, A. I., K. Focht, et al. (2011). "Pharmacokinetics of venlafaxine extended release 75 mg and desvenlafaxine 50 mg in healthy CYP2D6 extensive and poor metabolizers: a randomized, open-label, two-period, parallel-group, crossover study." Clin Drug Investig **31**(3): 155-167.
Oishi, M., K. Chiba, et al. (2010). "Effect of the CYP2D6*10 genotype on tolterodine pharmacokinetics." Drug Metab Dispos **38**(9): 1456-1463.
Pardo-Lozano, R., M. Farre, et al. (2012). "Clinical pharmacology of 3,4-methylenedioxymethamphetamine (MDMA, "ecstasy"): the influence of gender and genetics (CYP2D6, COMT, 5-HTT)." PLoS One **7**(10): e47599.
Park, J. Y., J. H. Shon, et al. (2006). "Combined effects of itraconazole and CYP2D6*10 genetic polymorphism on the pharmacokinetics and pharmacodynamics of haloperidol in healthy subjects." J Clin Psychopharmacol **26**(2): 135-142.
Pedersen, R. S., P. Damkier, et al. (2006). "Enantioselective pharmacokinetics of tramadol in CYP2D6 extensive and poor metabolizers." Eur J Clin Pharmacol **62**(7): 513-521.
Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March.
Plummer, M., N. Best, et al. (2006). "CODA: Convergence diagnosis and output analysis for MCMC." R news **6**(1): 7-11.
Pope, L. E., M. H. Khalil, et al. (2004). "Pharmacokinetics of Dextromethorphan After Single or Multiple Dosing in Combination With Quinidine in Extensive and Poor Metabolizers." The Journal of Clinical Pharmacology **44**(10): 1132-1142.
Preskorn, S., A. Patroneva, et al. (2009). "Comparison of the pharmacokinetics of venlafaxine extended release and desvenlafaxine in extensive and poor cytochrome P450 2D6 metabolizers." J Clin Psychopharmacol **29**(1): 39-43.
PRESKORN, S. H., J. ALDERMAN, et al. (1994). "Pharmacokinetics of Desipramine Coadministered With Sertraline or Fluoxetine." J Clin Psychopharmacol **14**(2): 90-98.
Renwick, A. G. (1993). "DATA-DERIVED SAFETY FACTORS FOR THE EVALUATION OF FOOD-ADDITIVES AND ENVIRONMENTAL CONTAMINANTS." Food Additives and Contaminants **10**(3): 275-305.
Renwick, A. G. and N. R. Lazarus (1998). "Human variability and noncancer risk assessment - An analysis of the default uncertainty factor." Regulatory Toxicology and Pharmacology **27**(1): 3-20.
Rigaux, C., J.-B. Denis, et al. (2013). "A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: Estimation of< i> D</i> at 121.1° C and pH 7,< i> zT</i> and< i> zpH</i> of< i> Geobacillus stearothermophilus</i>." International journal of food microbiology **161**(2): 112-120.
Salazar, D. E., R. C. Dockens, et al. (1995). "Pharmacokinetic and Pharmacodynamic Evaluation of Warfarin and Nefazodone Coadministration in Healthy Subjects." The Journal of Clinical Pharmacology **35**(7): 730-738.
Samer, C. F., Y. Daali, et al. (2010). "The effects of CYP2D6 and CYP3A activities on the pharmacokinetics of immediate release oxycodone." Br J Pharmacol **160**(4): 907-918.
Samer, C. F., Y. Daali, et al. (2010). "Genetic polymorphisms and drug interactions modulating CYP2D6 and CYP3A activities have a major effect on oxycodone analgesic efficacy and safety." Br J Pharmacol **160**(4): 919-930.
Sandberg, A., B. Abrahamsson, et al. (1993). "Steady-State Bioavailability and Day-to-Day Variability of a Multiple-Unit (CR/ZOK) and a Single-Unit (OROS) Delivery System of Metoprolol After Once-Daily Dosing." Pharmaceutical Research **10**(1): 28-34.
Seeringer, A., J. Brockmoller, et al. (2008). "Enantiospecific pharmacokinetics of metoprolol in CYP2D6 ultra-rapid metabolizers and correlation with exercise-induced heart rate." Eur J Clin Pharmacol **64**(9): 883-888.

Sehrt, D., I. Meineke, et al. (2011). "Carvedilol pharmacokinetics and pharmacodynamics in relation to CYP2D6 and ADRB pharmacogenetics." Pharmacogenomics **12**(6): 783-795.
Sharma, A., P. Pibarot, et al. (2005). "Modulation of metoprolol pharmacokinetics and hemodynamics by diphenhydramine coadministration during exercise testing in healthy premenopausal women." J Pharmacol Exp Ther **313**(3): 1172-1181.
Sharma, A., P. Pibarot, et al. (2010). "Toward optimal treatment in women: the effect of sex on metoprolol-diphenhydramine interaction." J Clin Pharmacol **50**(2): 214-225.
Siddoway, L. A., K. A. Thompson, et al. (1987). "Polymorphism of propafenone metabolism and disposition in man: clinical and pharmacokinetic consequences." Circulation **75**(4): 785-791.
Simon, H. U. and B. Malhotra (2009). "The pharmacokinetic profile of fesoterodine: similarities and differences to tolterodine." Swiss Med Wkly **139**(9-10): 146-151.
Skinner, M. H., H. Y. Kuan, et al. (2003). "Duloxetine is both an inhibitor and a substrate of cytochrome P4502D6 in healthy volunteers." Clin Pharmacol Ther **73**(3): 170-177.
Slanar, O., M. Nobilis, et al. (2007). "Pharmacokinetics of tramadol is affected by MDR1 polymorphism C3435T." Eur J Clin Pharmacol **63**(4): 419-421.
Somer, M., J. Kallio, et al. (2000). "Influence of hydroxychloroquine on the bioavailability of oral metoprolol." Br J Clin Pharmacol **49**(6): 549-554.
Spina, E., A. M. Pollicino, et al. (1993). "Effect of fluvoxamine on the pharmacokinetics of imipramine and desipramine in healthy subjects." Ther Drug Monit **15**(3): 243-246.
Sutton, A. J. and J. Higgins (2008). "Recent developments in meta-analysis." Statistics in medicine **27**(5): 625-650.
Truhaut, R. (1991). "The concept of the acceptable daily intake: An historical review∗." Food Additives & Contaminants **8**(2): 151-162.
Turgeon, J., A. C. G. Uprichard, et al. (1991). "Resolution and Electrophysiological Effects of Mexiletine Enantiomers." Journal of Pharmacy and Pharmacology **43**(9): 630-635.
Ujhelyi, M. R., E. A. O'Rangers, et al. (1993). "The pharmacokinetic and pharmacodynamic interaction between propafenone and lidocaine." Clin Pharmacol Ther **53**(1): 38-48.
Walton, K., J. L. Dorne, et al. (2001). "Uncertainty factors for chemical risk assessment: interspecies differences in glucuronidation." Food and Chemical Toxicology **39**(12): 1175-1190.
Walton, K., J. L. Dorne, et al. (2001). "Uncertainty factors for chemical risk assessment: interspecies differences in the in vivo pharmacokinetics and metabolism of human CYP1A2 substrates." Food and Chemical Toxicology **39**(7): 667-680.
Werner, U., C. Lamprecht, et al. (2006). "Valdecoxib does not interfere with the CYP2D6 substrate metoprolol." Int J Clin Pharmacol Ther **44**(9): 397-400.
Werner, U., D. Werner, et al. (2003). "Celecoxib inhibits metabolism of cytochrome P450 2D6 substrate metoprolol in humans." Clin Pharmacol Ther **74**(2): 130-137.
Wu, X., L. Yuan, et al. (2013). "The impact of CYP2D6 polymorphisms on the pharmacokinetics of codeine and its metabolites in Mongolian Chinese subjects." Eur J Clin Pharmacol.
Xu, Z. H., H. G. Xie, et al. (1996). "In vivo inhibition of CYP2C19 but not CYP2D6 by fluvoxamine." Br J Clin Pharmacol **42**(4): 518-521.
Yasui-Furukori, N., M. Saito, et al. (2007). "Terbinafine increases the plasma concentration of paroxetine after a single oral administration of paroxetine in healthy subjects." Eur J Clin Pharmacol **63**(1): 51-56.
Yin, O. Q., V. W. Mak, et al. (2012). "Impact of CYP2D6 polymorphisms on the pharmacokinetics of lovastatin in Chinese subjects." Eur J Clin Pharmacol **68**(6): 943-949.
Yoo, H. D., H. Y. Cho, et al. (2012). "Population pharmacokinetic analysis of risperidone and 9-hydroxyrisperidone with genetic polymorphisms of CYP2D6 and ABCB1." J Pharmacokinet Pharmacodyn **39**(4): 329-341.
Yoo, H. D., S. N. Lee, et al. (2011). "Influence of ABCB1 genetic polymorphisms on the pharmacokinetics of risperidone in healthy subjects with CYP2D6*10/*10." Br J Pharmacol **164**(2b): 433-443.

Zhou, H.-H., L. B. Anthony, et al. (1990). "Quinidine reduces clearance of (+)-propranolol more than (-)-propranolol through marked reduction in 4-hydroxylation." Clin Pharmacol Ther **47**(6): 686-693.

Zhou, S. F. (2009). "Polymorphism of human cytochrome P450 2D6 and its clinical significance: part II." Clin Pharmacokinet **48**(12): 761-804.

Zhu, X.-W., A. Sedykh, et al. (2013). "The use of pseudo-equilibrium constant affords improved QSAR models of human plasma protein binding." Pharmaceutical research **30**(7): 1790-1798.

4 Discussion

4.1 Main results and limits

A first work proposed a new dynamic model for exposure and a statistical method to calibrate it using biomonitoring data and toxicokinetic modeling. This model was applied to exposure to dioxins, furans and dl-PCBs and calibrated with blood serum levels recorded in a French population of fishermen families using a single compartment first order TK model. It results in a dynamical model that is able to predict body burden from dietary intake taking the historical evolution of dietary exposure into account. This calibrated model was then applied, in a second work, to the general French population in order to predict the body burden from dietary exposure and assess the risk related to the body burden. In this work, the case of cumulated exposure to chemicals with different half-lives was investigated with dioxins, furans and dl-PCBs by considering specific kinetics and historic evolution of exposure for each congener (or group of congeners). In a third work, the same approach was used in a context of aggregated exposure and applied to dietary and cigarette smoke exposure to cadmium of the French population. A dynamic model of exposure is used and the same statistical method is used to calibrate the model from biomonitoring data. A PBTK model is used to describe the kinetic from different pathways of exposure. These works demonstrate that it is possible to deal with the relation of current exposure assessment and internal doses measurements when considering persistent chemicals. Indeed, taking the kinetics into account is even more critical when the half-life is long. Consequently such modeling that considers life-long exposure is even more suitable for persistent chemicals. The two case studies proposed in this manuscript were thus chosen for the long half-life of the chemicals and because both individual biomonitoring data and data on individual habits (i.e. food consumption and smoking habits for cadmium) were available. Such work does not aim to convince that all this data are needed to evaluate the exposure but rather to investigate what can be learned when combining all the data. Figure 3 shows schematically, from the example of our work, how biomonitoring data can be used to inform an exposure model which can be reused to simulate future exposure or exposure of other populations. This work also brings other thoughts especially on the integration of the kinetics and inter and intra-individual variability.

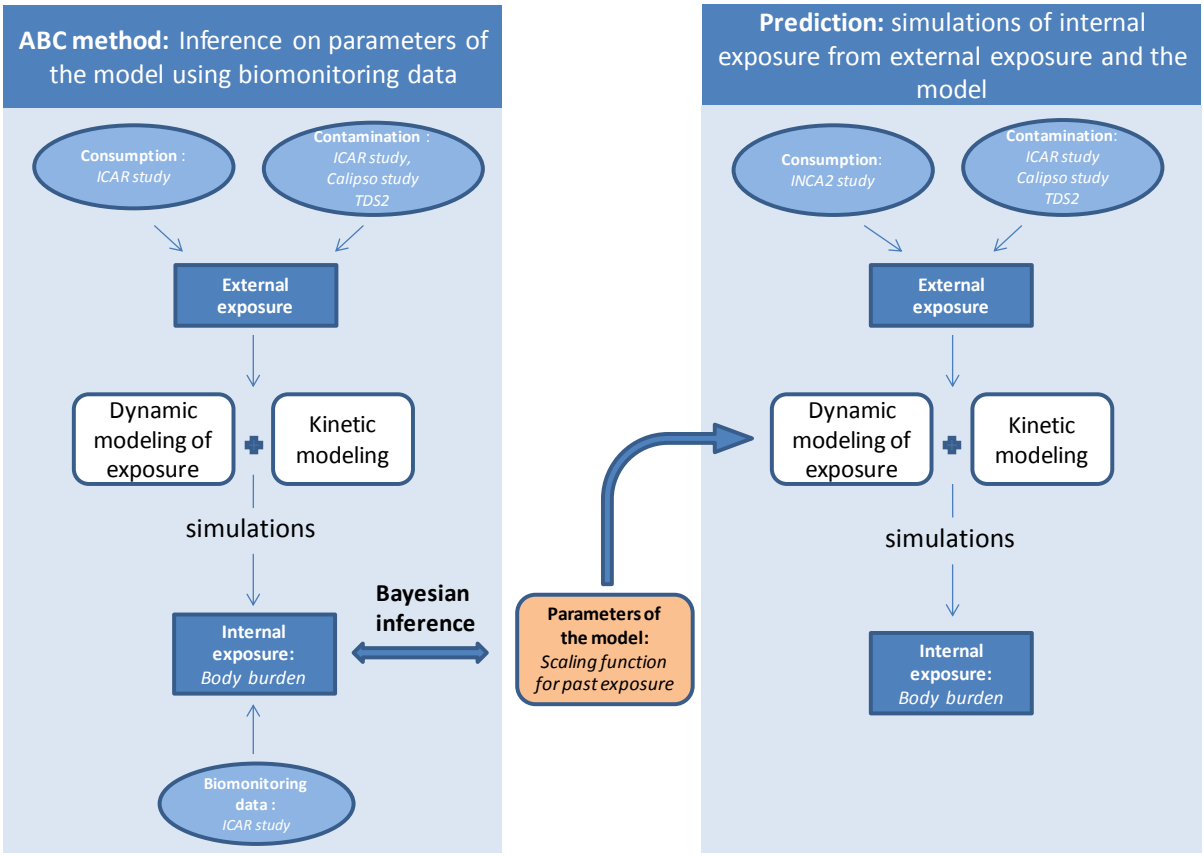


Figure 3: Schematic representation of the combination of several data to improve exposure assessment: example of the exposure to dioxins, furans and dl-PCBs in France combining data from ICAR (Anses/InVS, 2011), Calipso (Leblanc, 2006), TDS2 (Sirot et al.,2012), INCA2 (Dubuisson et al., 2010).

4.1.1 Toxicokinetic modeling

When trying to link external exposure to biomonitoring data, the kinetic is essential. Indeed, every steps of the ADME (Absorption, Distribution, Metabolism and Excretion) are involved in the process that leads to an internal dose measured in a specific tissue. The cadmium case study illustrates for example the importance of the different absorption rate when dealing with exposure from different pathways. In this case, the higher pulmonary absorption compared to the gastrointestinal absorption explains why the contribution from cigarette smoke increases when considering internal dose compared to external exposure. The case study on cumulative exposure to dioxins, furans and dl-PCBs illustrates the importance of the differences in half-lives when dealing with combined exposure. Indeed, the difference in half lives of the different congeners explains why the contribution of some dl-PCBs is much higher when considering internal dose instead of external exposure. Cadmium case study also highlights the importance of distribution and excretion. Indeed, cadmium is eliminated from the body primarily in urine and because cadmium remains tightly bound to metallothionein which is almost completely reabsorbed in the renal tubules, the rate of excretion

is low. Therefore, cadmium accumulation in the kidneys is significant and explains why cadmium concentration in blood reflects recent exposure whereas urinary cadmium concentration more closely reflects total body burden.

This demonstrates that accurate information on the kinetic of the chemical and toxicokinetic modeling are needed. More and more methods allow to investigate ADME processes and to provide data to build TK models. Especially in the context of reduction of *in vivo* animal testing, a number of in vitro models have been developed, mostly in the pharmaceutical industry. Additionally, a number of in silico models have been developed to predict ADME parameters with chemical structural features such as Quantitative Structure-Activity Relationships (QSAR). These in vitro models have still received little attention in hazard assessment of chemicals for the food safety area (Coecke et al., 2013) but Efsa (2014) reports most of these new methods to investigate TK processes. For example, we can cite for absorption, Caco-2 cells which are the most widely in vitro cell models to estimate intestinal absorption (Alqahtani et al., 2013) but also Human P-gp (hP-gp) membranes prepared from baculovirus-infected insect cells and control membranes (Meyer et al. 2013). A number of models and related software are also available for the simulation and modeling of intestinal absorption and metabolism such as CAT (compartmental absorption transit) models, ACAT (Advanced compartmental absorption transit model model) in GastroPlus, ADAM (Advanced Dissolution model) in SimCYP or GastrointestinalSim in GI-Sim. To investigate metabolism, we can cite Hepatic microsomes which are the most frequently used subcellular fractions for drug metabolism studies but also Hepatocytes, Cell line expressing CYP cDNA ,Immortalised cell lines and Liver slices. Regarding excretion, it seems that *in vitro* estimation of excretion has not been investigated much and no *in vitro* method is currently available (Coecke et al., 2013) but Efsa (2014) highlights that some progress has been made with pharmaceuticals (Kusuhara and Sugiyama, 2009).

4.1.2 Inter-individual variability

Inter-individual variability is a crucial aspect whatever the risk assessment that is conducted. Conventionally in risk assessment, a default uncertainty factor is used to cover the human variability (Truhaut 1991). However, using modeling as we did makes it possible to input some mechanistic and data based information about inter-individual variability. Indeed, the comparison of the internal dose estimated from assessed external exposure and TK modeling with internal dose provided by biomonitoring data allow to answer the questions: is the variability provided by data on individual behaviors and contamination of the media added to the variability provided by mechanistic

knowledge included in the TK model sufficient to explain the observed variability in internal dose? Are we missing some variability or inversely are we overestimating the variability? In both case studies proposed in this work i.e. dioxins (furans and dl-PCBs) and cadmium, inter-individual variability regarding food consumption (and smoking habits for cadmium) were taken into account since individual consumptions provided by ICAR, INCA2 or ENNS studies were used. These individual consumptions were combined with average national data on food contamination provided by the Second French Total Diet Study. Consequently no inter-individual variability was considered regarding the food (and cigarette) contamination. This is the classical assumption when considering chronic exposure. However, due to brand loyalty or particular local consumption, some variability can exist that could be interesting to study. Moreover, when dealing with other source of exposure such as indoor air or dust, variability of contamination level inter-house can be very high and has to be included in the modeling. Regarding inter-individual variability in kinetic, for the dioxins case study no variability in absorption was considered and the variability in apparent half-life was modeled as a function of age and BMI (Milbrath et al. 2009). For the cadmium case study, inter-individual variability was modeled using distribution for all the parameters of the PBTK model. A specific variability in absorption between male and female was also modeled since it was highlighted as very significant in the data and explained in literature. In these two applications, the comparison between predictions and biomonitoring data shows a good agreement in the predicted and observed variability. However, this variability is also substance-dependent and other sources of variability may be investigated when considering other chemicals.

Bois et al. (2010) reviewed the sources of inter-individual variability in ADME parameters and how it can be predicted. They distinguished three main contributing sources to the total variability: the variation across a population of 'normal' individuals at the same age, e.g. young adults; the variation across the population resulting from their different ages, e.g. infants or the elderly; and the variation resulting from the existence of subpopulations that differ in some way from the 'normal' population, e.g. due to genetic polymorphisms. Since the two first sources of variability are already well documented whereas the inter-individual variability related to genetic polymorphisms remains little known, we decided to investigate it through the example of CYP2D6 polymorphism. From a meta-analyze of the pharmacokinetic (PK) data available for the CYP2D6 substrates, we highlighted a high variability related to the polymorphism within the same ethnic group but also inter-ethnic. This variability estimated in our case study was shown higher than the one covered by the default uncertainty factor for kinetic proposed by Renwick (1993). Such variability in kinetics started to be well known for pharmaceuticals but not yet for chemical risk assessment. Consequently, inter-individual variability provided by such meta-analysis of pharmaceutical studies could be used as input

for toxicokinetic models in order to improve the integration of inter-individual variability in risk assessment. There are already some cases where huge variability in kinetics may explain a higher susceptibility of some individuals. For example, recently, Singh et al. (2014) suggested that polymorphism in CYP2D6 and Glutathione S-Transferases Pi 1 Genes may explain a stronger association between Alzheimer disease organochlorine pesticides in some individuals. Deng et al. (2004) and Elbaz et al. (2004) both analyzed the combined effect of pesticide exposure and a polymorphism of CYP2D6 (which metabolizes organophosphates, the herbicide atrazine, and 1-methyl-4-phenyl-1, 2, 3, 6,-tetrahydropyridine). As a result, the effect of professional pesticide exposure on Parkinson disease was increased in poor CYP2D6 metabolizers, whose slower CYP2D6 metabolism rendered them more susceptible to the biological effects of pesticides (Freire and Koifman, 2012). The objective of our work was also to investigate the possible prediction of this variability by combining in vitro data on the substance (i.e. the affinity for the enzyme) and distribution provided by this work. This would make it possible to use these distributions as inputs for TK models in order to integrate this source of inter-individual variability.

In order to take inter-individual variability into account in the modeling, the Bayesian framework has been found effective. Indeed, using a hierarchical modeling made it possible to consider together individual data (i.e. age, food consumption, smoking habits, biomonitoring data) and data at a population level (i.e. Contamination data) in order to infer on parameters at a population level. It also provided a natural framework to use distributions for the parameters of the PBTK model and therefore consider inter-individual variability in the kinetic modeling. Many previous works highlighted the advantage of Bayesian approach for this purpose (Bois et al., 2010; Krauss et al., 2013) and some software thus propose to implemented PBTK models in a Bayesian context like MCSIM (Bois, 2009). Increasingly, some anatomical and physiological parameter distributions that can be used as input for MCMC simulations can be obtained from freely available web-based application (McNally et al., 2012). For example, we can cite PopGen which is a virtual human population generator (http://xnet.hsl.gov.uk/popgen/) used by McNally et al. (2012) to reconstruct the exposure to m-Xylene. ABC method was also proven efficient to deal with the dynamical modeling of the exposure we proposed by avoiding the exact evaluation of the likelihood. Indeed, by using appropriate summary statistics and distances for our specific context and advance algorithm to reduce computation time, ABC succeeds to infer on parameters of our models. ABC already became popular in some fields like population genetics (Csillery et al., 2010) and mathematical epidemiology (Blum et al., 2010) but, to the best of our knowledge, this is the first time this method is applied to chemical risk assessment. Due to its intuitive nature and its ability to simply deal with complex modeling, the use of such method in this context could increase.

4.1.3 Modeling of lifelong exposure

When considering persistent chemicals with long half-life in human body, we demonstrated the relevance of life-long exposure. Indeed, both serum concentration for dioxins and urinary concentration for cadmium reflect a long-term exposure. In both cases, the current external exposure was not sufficient to explain the measured internal exposure which highlights the need to consider also past exposure. When considering life-long exposure, the intra-individual variability is a crucial aspect. It was demonstrated in literature that contamination of the environment and thus exposure to persistent chemicals can significantly change over decades due to changes in production, use and successive regulations. This evolution needed to be integrated in a life-long exposure modeling and exposure may thus change according to the age of the individual in relation to the changes in environment. Regarding external exposure, changes in behavior (e.g. food consumption, smoking habits) can also be source of intra-individual variability. In both case studies, changes in exposure related to historic changes in environment were modeled. In dioxins case study, age-dependent changes in food consumption were modeled by combining consumption data on children and adults in individual trajectories of exposure. In the cadmium case study, only the age-dependent change in body weight was considered for food exposure and for cigarette smoke exposure, the ages of starting and stopping smoking were used for each individual. Regarding kinetic, there are many potential sources of intra-individual variability; especially a high age-dependent variability was highlighted for many parameters in Bois et al. (2010). In the dioxins case study, the age dependency of the apparent half-life was considered from the work of Milbrath et al. (2009). Regarding cadmium, the modification of the body weight and the renal function with the age were included in the PBPK model. However, it may be useful to investigate and model other intra-individual variability in order to improve the modeling. For example, the difference between male and female cadmium absorption is explained in Järup and Akesson (2009) by the close relation between cadmium absorption and the expression of the divalent metal transporter 1 (DMT-1), which transports cadmium and iron into the mucosa cell in a competitive manner. Therefore, this difference may be the result of lower iron body stores in females compared to males (Choudhury et al. 2001). The individual iron status was recorded in the study used for the cadmium case study (i.e. ENNS). However, we found no significant correlations between urinary cadmium excretion and ferritin concentration. As for NHANES III study (Choudhury et al. 2001), the lack of correlation between urinary cadmium and iron status may reflect the very different kinetics of iron status and urinary cadmium. Indeed, urinary cadmium excretion reflects the long-term accumulation of cadmium in the body from long-term cadmium absorption, whereas serum ferritin concentration could change in response to relatively recent short-term changes in whole-body iron levels. A strong correlation

between urinary cadmium and iron status would thus be expected only if the measured iron status reflected the long-term iron status of the individuals. Unfortunately, there was no way to identify this from the data. Moreover, this situation of low iron status seems to be exacerbated during pregnancy when enterocytes have an increased DMT-1 density at the apical surface to optimize micronutrients absorption (Åkesson et al., 2002; Vahter et al., 2007). This is also an identified source of variability that would be interesting to add to the model to precisely investigate the risk for pregnant women.

Incidentally, PBTK modeling is a relevant tool to investigate internal exposure during specific stage of life. For example, PBTK models can be developed for the gestation and lactation life stages in humans to understand how the physiological changes associated with development affect kinetics of these compounds in the mother, fetus, and infant (Loccisano et al. 2012; 2013). One limit of our work is the huge uncertainty regarding the exposure during childhood that does not allow to assess the risk for children. A sensitive analyze was conducted and it resulted that exposure under three years doesn't influence the output of our modeling after 18 years old. Thus, since the biomonitoring data used in our case studies was only on adult population, we didn't investigate the exposure of young children. However, early stages of life, i.e. the fetal and early postnatal periods have been highlighted as periods of increased susceptibility (Sly and Flack, 2008; Diamanti Kandarakis et al., 2009). Therefore, investigating internal exposure during this specific period may be an important development of our work. We identified three missing points in our modeling that make the children exposure assessment very uncertain: modeling exchange between mother and fetus and the related burden at birth, modeling external exposure specific of children under 3 years (breast milk and infant food), use specific kinetic for children. Regarding dietary external exposure, infant foods available on the market are very specific and differ from those of population groups considered in the main studies (e.g. Total Diet Study, TDS) and thus specific data are needed. In France, a specific TDS on infants analyzed more than 500 chemical substances in food products consumed by children under 3 years old (Hulin et al. 2014). Combined with a national consumption survey conducted specifically on this population (Fantino, 2005), dietary exposure of children under 3 years old can be assessed specifically. Then PBTK models specific for children could be used to implement the same kind of work we proposed for adults.

4.2 Is it possible to extrapolate the proposed methods to mixtures?

4.2.1 Combined exposure

The work proposed in this manuscript was made in the context of single exposure or cumulative exposure. However, individuals are exposed to a very complex and variable combination of chemicals from different sources and risk assessment of combined exposure to multiple chemicals i.e. chemical mixture forms a new challenge for risk assessment. We thus propose to investigate how our work could be extended to a context of combined exposure.

First, regarding the exposure, the number of chemicals to which an individual can be exposed is very large. The number of associated combination of several chemicals that may form a mixture is thus huge and depends on multiple factors including dietary, behavioral and environmental factors. Therefore, the identification of the mixtures to which individuals are mainly exposed constitutes a first challenge. One way to identify such mixtures would be to study combined exposure in the population. Crépet et al. (2013) for example proposed to model co-exposure to extract cluster of individuals with similar profile of exposure to many pesticides and then study correlation between exposures in high exposed clusters to highlight mixture to which the French population is exposed. However, since the overall exposure of a particular individual depends on multiple factors including dietary, behavioral and environmental factors, starting from exposure to identify the mixture makes it difficult to characterize the determinants of exposure. In the following work, we proposed another way to study the combined exposure starting from the modeling of the source of exposure. The method is applied to identify the main mixtures of pesticides to which the French population is exposed and to connect these combined exposures with the diet. The non-negative-matrix factorization (NMF) is used to extract the main consumption patterns from individual consumption data of the French population and the exposure to associated mixtures.

Food and Chemical Toxicology 59 (2013) 191–198

Contents lists available at SciVerse ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Identification of pesticide mixtures and connection between combined exposure and diet

Camille Béchaux [a,*], Mélanie Zetlaoui [b], Jessica Tressou [c], Jean-Charles Leblanc [a], Fanny Héraud [d], Amélie Crépet [a]

[a] ANSES, Risk Assessment Department, 27-31 Avenue du Général Leclerc, 94701 Maisons-Alfort Cedex, France
[b] Université Paris Ouest, MODAL'X, 200 Avenue de la République, 92001 Nanterre Cedex, France
[c] INRA, Unité Met@risk, 16 rue Claude Bernard, 75231 Paris Cedex 05, France
[d] European Food Safety Authority (EFSA), Dietary and Chemical Monitoring Unit, Via Carlo Magno 1A, I-43100 Parma, Italy

ARTICLE INFO

Article history:
Received 19 April 2013
Accepted 5 June 2013
Available online 14 June 2013

Keywords:
Pesticides mixtures
Dietary patterns
Non-Negative Matrix Factorization
Combined exposure
Food risk assessment

ABSTRACT

The identification of the major associations of pesticides to which the population is exposed is the first step for the risk assessment of mixtures. Moreover, the interpretation of the mixtures through the individuals' diet and the characterization of potentially high-risk populations constitute a useful tool for risk management. This paper proposes a method based on Non-Negative Matrix Factorization which allows the identification of the major mixtures to which the French population is exposed and the connection between this exposure and the diet.

Exposure data of the French population are provided by the Second French Total Diet Study. The NMF is implemented on consumption data to extract consumption systems which are combined with the residue levels to link dietary behavior with exposure to mixtures of pesticides. A clustering of the individuals is achieved in order to highlight clusters of individuals with similar exposure to pesticides/consumption habits.

The model provides 6 main consumption systems, 6 associated mixtures of pesticides and the description of the population which is most exposed to each mixture. Two different ways to estimate the matrix providing the mixtures of pesticides to which the population is exposed are suggested. Their advantages in different contexts of risk assessment are discussed.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Assessing the risk for human health related to the exposure to chemical mixtures is one of the main challenge of the risk assessment bodies (EFSA, 2008; WHO, 2009). The risk assessments conducted up-to-date focused on chemicals belonging to the same chemical family and/or sharing the same mechanisms of action (Boobis et al., 2008; Bosgra et al., 2009; Muller et al., 2009; Reffstrup et al., 2010). These cumulative risk assessments have limitations since they do not consider possible interaction between substances whereas these interactions between pesticides cannot be ruled out (EFSA, 2012). Moreover, they do not consider the probability of the population to be really exposed to these mixtures through actual food consumption. Boobis et al. (2011) concluded that synergistic effects at low doses were observed in very few studies available in literature but more work is required to determined how frequently synergy is likely to occur at real word exposures. Therefore, the identification of mixture to which the population is really exposed is an important first step in the risk assessment related to mixtures. This first step will help to prioritize the combinations of pesticides for which it would be crucial to investigate the possible combined effects, as toxicological information about mixtures is essential to carry out risk assessment.

Determining the chemical mixtures to which the population is really exposed is not an easy task considering the number of chemicals potentially present in the diet and the subsequent number of possible mixtures. Indeed, the Second French Total Diet Study recently revealed the presence of at least 283 different residues of pesticides in the diet of the French population (Nougadère et al., 2012). Crépet and Tressou (2010) recently proposed a method to determine the mixtures of pesticides to which a population is exposed from the population combined exposure. However, that approach did not allow to easily characterize the determinants of exposure to the mixtures, such as the main foods contributing to the combined exposure.

Abbreviations: BIC, Bayesian information criterion; BMI, Body Mass Index; CS, consumption system; LB, lower-bound; LOD, limit of detection; NMF, Non-Negative Matrix Factorization; TDS, Total Diet Study; UB, upper-bound.

* Corresponding author.
E-mail address: camille.bechaux@anses.fr (C. Béchaux).

0278-6915/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fct.2013.06.006

The objective of this paper is to develop a method which allows the identification of the major mixtures to which the French population is exposed and simultaneously the description of its dietary determinants. This new approach assumes that the association between pesticides in the exposure of an individual is related to the way the individual combines the foods in his diet. Therefore, the consumption of the individuals will be first modeled in order to determine the major consumption patterns in the population. Zetlaoui et al. (2011) assume that even though a very large number of different foods are involved in individual consumption patterns, all possible food combinations are not observed in practice. Then, the consumption of the population is described by few consumption patterns that are a linear combination of specific foods. These combinations of foods are interpreted as statistical latent variables that are called consumption systems. According to this modeling, an individual diet must be seen as a linear combination of several consumption systems. The identification of the consumption systems combined with residue levels in foods will allow for the identification of the mixtures of pesticide.

Such latent variables can be provided by traditional methods as Principal Component Analysis or Factor Analysis. Incidentally they are widely used to determine dietary patterns (Fung et al., 2001; Hu et al., 2000; Oslera et al., 2001). However, because latent variable and noise are modeled by Gaussian random vectors, these methods are not really suitable for non-negative data with excess zeros. A latent variable based method called Non-Negative Matrix Factorization (NMF) has been proposed by Lee and Seung (Lee and Seung, 1999; Saul and Lee, 2002) to specifically deal with non-negative data using non-negativity constraints. This method is also particularly adapted to non-negative data that have excess zeros and measurement error such as exposure and food consumption data.

In this work, NMF is applied on the Second French Total Diet Study data to extract mixture of pesticides to which the French population is exposed related to the major consumption systems found in the French diet. A clustering of the individuals is achieved in the resulting latent consumption and exposure space in order to highlight clusters of individuals with similar consumption habits and exposure to pesticides. These clusters are then characterized from individual variables such as age, sex, Body Mass Index (BMI) and household income.

2. Materials and method

2.1. Exposure data from the Second French Total Diet Study

Pesticides exposure data of the French population are provided by the Second French Total Diet Study (Nougadère et al., 2012). "Total Diet Studies" (TDS) follow a standardized international methodology and aim at assessing the dietary exposure of individuals for many contaminants taking into account residue levels in foods as consumed at home. In a TDS, exposure is assessed by combining individual food consumption data and residue levels data from food sample analysis.

The individual food consumption data used in the Second French Total Diet Study (Sirot et al., 2009) were provided by the second "Individual and National Study on Food Consumption", INCA2 survey, carried out by the French Food Safety Agency (Dubuisson et al., 2010; Lioret et al., 2010) between late 2005 and April 2007. Two independent population groups were included in the survey: 2624 adults aged 18–79 years and 1455 children aged 3–17 years. Participants were selected using a three-stage random probability design stratified by region of residence, size of urban area and population category (adults or children). This design was summarized into sampling weights provided for each surveyed individual representing its frequency in the entire French population. Each participant was asked to complete a 7-day food diary as well as other questionnaires on anthropometric and socio-economic factors. Foods declared were subsequently categorized into 1305 "as consumed" food items. To be representative of the French adult population, a sample of adults is built out from the original samples by carrying out random trials with replacement respecting the individual sampling weights. The mean of the quantities of the same food consumed by each individual during the week is used in order to assess chronic exposure.

In the Second French Total Diet study, 283 pesticides were measured in foods as consumed at home. Among these, some were considered as priorities following the work done by the French Food Safety Agency using a conservative ranking method recommended by the WHO (Menard et al., 2008). In this way, a list of 72 priority pesticides was drawn up. Because of analytical issues (feasibility, cost and grouping of substances) 62 priority pesticides were analyzed. These pesticides were sought in 1235 composite samples corresponding to 194 food items that cover 90% of the adult and child diet from the INCA2 survey. To make up the composite samples, about 19,000 food products were bought during different seasons from 2007 to 2009 in 36 French cities and prepared according to the food preparation practices recorded in the INCA2 survey. The method proposed to extract mixture in this work requires a single value for residue levels data for all individuals. Therefore, national residue levels data are used which are the mean of the regional data.

To handle left-censored data, i.e., results which are below the detection limit, exposure is estimated under two scenarios: lower-bound scenario (LB) where undetected results are set to 0 and upper-bound scenario (UB) where undetected results are set to the detection limit. With the LB scenario, for 36 pesticides among the 62, the exposure of the whole population is equal to zero. Therefore, only the remaining 26 pesticides are considered.

2.2. Methodology

2.2.1. Combined exposure estimation

The combined exposure of an individual i to P pesticides is represented by a vector of length P, $e_i = (e_i^{(1)}, \ldots, e_i^{(P)})$. The pth element $e_i^{(p)}$, which represents the exposure to the pesticide $p \in (1,\ldots,P)$ of individual i, is calculated by combining individual consumption quantities with the concentration residue levels

$$e_i^{(p)} = \frac{\sum_{f=1}^{\mu} q_{if} \times c_{fp}}{bw_i}, \tag{1}$$

where q_{fi} is the quantity of food f consumed by individual i, C_{pf} is the residue level for pesticide p in food f and bw_i is the body weight of individual i.

Regarding a population of N individuals, the combined exposure matrix is the product of the $(P \times F$ matrix C of residue levels C_{fp} and the $(F \times N)$ matrix Q of consumption quantities q_{fi}. Thus, the $(P \times N)$ combined exposure matrix E is the juxtaposition of the N column vectors e_i

$$E = CQ. \tag{2}$$

2.2.2. Identification of mixtures

The food consumption Q is modeled by non-negative matrices to provide K consumption systems, as proposed in Zetlaoui et al. (2011). The consumption systems are combinations of food describing the way individuals associate foods in their diet. Since consumed quantities are not comparable between foods (e.g., drinking 1 kg of water is not comparable to eating 1 kg of meat), the consumption is normalized to avoid scaling effect. Thus, the normalized consumption matrix is written as

$$S_Q^{-1}Q = WH + \varepsilon, \tag{3}$$

where S_Q is the $(F \times F$ diagonal matrix of the standard deviations of the consumption of each food across individuals, W is the $(F \times K$ matrix of consumption systems with the identifiability constraint $\sum_{f=1}^{F} W_{fk} = 1$ for $k = 1,\ldots,K$, H is the $(K \times N$ matrix which provides how individual combine the consumption systems to constitute their diet and ε is the $(F \times N)$ matrix of residuals, the columns of which $\varepsilon_i = (\varepsilon_i^{(1)}, \ldots, \varepsilon_i^{(F)})$ are Gaussian random vectors with mean 0 and covariance Γ_i.

From Eqs. (2) and (3), the following model for combined exposure is proposed:

$$\begin{cases} E = CQ \\ S_Q^{-1}Q = WH + \varepsilon. \end{cases} \tag{4}$$

To avoid scaling effect in the pesticide exposure, E is multiplied by the inverse of the matrix S_E, which is the $(P \times P)$ diagonal matrix of the standard deviations of the exposure to each pesticide across individuals. Thereby, as illustrated in Fig. 1, the normalized exposure is provided by multiplying $S_Q^{-1}Q$ by $S_E^{-1}CS_Q$ and

$$S_E^{-1}E = W^*H^* + \varepsilon^*, \tag{5}$$

where $W^*H^* = S_E^{-1}CS_QWH$ and $\varepsilon^* = S_E^{-1}CS_Q\varepsilon$.

To satisfy the identifiability constraint $\sum_{p=1}^{p} W_{pk}^* = 1$, a normalization $(K \times K$ diagonal matrix R with elements corresponding to the sum of each column k of $S_E^{-1}CS_QW$ is introduced. Therefore $W^* = S_E^{-1}CS_QWR^{-1}$ and $H^* = RH$. Each column of W^* represents an exposure system and each $W_{p,k}$ cell gives the proportion of pesticide p in the exposure system k. Each $H_{k,i}^*$ cell gives the weight of the exposure system k in the exposure of individual i. In this model, H^* is equal to H up to the renormalization, thereby, the more an individual uses a consumption system k in his diet, the more exposed to the corresponding exposure system k.

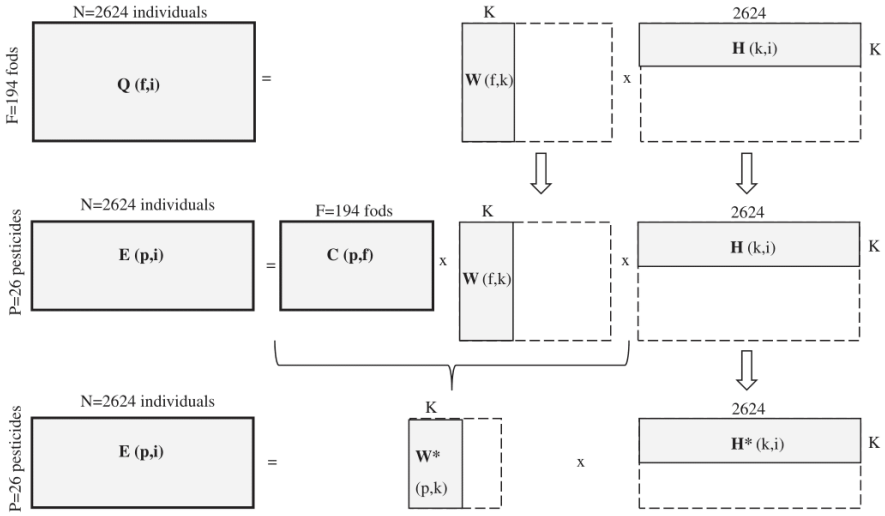


Fig. 1. Schematic representation of the model (normalization and residuals are not shown). Dashed lines indicate that *K*, the number of systems, may be variable.

2.3. NMF implementation

In order to estimate (W^*, H^*, two different estimators can be considered. When consumption data Q and residue levels data C are available, a consumption based estimation $(\widehat{W}^{*(Q)}, \widehat{H}^{*(Q)})$ is given by minimizing the product of C and $D_K(Q,(WH))$, where

$$D_K(Q,(WH)) = \sum_{n=1}^{N}\sum_{f=1}^{F}\left(Q_{f,n} - \sum_{k=1}^{K} W_{f,k}H_{k,n}\right)^2.$$

When only exposure data E are available, an exposure based estimation $(\widehat{W}^{*(E)}, \widehat{H}^{*(E)})$is given by minimizing $D_K(E,(W^*H^*))$, where

$$D_K(E,(W^*H^*)) = \sum_{n=1}^{N}\sum_{p=1}^{P}\left(E_{p,n} - \sum_{k=1}^{K} W^*_{p,k}H^*_{k,n}\right)^2.$$

To minimize the cost function D_K the algorithm proposed in Lee and Seung (2001) is implemented using the R package NMF (Gaujoux and Seoighe, 2010). This multiplicative algorithm is based on a gradient descent approach: it starts with an initial choice for (WH), or (W^*H^*), then iterates until the change in value D_K is negligible.

The NMF model requires fixing on the number K of underlying systems that is typically chosen such that $K \ll F$ in the consumption based estimation and $K \ll P$ in the exposure based estimation. To choose the optimal K value, NMF is run for different values of K and different criteria are calculated for each run. Zetlaoui et al. (2011) suggest the residual sum of squares as a criterion. A Bayesian information criterion (BIC1) proposed by Bai and Ng (2002) is also calculated, see Fig. 2.

2.4. Clustering of individuals

Through NMF, each individual exposure and consumption are reduced to a set of exposure and consumption system coefficients stored in the matrix H^* (note that the matrix H could be used to obtain the same results). In order to cluster individuals with similar exposure and consumption pattern, a hierarchical clustering is implemented on the matrix H^* using the Agglomerative Nesting (AGNES) algorithm (Kaufman and Rousseeuw, 1990). At a first step, clusters composed by one individual are created, and then, such clusters are merged into greater ones according to their distance. The second step is repeated iteratively until only one cluster is obtained, composed by all data objects. Euclidean distances between individuals and ward distances between clusters are chosen. Since for each aggregation level, some inter-class inertia is lost, the number of clusters that precedes a significant loss of inertia is chosen as the optimal number of clusters.

To characterize the clusters, comparisons between age, Body Mass Index (BMI), gender and household income of the individuals of a cluster and the whole population are made. Regarding age and BMI, a comparison of means (t-test) is used. The household monthly income is divided into 3 classes, 1: less than 1300 Euros, 2: between 1300 and 3100 Euros, 3: greater than 3100 Euros. Regarding the proportion men/women and the proportions of the different income classes, a chi-squared test is used.

3. Results

3.1. Lower-bound scenario (LB)

3.1.1. Number of systems K

The residual sum of squares and the BIC are plotted as a function of K for the two estimators of (W^*, H^*, Fig. 2). The plot of the residual sum of squares shows an inflection point at $K = 10$ regarding the exposure based estimation and no inflection point regarding the consumption based estimation. The plot of the BIC shows two significant decreases at $K = 10$ and $K = 15$ regarding the exposure based estimation and no significant decrease regarding the consumption based estimation. Therefore, it is chosen to consider 10 exposure and consumption systems.

3.1.2. Consumption systems and mixtures

The NMF is implemented to the consumption matrix Q composed of 194 foods and 2624 individuals. This results in a (194×10) W matrix and a (10×2624) H matrix. Matrix H provides the share of each consumption system allowing to describe the diet of each individual, and matrix W provides the weight of each food in a given consumption system. For example, beef steak, pasta, baguette, ham and butter have high weights for consumption system no. 6. Therefore they are associated in the diet of individuals who have high weights on the line 6 of the matrix H. Residue levels data are then combined with the W matrix in order to provide the matrix W^* composed of the 10 exposure systems corresponding to the 10 consumption systems (consumption based estimation). The matrix W^* is represented in Fig. 3. This matrix shows the weights of each pesticide in each exposure system. The lighter the color, the higher the proportion of the pesticide in the exposure system. Pesticides with the highest value of W^* for the same exposure system are considered as components of a mixture. For example, individuals who are highly exposed to Malathion are also highly exposed to Phosalone, Azinphos-methyl and carbendazim in the exposure system no. 8. Malathion, Phosalone, Azinphos-methyl and carbendazim are therefore considered as a mixture. From the W^* matrix, pesticides that constitute more than 10% of the exposure system are gathered in a mixture.

The exposure based estimation is also implemented and provides directly 10 exposure systems (results not shown). The

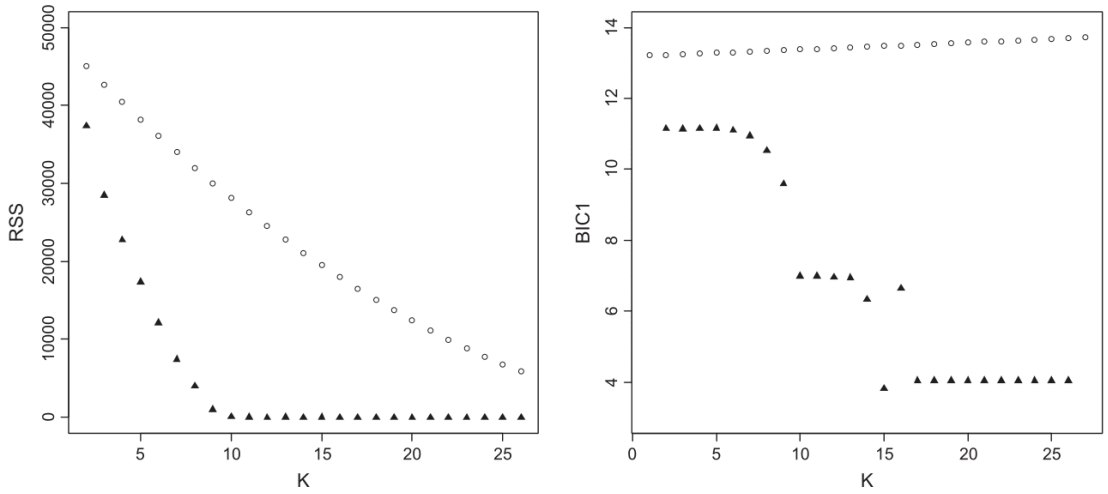


Fig. 2. Plot of the residual sum of square and the BIC as a function of k, points: consumption based estimation of (W^*, H^*), triangles: exposure based estimation of (W^*, H^*).

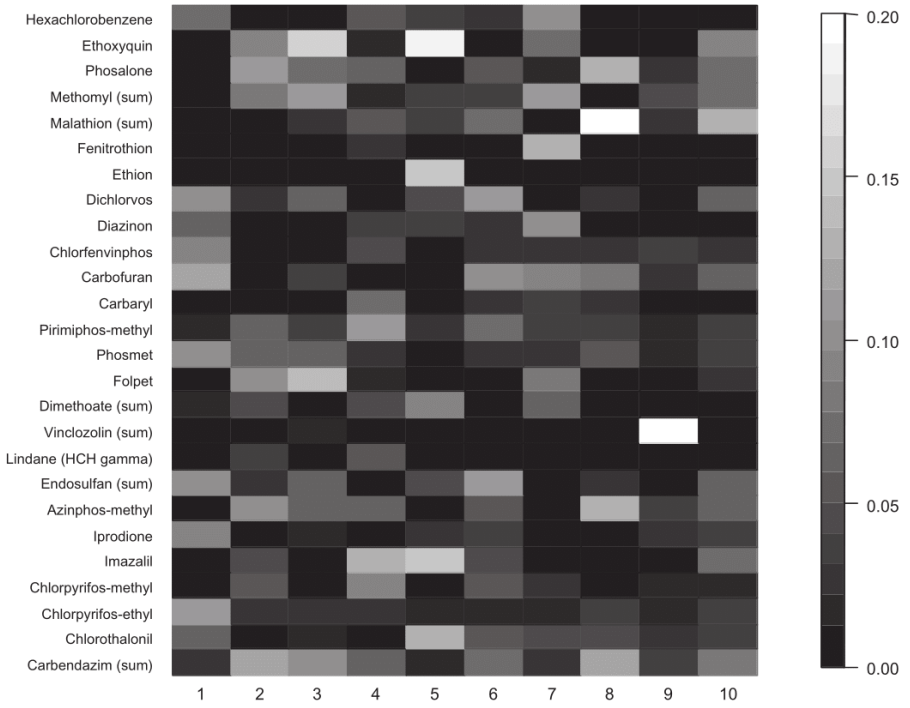


Fig. 3. Graphical representation of the W^* matrix which describes the proportions of each pesticide on the $k = 1, \ldots, 10$ exposure systems (LB scenario).

comparison of the mixtures provided by these two estimations shows that some associations are found in the two estimations: endolsulfan and dichlorvos, Carbendazim and phosalone (associated with folpet in the consumption based estimation and with azinphos-methyl in the other one), ethoxyquin, ethion, imazalil and chlorothalonil, Carbofuran, chlorpyrifos-ethyl and phosmet (associated with dichlorvos and endosulfan in the consumption based estimation). However two mixtures are not found with the exposure based estimation: Imazalil with pirimiphos-methyl and Malathion with ethoxyquin.

A clustering is realized on H^* matrix in order to determine clusters of individuals with similar consumption profiles and similar exposure to the mixtures. Indeed the matrix H^* provides both the weights of the consumption system in the diet of an individual and the weights of the exposure systems in his combined exposure to the 26 pesticides. Thereby, the more an individual uses a consumption system, the more exposed to the mixture associated to that consumption system. Six clusters of individuals with similar profiles of consumption are determined. Result of the clustering, major consumption systems for each cluster, major foods for each consumption system, pesticides forming the associated mixture and description of the individuals are summarized in Table 1. With this model, the description of the clusters through the individual variables and the major consumption systems allow to describe

the main determinants of the exposure to the mixtures. For example, cluster 1 is characterized by consumption system no. 6 which contributes at 42% to the diet of the individuals. The main foods associated in this consumption system are beef steak, pasta, baguette, cooked ham, butter, etc. (firs line of Table 1). Individuals associated to this consumption system are mostly men, younger than the average population with a household income lower than the average population. Through this diet, these individuals are mainly exposed to the mixture composed of Endosulfan, Dichlorvos and pirimiphos-methyl. Consumption system no. 6 is also an important consumption system in the diet of the individuals belonging to the 5 other clusters (Contributing at respectively 14%, 12%, 16%, 12% and 21% of the total consumption); it could be considered as a staple diet for the population. Consequently, the entire population is exposed to the corresponding mixture.

All systems are needed to fully describe all the individual diets but some are more important in some population groups. Thus, consumption systems no. 4 and no. 2 principally composed of fruit and vegetables (of different kinds according to the system) and fish (salmon for system no. 4 and tuna for system no. 2) are more associated with women, older than the average population and with a higher household income. Consumption system no. 5 which associates alcoholic beverages (wine, champagne, and beer) and other foods like coffee, ham, baguette, etc. is more associated with men, older than the average population and who have a higher household income. System no. 10 composed of various foods (some fruits, vegetables, cheese, drink and meat) is associated with women, older, with a higher BMI and who have a lower household income. Lastly, System 1 is very characteristic (51% of the entire diet) of the youngest individuals (men and women).

Pesticides are combined in a mixture related to a consumption system either because they are found in the same food, or because they are found in foods which are combined in a consumption system. For example, regarding mixture no. 6 (the most used in the population), Endosulfan and Dichlorvos are mostly found in beef steak but Endosulfan is also found in cooked ham, pirimiphos-methyl in baguette, butter and water and Carbofuran in pasta.

3.2. Upper-bound scenario (UB)

The exposure systems resulting of the modeling of the exposure with UB scenario does not provide strong association between pesticides. Some exposure systems do but the clustering of the individuals highlights that these systems are the less used in the population. Indeed, in this scenario, if a pesticide has not been searched in a food, it is set at zero and if it has been searched, it is either detected or set at the value of the limit of detection (LOD). Most pesticides were searched in all foods and the values of LOD are quite close for the foods. Therefore the exposure is fairly homogeneous in the population and mixtures can't be extracted. A small number of pesticides have only been searched in very few foods (25 among 194). Therefore the exposure to these pesticides is less homogeneous in the population and these pesticides are strongly associated in some exposure systems. Under this upper bound assumption, it seems that our results are mostly guided by the design of the TDS rather than by the actual residue levels in food, therefore, there is no interest in describing further the results.

4. Discussion

A new approach based on the NMF has been proposed to model the combined exposure in order to extract the main mixtures of pesticides to which the French population is exposed and the dietary determinants of this exposure. The choice of a method particularly adapted to nonnegative data that have excess zeros (i.e., the NMF) in this work is suitable since the matrix of the combined exposure and the matrix of consumption are composed of respectively 75% and 81% of zeros. Regarding the exposure matrix, the zeros result from the assumption made in the residue levels matrix. Indeed, the residue level of a pesticide in a food was regarded as zero either when a pesticide was not sought in a food, or when a pesticide was not detected in a food for the lower-bound scenario. Regarding the consumption matrix, the consumption of a food by an individual was regarded as zero when the individual did not consumed this food during the 7 days of the survey.

Several choices and source of uncertainty may affect the results of the proposed method. First, regarding the exposure assessment, a single residue level for each food was considered in this work which corresponds to the mean of the measured residue levels from the Second French Total Diet Study. However, uncertainty remains about the residue levels data used to assess the exposure. As described in Bertail and Tressou (2006), different concentration values could be taken into account in order to provide a more accurate exposure assessment. However the identification of the mixtures in this case would be more difficult as highlighted in Crépet and Tressou (2010). The choice of the optimal K is also a crucial issue when dealing with NMF. Indeed, choosing a different K would have led to a different number and composition of identified mixtures. In this work, regarding the exposure based estimation, the determination of the optimal K was satisfactory since $K = 10$ corresponds to a significant decrease of the residuals and the BIC. Regarding the consumption based estimation, the choice was not so obvious even though $K = 10$ was a possibility. Therefore, it was chosen to greatly reduce the dimension as suggested by the exposure based estimation. This significant reduction of dimension regarding the consumption habits of the population also allowed for an easier interpretation. Indeed, it provided 6 consumption systems which are characteristics of the corresponding specific individuals but in a sufficient number not to be considered as marginal population. Finally, the weight from which a pesticide is considered as belonging to a mixture directly impacts the composition of the mixtures. In this work, a weight was chosen to ensure a limited number of pesticides in a mixture. However, from Fig. 3, any user can form mixtures from his chosen criteria.

Two different ways to estimate the matrix providing the mixtures were suggested, one based on the consumption matrix and one based on the exposure matrix. The comparison of the mixtures provided by these two estimations shows that some mixtures are shared whereas some others are different. The differences in the results can partly be explained by the complexity of the consumption matrix which considers 194 foods for each individual whereas the exposure matrix considers 26 pesticides for each individual. Therefore, the aggregation of consumption data in few groups of foods could improve the modeling. However, the pesticides found in the foods gathered in a same food group will be necessarily gathered in the same mixture e.g., if sweet pepper and tomato are gathered in a food group "vegetable", pirimiphos-methyl found in sweet pepper but not in tomato and iprodione found in tomato but not in sweet pepper will necessarily form a mixture, which is not fully relevant. Consequently, further work would be interesting to determine the optimal level of aggregation of consumption data in order to improve the modeling, ensure proper interpretation of the consumption systems and identify relevant mixtures of pesticides.

The two different estimations proposed to extract the mixtures of pesticides are useful in different contexts. Individuals can potentially be exposed to mixtures of contaminants through several pathways of exposure. For example, individuals can be exposed to organophosphorus pesticides through indoor and outdoor air, drinking water, foods, soil and house dust (Lu et al., 2004). The exposure based estimation implemented in this work in the

Table 1
Result of the clustering of the individuals: number of individuals (*N*), characteristics of the individuals of the cluster (variables with * present a mean or a proportion significantly different from the whole population i.e., *p* value < 0.05), consumption systems contributing at least at 10% to the consumption of the individuals, the 15 major foods of these CS and the major pesticides associated to these CS (%pest > 10%).

Cluster	Description	Major CS	Major foods	% Food	Major pesticides (mixtures)	% Pest
1	*N* = 696	6 (42%)	Beef steak	4.6	Endosulfan (sum)	11.3
			Pasta	4.2	Dichlorvos	11.2
	Age = 43*		Baguette	3.8	Pirimiphos-methyl	10.3
			Cooked ham	3.0	Carbofuran	10.1
	BMI = 24.8		Butter	3.0		
			Egg	2.8		
	%Men = 61*		Coffee	2.6		
			Mashed potatoes	2.5		
	Income*:		Sauteed potatoes of chips	2.5		
	%Low = 33		Gruyere	2.4		
	%Medium = 39		Camembert	2.3		
	%High = 9		Bean	2.2		
			Spring water	2.1		
			Pork chop	2.1		
			Turkey cutlet	2.0		
2	*N* = 708	4 (37%)	Vegetable soup	3.7	Imazalil	12.8
			Apple	3.4	Pirimiphos-methyl	10.1
	Age = 53*		Endive	2.7		
			Carrot	2.6		
	BMI = 24.3		Kiwi	2.5		
			Lettuce	2.4		
	%Women = 65*		Honney	2.4		
			Clementine	2.4		
	Income*:		Wholemeal bread	2.3		
	%Low = 25		Apple compote	2.3		
	%Medium = 39		Tea or herbal tea	2.2		
	%High = 20		Salmon	2.2		
			Soy drink	2.0		
			Fresh orange	1.9		
			Spinach	1.8		
	6 (14%)		See Cluster 1		See Cluster 1	
3	*N* = 445	2 (36%)	Peach	5.2	Carbendazim (sum)	12.1
			Tomato	4.9	Phosalone	11.0
	Age = 51*		Cucumber	4.2		
			Apricot	4.2		
	BMI = 25.3*		Courgette	3.4		
			Bean	3.3		
	%Women = 71*		Boiled egg	2.7		
			Tap water	2.2		
	Income*:		Lettuce	2.1		
	%Low = 24		Tabbouleh	2.0		
	%Medium = 38		Canned tuna	1.9		
	%High = 21		Cherry	1.9		
			Orange juice	1.8		
			Ham	1.8		
		6 (12%)	See Cluster 1		See Cluster 1	
4	*N* = 344	5 (35%)	Wine	6.2	Ethoxyquin	18.4
			Champagne	3.8	Ethion	15.2
	Age = 49*		Coffee	3.4	Imazalil	14.3
			Beer	3.0	Chlorothalonil	12.5
	BMI = 24.9		Ham	3.0		
			Baguette	2.7		
	%Men = 76*		Dried sausage	2.7		
			Perrier	2.6		
	Income*:		Poutry	2.6		
	%Low = 18		Spring water National brand 1	2.5		
	%Medium = 42 %High = 23		Roquefort	2.4		
			Oilseed	2.1		
			Pâté	2.0		
			Aperitif biscuit	2.0		
			Goat cheese	1.8		
		6 (16%)	See Cluster 1		See Cluster 1	
5	*N* = 128	10 (35%)	Strawberry	7.7	Malathion (sum)	12.9
			Radish	7.4	Ethoxyquin	10
	Age = 54*		Liver	3.3		
			Camembert	3.2		
	BMI = 26.6*		Veal	3.2		
			Tuna	3.0		
	%Women = 59*		Low-fat butter	2.8		
			Cherry	2.1		

Table 1 (*continued*)

Cluster	Description	Major CS	Major foods	% Food	Major pesticides (mixtures)	% Pest
	Income*:		Leek	2.1		
	%Low = 35		Wine	2.1		
	%Medium = 33		Fromage blanc	1.9		
	%High = 14		Chipolata	1.7		
		6 (12%)	See Cluster 1		See Cluster 1	
6	*N* = 303	1 (51%)	Pizza	4.0	Carbofuran	11.5
			Hamburger	3.9	Chlorpyrifos-ethyl	11.2
	Age = 28*		Soda	3.8	Dichlorvos	10.4
			Sandwich	3.2	Endosulfan (sum)	10.4
	BMI = 22.3*		Sauteed potatoes or chips	3.1	Phosmet	10.0
			Semi-skimmed milk	2.8		
	%Men = 51		Instant hot chocolate drink	2.8		
			Mayonnaise	2.5		
	Income*:		Orange juice	2.5		
	%Low = 24		Ravioli	2.5		
	%Medium = 48		Spicy lamb sausage (merguez)	2.4		
	%High = 9		Lemonade	2.3		
			Salted potato crisps	2.2		
			Ice tea	2.1		
			Chocolate croissant	2.0		
		6 (21%)	See Cluster 1		See Cluster 1	

context of dietary exposure is also relevant in this context of aggregated exposure. Through the clustering of individuals, this work highlighted populations who present a specific exposure to one or more mixtures which are explained by specific consumption habits. The characterization of these populations and the identification of the foods involved in their exposure are directly provided when the consumption based estimation is implemented. Therefore this choice allows for an easier interpretation of the mixtures and also to characterize risks according to food consumption patterns, which may constitute a useful tool to establish risk controlling measures when necessary. Moreover once consumption systems are extracted for a population, they can also be used to directly determine mixtures of other substances or to update mixtures with new data by combining with appropriate residue levels data. The method detailed in this work can thus be an interesting approach to go further with the cluster diet proposed by Sy et al. (2013). Sy et al. have summarized a world food consumption database into several consumption systems and then grouped together countries with similar pattern of consumption in 17 cluster diets composed of 2–30 countries. By combining national or international residue levels data with the consumptions systems associated with these clusters through the method proposed in this work, the mixtures of contaminants to which groups of countries are exposed could be extracted. However, this would require data from many countries and most are not available yet.

The choice of the scenario to handle left-censored data is a major issue when dealing with exposure to pesticides. The upper-bound scenario was tested but leads to model the limits of detection (LOD) which does not make much sense. Consequently, the lower-bound scenario is preferred. However with this scenario, it is possible that some pesticides that were not currently detected but occur in some foods were not considered. In this work, NMF was first applied to consumption data to extract the consumption systems representing the different dietary patterns among the population. These consumption systems were then combined with the residue levels data and a clustering method was used to provide 6 mixtures of pesticides related to 6 consumption systems characteristic of some specific populations. The only previous work that identified mixtures of pesticides to which the French population is exposed is Crépet et al. (2013) and the mixtures provided by the two works are different (none of the mixture is common). The same consumption data were used in both work but the residue levels data used in this work come from the Second French Total Diet Study whereas residue levels data used in Crépet et al. (2013) came from the annual monitoring programs implemented in 2006 by the French ministries in charge of consumer affairs, agriculture and health. There are many differences between these two residue levels data which can explain the differences in the results: the sample design, the date of sampling, the analytical methods and the samples preparation Therefore, residue levels in food but also pesticides considered are different. The model proposed in this work was also implemented with $K = 7$ (i.e., the number of mixture provided by Crépet et al. (2013)) on the combined exposure matrix used in Crépet et al. (2013). It results in the same mixtures for both methods, which constitutes a nice validation for these two data driven approaches and highlights that the choice of residue levels data is paramount in the identification of mixture to which the population is exposed.

Conflict of Interest

The authors declare that there are no conflicts of interest.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bertail, P., Tressou, J., 2006. Incomplete generalized U-statistics for food risk assessment. Biometrics 62, 66–74.

Boobis, A.R., Ossendorp, B.C., Banasiak, U., Hamey, P.Y., Sebestyen, I., Moretto, A., 2008. Cumulative risk assessment of pesticide residues in food. Toxicol. Lett. 15, 137–150.

Boobis, A., Budinsky, R., Collie, S., Crofton, K., Embry, M., Felter, S., Hertzberg, R., Kopp, D., Mihlan, G., Mumtaz, M., Price, P., Solomon, K., Teuschler, L., Yang, R., Zaleski, R., 2011. Critical analysis of literature on low-dose synergy for use in screening chemical mixtures for risk assessment. Crit. Rev. Toxicol. 41 (5), 369–383.

Bosgra, S., van der Voet, H., Boon, P.E., Slob, W., 2009. An integrated probabilistic framework for cumulative risk assessment of common mechanism chemicals in food: an example with organophosphorus pesticides. Regul. Toxicol. Pharmacol. 54, 124–133.

Crépet, A., Tressou, J., 2010. Bayesian nonparametric model for clustering individual co-exposure to pesticides founds in the French diet. Bayesian Anal., 127–144.

Crépet, A., Tressou, J., Graillot, V., Béchaux, C., Pierlot, S., Héraud, F., Leblanc, J.C., 2013. Identification of the main pesticide residue mixtures to which the French population is exposed. Environ. Res., <http://dx.doi.org/10.1016/j.envres.2013.03.008i>.

Dubuisson, C., Lioret, S., Touvier, M., Dufour, A., Calamassi-Tran, G., Volatier, J.L., Lafay, L., 2010. Trends in food and nutritional intakes of French adults from 1999 to 2007: results from the INCA surveys. Br. J. Nutr. 103, 1035–1048.

EFSA, 2008. Opinion of the Scientific Panel on Plant Protection products and their Residues to evaluate the suitability of existing methodologies and, if

appropriate, the identification of new approaches to assess cumulative and synergistic risks from pesticides to human health with a view to set MRLs for those pesticides in the frame of Regulation (EC) 396/20051, pp. 1–84.
EFSA, 2012. Scientific Report on Investigation of the state of the science on combined actions of chemicals in food through dissimilar modes of action and proposal for science-based approach for performing related cumulative risk assessment. Supporting Publications 2012:EN-232. Available from: www.efsa.europa.eu.
Fung, T.T., Willett, W.C., Stampfer, M.J., Manson, J.E., Hu, F.B., 2001. Dietary patterns and the risk of coronary heart disease in women. Arch. Int. Med. 161, 1857–1862.
Gaujoux, R., Seoighe, C., 2010. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11, 367.
Hu, F.B., Rimm, E.B., Stampfer, M.J., Ascherio, A., Spiegelman, D., Willett, W.C., 2000. Prospective study of major dietary patterns and risk of coronary heart disease in men. Am. J. Clin. Nutr. 72, 912–921.
Kaufman, L., Rousseeuw, P., 1990. Finding Groups in Data: An Introduction to Cluster Analysis Wiley.
Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791.
Lee, D.D., Seung, H.S., 2001. Algorithms for non-negative matrix factorization. Adv. Neural Inf. Process. Syst., 556–562.
Lioret, S., Dubuisson, C., Dufour, A., Touvier, M., Calamassi-Tran, G., Maire, B., Volatier, J.L., Lafay, L., 2010. Trends in food intake in French children from 1999 to 2007: results from the INCA (etude Individuelle Nationale des Consommations Alimentaires) dietary surveys. Br. J. Nutr. 103, 585–601.
Lu, C., Kedan, G., Fisker-Andersen, J., Kissel, J.C., Fenske, R.A., 2004. Multipathway organophosphorus pesticide exposures of preschool children living in agricultural and nonagricultural communities. Environ. Res. 96, 283–289.
Menard, C., Heraud, F., Nougadere, A., Volatier, J.-L., Leblanc, J.-C., 2008. Relevance of integrating agricultural practices in pesticide dietary intake indicator. Food Chem. Toxicol. 46, 3240–3253.
Muller, A.K., Bosgra, S., Boon, P.E., van der Voet, H., Nielsen, E., Ladefoged, O., 2009. Probabilistic cumulative risk assessment of anti-androgenic pesticides in food. Food Chem. Toxicol. 47, 2951–2962.
Nougadère, A., Sirot, V., Kadar, A., Fastier, A., Truchot, E., Vergnet, C., Hommet, F., Baylé, J., Gros, P., Leblanc, J.-C., 2012. Total diet study on pesticide residues in France: levels in food as consumed and chronic dietary risk to consumers. Environ. Int. 45, 135–150.
Oslera, M., Heitmanna, B., Gerdesa, L., Jørgensen, L., Schroll, M., 2001. Dietary patterns and mortality in Danish men and women: a prospective observational study. Br. J. Nutr. 85, 219–225.
Reffstrup, T.K., Larsen, J.C., Meyer, O., 2010. Risk assessment of mixtures of pesticides. Current approaches and future strategies. Regul. Toxicol. Pharmacol. 56, 174–192.
Saul, L.K., Lee, D.D., 2002. Multiplicative updates for classification by mixture models. Adv. Neural Inf. Process. Syst. 14.
Sirot, V., Volatier, J.L., Calamassi-Tran, G., Dubuisson, C., Ménard, C., Dufour, A., Leblanc, J.C., 2009. Core food of the French food supply: second Total Diet Study. Food Addit. Contam. 26, 623–639.
Sy, M.M., Feinberg, M., Verger, P., Barré, T., Clémençon, S., Crépet, A., 2013. New approach for the assessment of cluster diets. Food Chem. Toxicol. 52, 180–187.
WHO, 2009. Assessment of combined exposures to multiple chemicals: report of a WHO/IPCS international workshop on aggregate/cumulative risk assessment.
Zetlaoui, M., Feinberg, M., Verger, P., Clémençon, S., 2011. Extraction of food consumption systems by nonnegative matrix factorization (NMF) for the assessment of food choices. Biometrics 67, 1647–1658.

4.2.2 Toxicokinetic/dynamic interactions

The next step for considering the exposure-effect continuum in a context of combined exposure is the toxicokinetic. Indeed, in order to move from external combined exposure to internal dose, information on potential interaction in toxicokinetic is the key element. Again, PBTK modeling seems to become a particularly useful tool to investigate and model such interaction. Currently, computational modeling of the absorption, distribution, metabolism, and excretion of chemicals is now theoretically able to describe metabolic interactions in realistic mixtures of tens to hundreds of substances (Cheng and Bois, 2011). Such approach combines generic substance-independent knowledge from PBTK modeling, metabolic networks modeling and detailed description of enzymatic reactions (Bois 2010). The method was validated by Cheng and Bois (2011) on a mixture of benzene, toluene, ethylbenzene, m-xylene using Markov chain Monte Carlo simulations and single-substance exposure data. In the same spirit, Sasso et al. (2010) developed a generalized PBTK model for mixtures using a chemical independent approach based on modules that can be directly mapped to individual TK models for specific chemicals, while maintaining physiological consistency across different chemicals. The model was applied for validation to a mixture of methylmercury, cadmium, lead, arsenic (and metabolites), toluene, and benzene.

The role of the cytochrome P450 (CYP450) is also crucial when considering kinetics interaction. Indeed, many chemicals have been shown inducing or inhibiting CYP450. For example, cadmium induces 2A6, 2E1 and 2C9 or lead inhibits 2A6 (Satarug et al., 2004; 2006). Consequently, the kinetic of other substrates of these enzymes will be affected by the presence of such interactor. This kind of interaction can typically be modeled in a PBTK model. Regarding the inter-individual variability, as demonstrated in our case study on CYP2D6 for single a compound, the polymorphism is source of high variability. This is even truer when considering interaction. Indeed, if an inhibitor slows down the metabolism capacity of a polymorphic enzyme, the metabolism of another substrate of the enzyme individual would be even more slowed in a poor metabolizer. In the same way, if the inhibitor blocks the activity of the enzyme and the individual is poor metabolizer for the alternative pathway, the metabolism would slow down drastically. Therefore, extend the investigation of the inter-variability in kinetic related to polymorphism in P450 enzymes would be useful to inform PBTK model and risk assessment.

Applying TK models like PBTK models to mixture of several chemicals with high probability of co-exposure may lead to predict more realistic internal doses. It could then be interesting to compare internal doses with biomonitoring data. Indeed, in the work presented in this manuscript

biomonitoring data informed exposure in case of single and cumulative exposure; in the same way it could inform on combined exposure.

Finally, the last step of the continuum to achieve a complete risk assessment is the toxicodynamic (TD) that links internal dose with response is needed. Several PBTK/TD models have been recently developed to address combined exposure to multiple chemicals. For example, focusing on pesticide as it is our case study for combined exposure, PBTK/TD models for organophosphates (Knaak et al., 2004) and carbamates (Knaak et al., 2008) are available. Especially, a model has been developed for chlorpyrifos and diazinon which is an association we highlighted as forming a mixture in our work on combined exposure to pesticides in France. It could thus be interesting to apply such model to the co-exposure to chlorpyrifos and diazinon found in the population. The model was built from previous models for the individual chemicals. Then interactions in number of important metabolic steps such as CYP450 mediated activation/detoxification, B-esterases, butyrylcholinesterase (B-E) and AchE or PON-1 oxon detoxification were added. Since both insecticides were shown to inhibit the CYP-mediated metabolism *in vitro* in a concentration-dependent manner, the PBPK model was modified to reflect the TK of the CYP inhibition (competitive vs. non-competitive). In addition, B-esterase metabolism was described as dose-additive, and no PON-1 interactions were assumed (Timchalk and Poet, 2008). This binary model was then compared with previously published rodent oral TK data and TD data (AchE inhibition) for co-exposure to chlorpyrifos and diazinon. No differences between predicted TK and published TK data were shown for either chlorpyrifos or diazinon or their respective metabolites, while TD AchE inhibition was shown to be described using dose-addition. The authors concluded from this model that, at low environmentally relevant binary doses of chlorpyrifos and diazinon, the TK and TD of the mixture were expected to be linear and dose-additive (Timchalk and Poet, 2008). Indeed, PBTK and PBTK/TD models can also be useful tool to validate or invalidate previous assumption made for the risk assessment of chemicals like dose addition.

4.2.3 Proposition of a case study to be investigated

In order to extend our work to combined exposure, a very interesting case study may be the combined exposure to methylmercury, cadmium, lead and arsenic. Indeed, a PBTK model that account for TK interactions is available and has been validated by Sasso et al. (2010). Moreover, due to their persistence and bioaccumulation in water and sediment, methylmercury, cadmium, lead and arsenic are often found together in fish when looking at the contamination level (Noël et al., 2013).

Some exposure study also highlighted that consumption of seafood leads to a higher exposure to these four chemicals. Consequently, the co-exposure may be high especially regarding individuals with high consumption of fish and seafood. Cigarette smoke is also a major source of exposure to cadmium and lead, thus an aggregated exposure should be considered. Moreover, since such consumers constitute a population that is much studied through the world because of its specific exposure, many biomonitoring data are available for this population. For example, the Calipso study measured the internal dose through biomarker of exposure to these four substances in a population of French high fish consumers (Leblanc, 2006). Dietary exposure was also assessed for these individuals from the record of dietary habits and contamination level in fish and seafood (Leblanc, 2006). Exposure from cigarette smoke could be assessed as we did in the cadmium case study from individual smoking status recorded in the Calipso study. Furthermore, Sasso et al. (2010) highlighted that toxic metals may affect the toxicokinetics of other chemicals. Among them, they cited the PCBs to which high consumers of fish and seafood are also exposed. This interaction is explained by the accumulation of the toxic metals in the liver and kidneys due to their long half-lives, which affects the hepatic and renal levels of Cytochrome P450 enzymes and therefore the metabolism of the other chemicals. Other interactions can also be expected and should be studied at the Cytochrome P450 level through this example since many chemicals are known to induce or inhibit P450 including the PCBs.

5 References

Åkesson A, Berglund M, Schütz A, Bjellerup P, Bremme K, Vahter M. 2002. Cadmium exposure in pregnancy and lactation in relation to iron status. American Journal of Public Health, 92, 284–287.

Albert, I., Villeret, G., Paris, A. and Verger, P. 2010. Integrating variability in half-lives and dietary intakes to predict mercury concentration in hair. Regulatory Toxicology and Pharmacology, 58(3), 482-489.

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K. and Rousseau, J. 2012. Combining expert opinions in prior elicitation. Bayesian Analysis, 7(3), 503-532.

Albertini, R., Bird, M., Doerrer, N., Needham, L., Robison, S., Sheldon, L. and Zenick, H. 2006. The use of biomonitoring data in exposure and human health risk assessments. Environmental health perspectives, 1755-1762.

Alqahtani S, Mohamed LA and Kaddoumi A, 2013. Experimental models for predicting drug absorption and metabolism. Expert Opinion on Drug Metabolism and Toxicology, 9, 1241-1254.

Anses/InVS, 2011. Etude nationale d'imprégnation aux polychlorobiphényles des consommateurs de poissons d'eau douce (ICAR-PCB). Rapport d'étude scientifique et synthèse. Maisons-Alfort, Saint-Maurice, France.

Amzal B, Julin B, Vahter M, Wolk A, Johanson G, Åkesson A. 2009. Population toxicokinetic modeling of cadmium for health risk assessment. Environmental health perspectives, 117(8), 1293.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE and Villeneuve DL, 2010. Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment. Environmental Toxicology and Chemistry, 29, 730-741.

Aylward LL, Collins JJ, Bodner K., Wilken M and Bodnar CM. Elimination rates of dioxin congeners in former chlorophenol workers from Midland, Michigan. Environ Health Perspect 2013; 121(1), 39.

Angerer, J., Ewers, U. and Wilhelm, M. 2007. Human biomonitoring: state of the art. International journal of hygiene and environmental health, 210(3), 201-228.

Beaumont M, Zhang W, Balding D. 2002. Approximate Bayesian computation in population genetics. Genetics;162:2025–35.

Beaumont, M. A., Cornuet, J. M., Marin, J. M. and Robert, C. P. 2009. Adaptive approximate Bayesian computation. Biometrika, asp052.

Béchaux, C., Zetlaoui, M., Tressou, J., Leblanc, J. C., Héraud, F., & Crépet, A. (2013). Identification of pesticide mixtures and connection between combined exposure and diet. *Food and Chemical Toxicology*, *59*, 191-198.

Béchaux, C., Crépet, A., & Clémençon, S. (2014a). Improving Dietary Exposure Models by Imputing Biomonitoring Data through ABC Methods. *The international journal of biostatistics*. In press.

Béchaux, C., Zeilmaker, M., Merlo, M., Bokkers, B., & Crépet, A. (2014b). An integrative risk assessment approach for persistent chemicals: a case study on dioxins, furans and dioxin-like PCBs in France. *Regulatory Toxicology and Pharmacology*. 70, 261-269

Béchaux, C., Bodin, L., Clémençon, S., & Crépet, A. (2014c). PBPK and population modelling to interpret urine cadmium concentrations of the French population. *Toxicology and applied pharmacology, 279, 364–372.*

Bernillon P, Bois FY. 2000. Statistical issues in toxicokinetic modeling: a Bayesian perspective. Environmental health perspectives 108(suppl 5):883–893.

Bertail P, Clémençon S, Tressou J. A storage model with random release rate for modeling exposure to food contaminants. Math Biosci Eng 2008;35:35–60.

Bois, F. Y. (2009). GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. Bioinformatics, 25(11), 1453-1454.

Bois, F. Y. (2010). Physiologically based modelling and prediction of drug interactions. Basic & clinical pharmacology & toxicology, 106(3), 154-161.

Bois F. Y, Jamei M, Clewell H. J. 2010. PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology, 278(3), 256-267.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Vickers C, Willcocks D and Farland W, 2006. IPCS framework for analysing the relevance of a cancer mode of action for humans. Toxicology Letters, 164, S254-S255.

Boobis, A. R., Ossendorp, B. C., Banasiak, U., Hamey, P. Y., Sebestyen, I. and Moretto, A. 2008. Cumulative risk assessment of pesticide residues in food. Toxicology letters, 180(2), 137-150.

Centers for disease Control & Prevention. 2000. National health and nutrition examination survey. http://www.cdc.gov/nchs/about/major/nhanes/nhanes9900:htmS

Cheng S and Bois FY, 2011. A Mechanistic Modeling Framework for Predicting Metabolic Interactions in Complex Mixtures. Environmental Health Perspectives, 119, 1712-1718.

Choudhury H, Harvey T, Thayer W.C, Lockwood T.F. 2001. Urinary cadmium elimination as a biomarker of exposure for evaluating a cadmium dietary exposure-biokinetics model. Journal of Toxicology and Environmental Health A 63, 321–350.

Coecke S, Pelkonen O, Leite SB, Bernauer U, Bessems JGM, Bois FY, Gundert-Remy U, Loizou G, Testai E and Zaldivar JM, 2013. Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches. Toxicology in Vitro, 27, 1570-1577.

Commission Regulation (EC) No 1488/94 of 28 June 1994 laying down the principles for the assessment of risks to man and the environment of existing substances in accordance with Council Regulation (EEC) No 793/93.

Crépet, A., Denys, S. and Hulin, M. 2012. Exposition agrégée, combinée, cumulée à des substances chimiques et risques associés. Environnement risques & santé, 11(5), 424-425.

Crépet, A., Tressou, J., Graillot, V., Béchaux, C., Pierlot, S., Héraud, F., and Leblanc, J. C. 2013. Identification of the main pesticide residue mixtures to which the French population is exposed. Environmental research, 126, 125-133.

Csilléry K, Blum M, Gaggiotti O, François O. 2010. Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol;25:490–1.

Diamanti Kandarakis E, Bourguignon JP, Giudice LC, Hauser R, Prins GS, Soto AM, Zoeller RT, Gore AC. 2009. Endocrine disrupting chemicals: an endocrine society scientific statement. Endocr Rev. 30:293–342.

Doerge DR, Twaddle NC, Woodling KA and Fisher JW, 2010. Pharmacokinetics of bisphenol A in neonatal and adult rhesus monkeys. Toxicology and Applied Pharmacology, 248, 1-11.

Dubuisson C, Lioret S, Touvier M, Dufour A, Calamassi-Tran G, Volatier JL. Trends in food and nutritional intakes of French adults from 1999 to 2007: results from the INCA surveys. Br J Nutr 2010; 103: 1035-1048.

EFSA (European Food Safety Authority), 2011. Use of the EFSA Comprehensive European Food Consumption Database in Exposure Assessment. EFSA Journal (2011);9(3):2097.

EFSA (European Food Safety Authority), 2014. Modern methodologies and tools for human hazard assessment of chemicals. EFSA Journal 2014;12(4):3638, 87 pp.

EPA (US Environmental Protection Agency), 2006a. Organophosphorus cumulative risk assessment. 2006 update. Washington, DC.

EPA (US Environmental Protection Agency), 2006b. Cumulative risk from triazine pesticides.Washington, DC.

EPA (U.S. Environmental Protection Agency), 2014. Guidance for Applying Quantitative Data to Develop Data Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. EPA/100/R-11/003F.

Falq, G., Zeghnoun, A., Pascal, M., Vernay, M., Le Strat, Y., Garnier, R. and Fréry, N. 2011. Blood lead levels in the adult population living in France the French Nutrition and Health Survey (ENNS 2006–2007). Environment international, 37(3), 565-571.

Fantino M. 2005. Etude SFAE sur la consommation alimentaire des nourrissons et enfants en bas âge français de 1 mois à 36 mois – Analyse des données nutritionnelles. Not published, Etude SOFRES 2005/Université de Bourgogne – Pr M. Fantion for the French Association for Children's Food.

Filippi, S., Barnes, C. P., Cornebise, J. and Stumpf, M. P. 2013. On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. Statistical applications in genetics and molecular biology, 12(1), 87-107.

Fisher JW, Twaddle NC, Vanlandingham M and Doerge DR, 2011. Pharmacokinetic modeling: Prediction and evaluation of route dependent dosimetry of bisphenol A in monkeys with extrapolation to humans. Toxicology and Applied Pharmacology, 257, 122-136.

Freire, C. and Koifman, S. 2012. Pesticide exposure and Parkinson's disease: epidemiological evidence of association. Neurotoxicology, 33(5), 947-971.

Fromme, H., Körner, W., Shahin, N., Wanner, A., Albrecht, M., Boehmer, S. and Bolte, G. 2009. Human exposure to polybrominated diphenyl ethers (PBDE), as evidenced by data from a duplicate diet study, indoor air, house dust, and biomonitoring in Germany. *Environment international*, *35*(8), 1125-1135.

Gelfand, A. and Smith, A. 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398-409

Gelman A, Bois F, Jiang JM. 1996. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. J Am Stat Assoc 91(436):1400–1412.

Geman, S. and Geman, D. 1984. Stochastic relaxation, gibbs distribution and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6, 721-741.

Hastings, W. 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57, 97-109.

Hulin, M., Bemrah, N., Nougadère, A., Volatier, J. L., Sirot, V and Leblanc, J. C. 2014. Assessment of infant exposure to food chemicals: the French Total Diet Study design. Food Additives & Contaminants: Part A, (just-accepted).

Järup L, Åkesson A. 2009. Current status of cadmium as an environmental health problem. Toxicology and applied pharmacology, 238(3), 201-208.

Knaak JB, Dary CC, Okino MS, Power FW, Zhang X, Thompson CB, Tornero-Velez R and Blancato JN, 2008. Parameters for Carbamate Pesticide QSAR and PBPK/PD Models for Human Risk Assessment. Review of Environmental Contaminants in Toxicology, 193, 53-212.

Knaak JB, Dary CC, Power F, Thompson CB and Blancato JN, 2004. Physicochemical and biological data for the development of predictive organophosphorus pesticide QSARs and PBPK/PD models for human risk assessment. Critical Reviews in Toxicology, 34, 143-207.

Krauss, M., Burghaus, R., Lippert, J., Niemi, M., Neuvonen, P., Schuppert, A., Görlitz, L. (2013). Using Bayesian-PBPK modeling for assessment of inter-individual variability and subgroup stratification. In Silico Pharmacology, 1(1), 1-11.

Kusuhara H and Sugiyama Y, 2009. In vitro-in vivo extrapolation of transporter-mediated clearance in the liver and kidney. Drug Metabolism and Pharmacokinetics, 24, 37-52.

Leblanc JC. 2006. CALIPSO, fish and seafood consumption study and biomarker of exposure to trace elements, pollutants and omega3.

Loccisano AE, Campbell JL, Andersen ME and Clewell HJ, 2011. Evaluation and prediction of pharmacokinetics of PFOA and PFOS in the monkey and human using a PBPK model. Regulatory Toxicology and Pharmacology, 59, 157-175.

Loccisano, A. E., Longnecker, M. P., Campbell Jr, J. L., Andersen, M. E., & Clewell III, H. J. (2013). Development of PBPK models for PFOA and PFOS for human pregnancy and lactation life stages. Journal of Toxicology and Environmental Health, Part A, 76(1), 25-57.

Loizou G and Hogg A, 2011. MEGen: A Physiologically Based Pharmacokinetic Model Generator. Frontiers in Pharmacology, 2, 56.

Lorber M. A pharmacokinetic model for estimating exposure of Americans to dioxin-like compounds in the past, present, and future. 2002. Sci Total Environ; 288(1-2): 81-95.

Lorber M. 2007. Exposure of Americans to polybrominated diphenyl ethers. Journal of Exposure Science and Environmental Epidemiology, 18(1), 2-19.

Lyons M.A, Yang R.S, Mayeno A.N, Reisfeld B. 2008. Computational toxicology of chloroform: reverse dosimetry using Bayesian inference, Markov chain Monte Carlo simulation, and human biomonitoring data. Environmental health perspectives, 116(8), 1040-1046.

Marjoram P. Markov Chain Monte Carlo without likelihoods. 2003. Proc Natl Acad Sci USA;100:15324–8.

McNally, K., Cotton, R., Cocker, J., Jones, K., Bartels, M., Rick, D. and Loizou, G. 2012. Reconstruction of exposure to m-xylene from human biomonitoring data using PBPK modelling, Bayesian inference, and Markov Chain Monte Carlo simulation. Journal of toxicology, 2012.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J and Vickers C, 2013. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. Journal of Applied Toxicology, 34, 1-18.

Metropolis, N., Rosenbluth, A., Rosenbluth,M., Teller, A and Teller, E. 1953. Equation of state calculations by fast computing machine. Journal of Chemical Physics 21, 1087-1091.

Meyer MR, Orschiedt T and Maurer HH, 2013. Michaelis-Menten kinetic analysis of drugs of abuse to estimate their affinity to human P-glycoprotein. Toxicology Letters, 217, 137-142.

Milbrath MO, Wenger Y, Chang CW, Emond C, Garabrant D, Gillespie BW et al. Apparent half-lives of dioxins, furans, and polychlorinated biphenyls as a function of age, body fat, smoking status, and breast-feeding. Environ Health Perspect 2009; 117(3): 417-425.

Mumtaz, M., Fisher, J., Blount, B., & Ruiz, P. (2012). Application of Physiologically Based Pharmacokinetic Models in Chemical Risk Assessment. Journal of toxicology, 2012.

National Research Council (NRC). Human Biomonitoring for Environmental Chemicals. Washington, DC: The National Academies Press, 2006.

Noël, L., Chekri, R., Millour, S., Merlo, M., Leblanc, J. C. and Guérin, T. 2013. Distribution and relationships of As, Cd, Pb and Hg in freshwater fish from five French fishing areas. Chemosphere, 90(6), 1900-1910.

OECD (Organisation for Economic Co-Operation and Development), 2013. Guidance document on developing and assessing adverse ouctome pathways. Paris, France, OECD Environment Directorate. OECD Environment, Health and Safety Publications. Series on Testing and Assessment. No 184.

Pouillot, R., Albert, I., Cornu, M. and Denis, J. B. 2003. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to Listeria monocytogenes. International journal of food microbiology, 81(2), 87-104.

Renwick, A.G., 1993. Data-derived safety factors for the evaluation of food additives and environmental contaminants. Food Additives and Contaminants 10, 275–305.

Sasso AF, Isukapalli SS and Georgopoulos PG, 2010. A generalized physiologically-based toxicokinetic modeling system for chemical mixtures containing metals. Theoretical Biology and Medical Modelling, 7. doi: 10.1186/1742-4682-7-17

Sasso AF, Georgopoulos PG, Isukapalli SS and Krishnan K, 2012. Bayesian Analysis of a Lipid-Based Physiologically Based Toxicokinetic Model for a Mixture of PCBs in Rats. Journal of Toxicology, 2012, 895391.

Satarug S, Ujjin P, Vanavanitkun Y, Nishijo M, Baker JR, Moore MR: Effects of cigarette smoking and exposure to cadmium and lead on phenotypic variability of hepatic CYP2A6 and renal function biomarkers in men. Toxicology 2004, 204(2-3):161-73.

Satarug S, Nishijo M, Lasker JM, Edwards RJ, Moore MR: Kidney dysfunction and hypertension: role for cadmium, p450 and heme oxygenases? Tohoku J Exp Med 2006, 208(3):179-202.

Schmidt CW, 2009. TOX 21: new dimensions of toxicity testing. Environmental Health Perspectives, 117(8), A348-353.

Scholz S, Sela E, Blaha L, Braunbeck T, Galay-Burgos M, Garcia-Franco M, Guinea J, Kluver N, Schirmer K, Tanneberger K, Tobor-Kaplon M, Witters H, Belanger S, Benfenati E, Creton S, Cronin MT, Eggen RI, Embry M, Ekman D, Gourmelon A, Halder M, Hardy B, Hartung T, Hubesch B, Jungmann D, Lampi MA, Lee L, Leonard M, Kuster E, Lillicrap A, Luckenbach T, Murk AJ, Navas JM, Peijnenburg W, Repetto G, Salinas E, Schuurmann G, Spielmann H, Tollefsen KE, Walter-Rohde S, Whale G, Wheeler JR and Winter MJ, 2013. A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment. Regulatory Toxicology and Pharmacology, 67, 506-530.

Schulz, C., Conrad, A., Becker, K., Kolossa-Gehring, M., Seiwert, M., and Seifert, B. 2007. Twenty years of the German Environmental Survey (GerES): human biomonitoring–temporal and spatial (West Germany/East Germany) differences in population exposure. International journal of hygiene and environmental health, 210(3), 271-297.

Shuey DL, Lau C, Logsdon TR, Zucker RM, Elstein KH, Narotsky MG, Setzer RW, Kavlock RJ and Rogers JM, 1994. Biologically-Based Dose-Response Modeling in Developmental Toxicology - Biochemical and Cellular Sequelae of 5-Fluorouracil Exposure in the Developing Rat. Toxicology and Applied Pharmacology, 126, 129-144.

Singh, N. K., Banerjee, B. D., Bala, K., Basu, M., & Chhillar, N. (2014). Polymorphism in Cytochrome P450 2D6, Glutathione S-Transferases Pi 1 Genes, and Organochlorine Pesticides in Alzheimer Disease A Case–Control Study in North Indian Population. Journal of geriatric psychiatry and neurology, 0891988714522698.

Sirot, V., Guérin, T., Mauras, Y., Garraud, H., Volatier, J. L. and Leblanc, J. C. 2008. Methylmercury exposure assessment using dietary and biomarker data among frequent seafood consumers in France: CALIPSO study. Environmental research, 107(1), 30-38.

Sirot V, Tard A, Venisseau A, Brosseaud A, Marchand P, Le Bizec B et al. Dietary exposure to polychlorinated dibenzo-p-dioxins, polychlorinated dibenzofurans and polychlorinated biphenyls of the French population: results of the second French Total Diet Study. Chemosphere 2012 ; 88(5):492-500.

Sisson S, Fan Y, Tanaka M. Sequential Monte Carlo without likelihoods. 2007. Proc Natl Acad Sci USA;104:1760–5.

Sly PD, Flack F. 2008. Susceptibility of children to environmental pollutants. Ann N Y Acad Sci. 1140:163–183.

Timchalk C and Poet TS, 2008. Development of a physiologically based pharmacokinetic and pharmacodynamic model to determine dosimetry and cholinesterase inhibition for a binary mixture of chlorpyrifos and diazinon in the rat. Neurotoxicology, 29, 428-443.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M. P. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface, 6(31), 187-202.

Vahter M, Åkesson A, Lidén C, Ceccatelli C, Berglund M. 2007. Gender differences in the disposition and toxicity of metals. Environmental Research, 104, 85–95.

Valcke M and Krishnan K, 2014. Characterization of the human kinetic adjustment factor for the health risk assessment of environmental contaminants. Journal of Applied Toxicology, 34(3): 227-240.

Van den Berg, M., Birnbaum, L. S., Denison, M., De Vito, M., Farland, W., Feeley, M. and Peterson, R. E. 2006. The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicological sciences, 93(2), 223-241.

Van Leeuwen, FXR and MM Younes. 2000. Consultation on assessment of the health risk of dioxins: re-evaluation of the tolerable daily intake (TDI): Executive summary. Food Addit Contam; 17 (4):223-240.

Verger P, Tressou J, Clémençon S. 2007. Integration of time as a description parameter in risk characterisation: application to methyl mercury. Regul Toxicol Pharmacol;49:25–30.

Verner MA, Ayotte P, Muckle G, Charbonneau M and Haddad S, 2009. A Physiologically Based Pharmacokinetic Model for the Assessment of Infant Exposure to Persistent Organic Pollutants in Epidemiologic Studies. Environmental Health Perspectives, 117, 481-487.

Wegmann, D., Leuenberger, C., and Excoffier, L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics, 182(4), 1207-1218.

WHO (World Health Organization), 1995. Application of Risk Analysis to Food Standards Issues, a Joint FAO/WHO Expert Consultation, Geneva. Available online: http://www.who.int/foodsafety/publications/micro/march1995/

WHO (World Health Organization), 2009. Assessment of Combined Exposures to Multiple Chemicals, Report of a WHO/IPCS International Workshop on Aggregate/Cumulative Risk Assessment, WHO,Geneva. Available online: http://www.who.int/ipcs/methods/harmonization/areas/workshopreportdocument7.pdf

WHO (World Health Organization), 2010. Characterization and application of physiologically based pharmacokinetic models in risk assessment. Available online: http://www.who.int/ipcs/methods/harmonization/areas/pbpk_models.pdf

6 Supplementary Material

6.1 ABC-MCMC algorithm and related functions written in R language used in the work "Improving dietary exposure models by imputing biomonitoring data through ABC methods"

```
############################################################
# input for function var_conta et calcul_bb
############################################################

# table which records the exposure corrected factors per year
tab_conta_init=matrix(NA,76,2)
colnames(tab_conta_init)=c("annee","coeff")
for (i in 1:76) tab_conta_init[i,1]=1933+i

X_0=0     ##initial body burden
delta_t=7    ## time between two intakes in days
chemin_tableau_initial="xxxxx"        ## working file


############################################################
# algorithm
############################################################

ABC_MCMC_h=function(nsim,delta){
date_1=date()

#-----------------------------------------------------------
# hasting Metropolis variance

var_lambda=1
var_a=0.0001
var_b=3
var_c=3
var_d=3
var_e=0.001



#-----------------------------------------------------------
# initialization

n_d=rep()
theta=list()
theta_1=list()
theta_2=list()
theta_3=list()
theta_4=list()

list_vec_I=list()
list_vec_I_1=list()
list_vec_I_2=list()
list_vec_I_3=list()
list_vec_I_4=list()


## initialize with values that verifies Ip < delta
n_d[1]=3
```

```
theta[[1]]=c(0.099,9.72,11.83,1)

theta_1[[1]]=c(0.1,1)
theta_2[[1]]=c(0.1,17,1)
theta_3[[1]]=c(0.099,9.72,11.83,1)
theta_4[[1]]=c(0.1,9.78,8.99,9.81,1.1)

ntheta=0
ntheta_1=0
ntheta_2=0
ntheta_3=0
ntheta_4=0

j=1
k=1
l=1
m=1

mu_prior_a=0.1
mu_prior_b=10
mu_prior_c=3
mu_prior_d=3
mu_prior_e=1

var_prior_a=0.01
var_prior_b=5
var_prior_c=1.5
var_prior_d=1.5
var_prior_e=0.1

#############################################################################################
# start

for (iter in 2:nsim){

##----------------  start of the first stage (model selection)   ---------------------------##


# new candidate for the model: n_d_prop
lambda=rtruncnorm(1,0,Inf,n_d[iter-1],var_lambda^0.5)
n_d_prop=rpois(1,lambda)


while (n_d_prop==0|n_d_prop>4) {
        lambda=rtruncnorm(1,0,Inf,n_d[iter-1],var_lambda^0.5)
        n_d_prop=rpois(1,lambda)
}

##-----  start of the second stage if m=1 (n_d=1)  ---------##

if(n_d_prop==1){
        j=j+1
        # new candidates for theta
        a_prop=rtruncnorm(1,0,Inf,theta_1[[j-1]][1],var_a^0.5)
        e_prop=rtruncnorm(1,0,Inf,theta_1[[j-1]][2],var_e^0.5)

        vec_theta_prop=c(a_prop,e_prop)

        tab_conta_prop=var_conta_1_compiled(vec_theta_prop,tab_conta_init)

        #simulated body burdens
        vec_bb_prop=calcul_bb_compiled(chemin_tableau_initial,tab_conta_prop, X_0,delta_t,vec_ind)
```

```
		# distances
		vec_l=calcul_lp_compiled(vec_bb_prop,vec_bb_obs_2,vec_classe_2)
		list_vec_l[[iter]]=vec_l
		list_vec_l_1[[j]]=vec_l

		theta[[iter]]=theta[[iter-1]]
		theta_1[[j]]=theta_1[[j-1]]
		n_d[iter]=n_d[iter-1]

		if (length(which(vec_l>delta))==0){
			h=min(1,(prior1(vec_theta_prop[1],mu_prior_a,var_prior_a)/prior1(theta_1[[j-
1]][1],mu_prior_a,var_prior_a))*(prior1(vec_theta_prop[2],mu_prior_e,var_prior_e)/prior1(theta_1[[j-
1]][2],mu_prior_e,var_prior_e)))
			u=runif(1,0,1)
			if (u<h) {
				theta[[iter]]=vec_theta_prop
				theta_1[[j]]=vec_theta_prop
				ntheta_1=ntheta_1+1
				ntheta=ntheta+1
				n_d[iter]=n_d_prop
			}
		}

}else if(n_d_prop==2){

##-----  start of the second stage if m=2 (n_d=2)  ---------##
		k=k+1

		# new candidates for theta
		a_prop=rtruncnorm(1,0,Inf,theta_2[[k-1]][1],var_a^0.5)
		b_prop=rtruncnorm(1,0,Inf,theta_2[[k-1]][2],var_b^0.5)
		e_prop=rtruncnorm(1,0,Inf,theta_2[[k-1]][3],var_e^0.5)

		vec_theta_prop=c(a_prop,b_prop,e_prop)

		tab_conta_prop=var_conta_2_compiled(vec_theta_prop,tab_conta_init)

		#simulated body burdens
		vec_bb_prop=calcul_bb_compiled(chemin_tableau_initial,tab_conta_prop, X_0,delta_t,vec_ind)

		# distances
		vec_l=calcul_lp_compiled(vec_bb_prop,vec_bb_obs_2,vec_classe_2)
		list_vec_l[[iter]]=vec_l
		list_vec_l_2[[k]]=vec_l


		theta[[iter]]=theta[[iter-1]]
		theta_2[[k]]=theta_2[[k-1]]
		n_d[iter]=n_d[iter-1]

		if (length(which(vec_l>delta))==0){
			h=min(1,(prior1(vec_theta_prop[1],mu_prior_a,var_prior_a)/prior1(theta_2[[k-
1]][1],mu_prior_a,var_prior_a))*(prior1(vec_theta_prop[3],mu_prior_e,var_prior_e)/prior1(theta_2[[k-
1]][3],mu_prior_e,var_prior_e))*(prior1(vec_theta_prop[2],mu_prior_b,var_prior_b)/prior1(theta_2[[k-
1]][2],mu_prior_b,var_prior_b)))
			u=runif(1,0,1)
			if (u<h) {
				theta[[iter]]=vec_theta_prop
				theta_2[[k]]=vec_theta_prop
				ntheta_2=ntheta_2+1
				ntheta=ntheta+1
				n_d[iter]=n_d_prop
```

```
                }
        }


}else if(n_d_prop==3){

##-----  start of the second stage if m=3 (n_d=3)  ---------##
        l=l+1

        # new candidates for theta
        a_prop=rtruncnorm(1,0,Inf,theta_3[[l-1]][1],var_a^0.5)
        b_prop=rtruncnorm(1,0,Inf,theta_3[[l-1]][2],var_b^0.5)
        c_prop=rtruncnorm(1,0,Inf,theta_3[[l-1]][3],var_c^0.5)
        e_prop=rtruncnorm(1,0,Inf,theta_3[[l-1]][4],var_e^0.5)

        vec_theta_prop=c(a_prop,b_prop,c_prop,e_prop)

        tab_conta_prop=var_conta_3_compiled(vec_theta_prop,tab_conta_init)

        #simulated body burdens
        vec_bb_prop=calcul_bb_compiled(chemin_tableau_initial,tab_conta_prop, X_0,delta_t,vec_ind)

        # distances
        vec_I=calcul_Ip_compiled(vec_bb_prop,vec_bb_obs_2,vec_classe_2)
        list_vec_I[[iter]]=vec_I
        list_vec_I_3[[l]]=vec_I


        theta[[iter]]=theta[[iter-1]]
        theta_3[[l]]=theta_3[[l-1]]
        n_d[iter]=n_d[iter-1]

        if (length(which(vec_I>delta))==0){
                h=min(1,(prior1(vec_theta_prop[1],mu_prior_a,var_prior_a)/prior1(theta_3[[l-
1]][1],mu_prior_a,var_prior_a))*(prior1(vec_theta_prop[4],mu_prior_e,var_prior_e)/prior1(theta_3[[l-
1]][4],mu_prior_e,var_prior_e))*(prior1(vec_theta_prop[2],mu_prior_b,var_prior_b)/prior1(theta_3[[l-
1]][2],mu_prior_b,var_prior_b))*(prior1(vec_theta_prop[3],mu_prior_c,var_prior_c)/prior1(theta_3[[l-
1]][3],mu_prior_c,var_prior_c)))
                u=runif(1,0,1)
                if (u<h) {
                        theta[[iter]]=vec_theta_prop
                        theta_3[[l]]=vec_theta_prop
                        ntheta_3=ntheta_3+1
                        ntheta=ntheta+1
                        n_d[iter]=n_d_prop
                }
        }


}else if(n_d_prop==4){

##-----  start of the second stage if m=4 (n_d=4)  ---------##
        m=m+1

        # new candidates for theta
        a_prop=rtruncnorm(1,0,Inf,theta_4[[m-1]][1],var_a^0.5)
        b_prop=rtruncnorm(1,0,Inf,theta_4[[m-1]][2],var_b^0.5)
        c_prop=rtruncnorm(1,0,Inf,theta_4[[m-1]][3],var_c^0.5)
        d_prop=rtruncnorm(1,0,Inf,theta_4[[m-1]][4],var_d^0.5)
        e_prop=rtruncnorm(1,0,Inf,theta_4[[m-1]][5],var_e^0.5)

        vec_theta_prop=c(a_prop,b_prop,c_prop,d_prop,e_prop)
```

```
		tab_conta_prop=var_conta_4_compiled(vec_theta_prop,tab_conta_init)

		#simulated body burden
		vec_bb_prop=calcul_bb_compiled(chemin_tableau_initial,tab_conta_prop, X_0,delta_t,vec_ind)

		# distances
		vec_I=calcul_Ip_compiled(vec_bb_prop,vec_bb_obs_2,vec_classe_2)
		list_vec_I[[iter]]=vec_I
		list_vec_I_4[[m]]=vec_I

		theta[[iter]]=theta[[iter-1]]
		theta_4[[m]]=theta_4[[m-1]]
		n_d[iter]=n_d[iter-1]

		if (length(which(vec_I>delta))==0){
			h=min(1,(prior1(vec_theta_prop[1],mu_prior_a,var_prior_a)/prior1(theta_4[[m-
1]][1],mu_prior_a,var_prior_a))*(prior1(vec_theta_prop[5],mu_prior_e,var_prior_e)/prior1(theta_4[[m-
1]][5],mu_prior_e,var_prior_e))*(prior1(vec_theta_prop[2],mu_prior_b,var_prior_b)/prior1(theta_4[[m-
1]][2],mu_prior_b,var_prior_b))*(prior1(vec_theta_prop[3],mu_prior_c,var_prior_c)/prior1(theta_4[[m-
1]][3],mu_prior_c,var_prior_c))*(prior1(vec_theta_prop[4],mu_prior_d,var_prior_d)/prior1(theta_4[[m-
1]][4],mu_prior_d,var_prior_d)))
			u=runif(1,0,1)
			if (u<h) {
				theta[[iter]]=vec_theta_prop
				theta_4[[m]]=vec_theta_prop
				ntheta_4=ntheta_4+1
				ntheta=ntheta+1
				n_d[iter]=n_d_prop
			}
		}

}

res=list(theta,ntheta,n_d,list_vec_I,iter,date_1,date_2)
res_1=list(theta_1,ntheta_1,list_vec_I_1)
res_2=list(theta_2,ntheta_2,list_vec_I_2)
res_3=list(theta_3,ntheta_3,list_vec_I_3)
res_4=list(theta_4,ntheta_4,list_vec_I_4)


save(res,file="res.RData")
save(res_1,file="res_1_suite.RData")
save(res_2,file="res_2_suite.RData")
save(res_3,file="res_3_suite.RData")
save(res_4,file="res_4_suite.RData")

}

# end
##################################################################################################
return(res)
}
```

```
library(truncnorm)  ## load R package "truncnorm" for truncated disctribution
library(splines)    ## load R package "splines"


#---------------------------------------------------------------------
# function to provide the exposure corrected factors per year for the m=1 model

var_conta_1=function(theta,tab_conta_init) {

base=bs(1934:2009, knots=0, degree=1)

a=theta[1]
e=theta[2]

f=a*base[,1]+e*base[,2]
for (i in 1:76) tab_conta_init[i,2]=f[i]

return(tab_conta_init)
}

#---------------------------------------------------------------------
# function to provide the exposure corrected factors per year for the m=2 model

var_conta_2=function(theta,tab_conta_init) {

base=bs(1934:2009, knots=0, degree=2)

a=theta[1]
b=theta[2]
e=theta[3]

f=a*base[,1]+b*base[,2]+e*base[,3]
for (i in 1:76) tab_conta_init[i,2]=f[i]

return(tab_conta_init)
}

#---------------------------------------------------------------------
# function to provide the exposure corrected factors per year for the m=3 model

var_conta_3=function(theta,tab_conta_init) {

base=bs(1934:2009, knots=0, degree=3)

a=theta[1]
b=theta[2]
c=theta[3]
e=theta[4]

f=a*base[,1]+b*base[,2]+c*base[,3]+e*base[,4]
for (i in 1:76) tab_conta_init[i,2]=f[i]

return(tab_conta_init)
}

#---------------------------------------------------------------------
# function to provide the exposure corrected factors per year for the m=4 model

var_conta_4=function(theta,tab_conta_init) {
base=bs(1934:2009, knots=0, degree=4)

a=theta[1]
b=theta[2]
```

```
c=theta[3]
d=theta[4]
e=theta[5]

f=a*base[,1]+b*base[,2]+c*base[,3]+d*base[,4]+e*base[,5]
for (i in 1:76) tab_conta_init[i,2]=f[i]

return(tab_conta_init)
}

#-----------------------------------------------------------------------------------------
###function to calculate the body burden with corrected KDEM

calcul_bb=function(chemin_tableau_initial,tab_conta, X_0,delta_t,vec_ind){

vec_bb_ind=c(NA,length(vec_ind))
r=1

for (ind in vec_ind){
        load(paste(chemin_tableau_initial,ind,".RData",sep=""))
        bb=rep()
        bb[1]=X_0
        for (t in 1:(nrow(tab_ind)-1)){
                bb[t+1]=bb[t]*exp((-
log(2)/tab_ind[t,"HL"])*delta_t)+tab_ind[(t+1),"expo_ext"]*tab_conta[which(tab_conta[,"annee"]==tab_ind[t,"annee"]),"co
eff"]
        }
vec_bb_ind[r]=bb[nrow(tab_ind)]
r=r+1
}
return(vec_bb_ind)
}
#-----------------------------------------------------------------------------------------
###function to calculate the distances

calcul_Ip=function(vec_bb_simu,vec_bb_obs,vec_classe){

vec_I=rep()

for (c in 1:5){
a=vec_bb_simu[which(vec_classe==c)]
b=vec_bb_obs[which(vec_classe==c)]

N=length(a)
## connaitre le rang des valeurs du vecteur a
vec_combined=c(a,b)
vec_combined2=vec_combined[order(vec_combined)]

a=a[order(a)]
b=b[order(b)]

r=rep()
s=rep()
for (i in 1:N){
        r=c(r,which(vec_combined2==a[i]))
        s=c(s,which(vec_combined2==b[i]))
}

somme=0
for (i in round(N/2):N){
        somme=somme+((r[i]-2*i)^2+(s[i]-2*i)^2)
}
```

```
I=4/(N^3)*somme
vec_I[c]=I
}
return(vec_I)
}

#-------------------------------------------------------------------------------------------
## prior1: troncated normal distribution
prior1=function(theta, mu, tau){
prob=dtruncnorm(theta,0,Inf, mean = mu, sd = tau)
return(prob)
}

#-------------------------------------------------------------------------------------------
library(compiler)   ## load R package "compiler" to speed the computation time

var_conta_1_compiled <- cmpfun(var_conta_1)
var_conta_2_compiled <- cmpfun(var_conta_2)
var_conta_3_compiled <- cmpfun(var_conta_3)
var_conta_4_compiled <- cmpfun(var_conta_4)

calcul_bb_compiled <- cmpfun(calcul_bb)
calcul_Ip_compiled <- cmpfun(calcul_Ip)
```

6.2 ABC-SMC algorithm and related functions written in R language used in the work "PBPK and population modeling to interpret urine cadmium concentrations of the French population"

```
w=matrix(NA,T,N)
theta=list()

theta[[1]]=matrix(NA,T,N)
theta[[2]]=matrix(NA,T,N)
theta[[3]]=matrix(NA,T,N)
#theta[[4]]=matrix(NA,T,N)

## ------------------------------------------------------------
## Etape à t0 (t=1)
t=1
i=1
while (i<(N+1)){

### proposal
theta_prop=rep()

        theta_prop[1]=rnorm(1,0.1,0.01)
        theta_prop[2]=rnorm(1,4,2)
        theta_prop[3]=rnorm(1,1,0.1)

        tab_conta_prop=var_conta_2(theta_prop,tab_conta_init)

        #simulatd body burden
        vec_bb_prop=calcul_bb_compiled(delta,tab_conta_prop,vec_age,vec_expo_tab,vec_age_deb,vec_age_fin,vec_U,
vec_BW)

        # comparison with the tolerance
        vec_I=calcul_Ip_compiled(vec_bb_prop,vec_bb_obs,vec_class)

        if (sum(vec_I)<=E[t]){
                w[t,i]=1
                theta[[1]][t,i]=theta_prop[1]
                theta[[2]][t,i]=theta_prop[2]
                theta[[3]][t,i]=theta_prop[3]
                #theta[[4]][t,i]=theta_prop[4]

                i=i+1
        }
}

#------------------------------------------------------------
# Etapes de t=1 (t=2) à t=T (t=T+1)

t=2
while (t<(T+2)){
i=1
while (i<(N+1)){

### proposal
theta_ss=rep()

        theta_ss[1]=sample(theta[[1]][t-1,], 1, prob =w[t-1,])
        theta_prop[1]=rnorm(1,theta_ss[1],sd_K)
```

```
        theta_ss[2]=sample(theta[[2]][t-1,], 1, prob =w[t-1,])
        theta_prop[2]=rnorm(1,theta_ss[2],sd_K2)

        theta_ss[3]=sample(theta[[3]][t-1,], 1, prob =w[t-1,])
        theta_prop[3]=rnorm(1,theta_ss[3],sd_K3)

        tab_conta_prop=var_conta_2(theta_prop,tab_conta_init)

        #simulated body burden
        vec_bb_prop=calcul_bb_compiled(delta,tab_conta_prop,vec_age,vec_expo_tab,vec_age_deb,vec_age_fin,vec_U,
vec_BW)

        # comparison with the tolerance
        vec_I=calcul_Ip_compiled(vec_bb_prop,vec_bb_obs,vec_class)

        if (sum(vec_I)<=E[t]){
                somme1=rep()
                somme2=rep()
                somme3=rep()
                #somme4=rep()

                for (j in 1:N){
                        somme1[j]=w[t-1,j]*dnorm(theta_prop[1],theta[[1]][t-1,j],sd_K)
                        somme2[j]=w[t-1,j]*dnorm(theta_prop[2],theta[[2]][t-1,j],sd_K2)
                        somme3[j]=w[t-1,j]*dnorm(theta_prop[3],theta[[3]][t-1,j],sd_K3)
                        #somme4[j]=w[t-1,j]*dnorm(theta_prop[4],theta[[4]][t-1,j],sd_K4)

                }

        w[t,i]=(dnorm(theta_prop[1],0.1,0.01)/sum(somme1))*(dnorm(theta_prop[2],4,2)/sum(somme2))*(dnorm(theta_
prop[3],1,0.1)/sum(somme1))#*(dgamma(theta_prop[4],0.01,0.01)/sum(somme2))
                theta[[1]][t,i]=theta_prop[1]
                theta[[2]][t,i]=theta_prop[2]
                theta[[3]][t,i]=theta_prop[3]
                #theta[[4]][t,i]=theta_prop[4]


                i=i+1
                print(i)
        }
}
## normalize the weigths
for(i in 1:N) w[t,i]=w[t,i]/sum(w[t,])

t=t+1
save(theta,file="theta_lambda_m2.RData")

}
```

```
#-----------------------------------------------------------------------
# table of exposure variation in time for model 1

var_conta_2=function(theta,tab_conta_init) {

base=bs(1900:2009, knots=4, degree=2)

a=theta[1]
b=theta[2]
e=theta[3]

f=a*base[,1]+b*base[,2]+e*base[,3]
for (i in 1:110) tab_conta_init[i,2]=f[i]

return(tab_conta_init)
}

#-----------------------------------------------------------------------
# table of exposure variation in time for model 1

var_conta_3=function(theta,tab_conta_init) {

base=bs(1900:2009, knots=4, degree=3)

a=theta[1]
b=theta[2]
c=theta[3]
e=theta[4]

f=a*base[,1]+b*base[,2]+c*base[,3]+e*base[,4]
for (i in 1:110) tab_conta_init[i,2]=f[i]

return(tab_conta_init)
}

#-----------------------------------------------------------------------
## pbpk model


pbpk=function(delta,tab_conta,TSTOP,EXPOAIR,STARTEXPO,ENDEXPO,G_init,BW,SEXE){

######                    variables du modèles PBPK          ###############

 C1 = 0.1 #fraction inhal depos muco ciliaire - 0.1 cigarette smoke, 0.7 factory dust
 C2 = 0.4 #fraction inhal depos poumons - 0.4 cigarette smoke, 0.13 factory dust
 C3 = 0.05*delta #taux absorption pulmonaire (/day)
 C4 = 0.005*delta #taux transport muco ciliaire vers TGI (/day)

 #fraction deposition TGI
if (SEXE==1) C5 = 0.05 else if (SEXE==2)  C5 = 0.075

 C6 = 0.05*delta  #taux d'absorption TGI (/day)
 C7 = 0.25 #fraction distribution vers MT
 C8 = 1 # quantite MT maximum (ug)

 C9 = 0.44 #debit plasma/tissus (/day)
 C10 = 0.00014*delta #relargage tissus/plasma (/day)
 C11 = 0.27 #debit tissus/excretion fecale (/day)
 C12 = 0.25 #debit plasma/foie  (/day)
 C13 = 0.00003*delta #relargage foie/plasma (/day)
 C14 = 0.00016*delta #debit MT/foie (/day)
 C15 = 0.00005*delta #debit foie/excretion fecale (/day)
```

```
C16 = 0.012*delta #relargage MT/eryth (/day)
C17 = 0.95#debit MT/excretion urinaire (/day) - decreases from age 30 to 80 by 33%      pour ATSDR C17= 2.99  et C17b = 0.05
C18 = 0.00001*delta #retour rein/plasma (/day)
C19 = 0.00014*delta #debit rein/excretion urinaire (/day) - increases from age 30 with C21 each year

FREIN= round(10950/delta) # age de modification renale 35 ans (day)

CX = 0.04 #debit plasma/erythrocytes
C20 = 0.1 #fraction de cadmium plasma + eryth
C21 = 0.0000011 *delta#taux augmentation excretion urinaire apres 30 ans (/day)

CORTX = 1.51 #Ratio of kidney/cortex concentration

ALW = 1500 #Average liver weight (g)
AKW = 281 #Average kidney weight (g)
ABV = 0.07 #Average blood volume (L/kg)
ATW = 67467 #Weight of other tissues (g)
ABSG = 1.06 #Average blood specific gravity (g/ml)

ADUA = 1.4 *delta#Average daily crat excretion adult(g/day)
ADUO = 0.9 *delta#Average daily urine excretion old (g/day)       # voir comment l'intégrer dans une boucle
ADUC = 0.5*delta #Average daily urine excretion child(g/day)

######                    initialisation des vecteurs          ###############

vec_C17=rep(0,TSTOP) # fonction renale
vec_AE1=rep(0,TSTOP) #lung
vec_AE2=rep(0,TSTOP) #intestine
vec_I=rep(0,TSTOP)   #daily absorbtion
vec_I2=rep(0,TSTOP)
vec_I1=rep(0,TSTOP)
vec_AL=rep(0,TSTOP) #liver
vec_AT=rep(0,TSTOP) #other tissue
vec_AK=rep(0,TSTOP) #kidney
vec_AB1=rep(0,TSTOP) #BC 1 plasma
vec_AB2=rep(0,TSTOP) #BC 2 erythrocytes
vec_AB3=rep(0,TSTOP) #BC 3 methalothionnein
vec_CB=rep(0,TSTOP) #conc in whole blood (ug/L)
vec_AF=rep(0,TSTOP) #feces excretion
vec_AU=rep(0,TSTOP) #urinary excretion
vec_CU=rep(0,TSTOP) #conc in daily urine  (ug/g cr)

age=TSTOP/round(365/delta)
naissance=2009-age

######                    modèle pbpk                    ###############
for (t in 2:TSTOP){
annee=naissance+trunc(t/round(365/delta))

G=G_init*tab_conta[which(tab_conta[,"annee"]==annee),"coeff"]

 ####  modif fonction renale
 if (t<=FREIN) vec_C17[t] = C17
 if (t>FREIN)  vec_C17[t] = C17*(1.198 - t*(1.8*10^(-05)))
 if (t<FREIN)  n = 0
 if (t>=FREIN) n = (t - FREIN)/365

 if (t<=round(7300/delta)) ADU = ADUC
 if (t<round(21900/delta)) ADU = ADUA
 if (t>=round(21900/delta))ADU = ADUO
 #if (CKBW>0) URCO=Ut/CKBW      #ratio urinary Cd/cortex Cd  XXXXXXXXX a quoi sert cette condition?
```

```
#### modif durée du tabagisme
if(t<STARTEXPO) EXPOAIR=0
if(t>ENDEXPO) EXPOAIR=0

#### LUNG

E1 = C2*EXPOAIR #rate exposing the lung (ug/day) XXXXXXXXXX J AI MODIFIE
E1t = E1 -C3*vec_AE1[t-1] -C4*vec_AE1[t-1]  #rate of loading (ug/day)
vec_AE1[t]=vec_AE1[t-1]+E1t   #amount remaining in the lung (ug) at time t

#### INTESTINE

E2 = C5*(G + C1*EXPOAIR +C4*vec_AE1[t])
E2t=E2-C6*vec_AE2[t-1]    #loading GI tract (ug/day)  XXXXXXXXXX J AI MODIFIE
vec_AE2[t]=vec_AE2[t-1]+ E2t       #amount remaining in the GI (ug)  at time t

#### DAILY ABSORBED AMOUNTS OF Cd

vec_I[t] = C3*vec_AE1[t] +C6*vec_AE2[t] #rate of internal uptake (ug/day)
vec_I2[t] = C7*vec_I[t]         #amount of daily intake to MT (ug)   at time t
vec_I1[t] = vec_I[t] - vec_I2[t]       #rate unloading to plasma (ug/day)

#### BLOOD COMPARTMENT 1 PLASMA

B1=vec_I1[t] +C10*vec_AT[t-1] +C13*vec_AL[t-1]+C18*vec_AK[t-1] #rate loading plasma (ug/day)
B1t= B1-vec_AB1[t-1]*(C9+C11+C12+CX)   #rate of change (ug/day)
vec_AB1[t]= vec_AB1[t-1]+ B1t                #amount in plasma (ug)

#### Blood Compartment 2 including ERYTHROCYTES

B2 = vec_AB1[t]*CX # rate loading blood 2 (ug/day)
B2t = B2-C16*vec_AB2[t-1]       #rate of change in blood 2 (ug/day)
vec_AB2[t]=vec_AB2[t-1]+ B2t     #amount in blood 2 (ug)

#### BLOOD COMPARTMENT 3 METALLOTHIONEIN

B3 = vec_I2[t]+C16*vec_AB2[t]+C14*vec_AL[t-1] #rate loading MT (ug/day)
B3t= B3-(vec_C17[t]*vec_AB3[t-1]+(1-vec_C17[t])*vec_AB3[t-1])
vec_AB3[t]=vec_AB3[t-1]+B3t                    #amount bound to MT (ug)

#### AMOUNT IN WHOLE BLOOD (ug)
AB4t = vec_AB1[t]+vec_AB3[t]+vec_AB2[t]

vec_CB[t] = AB4t/(ABV*BW) #Conc in whole blood (ug/L)

#### LIVER

L = C12*vec_AB1[t] #rate loading liver (ug/day)
Lt = L -vec_AL[t-1]*(C13+C14+C15) #rate of change in liver (ug/day)
vec_AL[t] = vec_AL[t-1]+Lt      #amount in liver (ug)

LBW= vec_AL[t]/ALW #concentration in liver (ug/g)

#### KIDNEY

C19t= C19 + C21*n
K = vec_C17[t]*vec_AB3[t] #rate loading kidney (ug/day)
Kt = K - vec_AK[t-1]*(C18+C19t) #rate of change in kidney (ug/day)
vec_AK[t]=vec_AK[t-1]+Kt       # amount in kidney (ug)

KBW= vec_AK[t]/AKW #concentration in kidney (ug/g)
CKBW= KBW*CORTX #concentration in kidney cortex (ug/g)
```

```
#### OTHER TISSUES

OT = C9*vec_AB1[t] #rate loading other tissues (ug/day)
Tt = OT- vec_AT[t-1]*C10 #rate of change in other tissues (ug/day)
vec_AT[t]=vec_AT[t-1]+ Tt  # amount in other tissue (ug)
TBW= vec_AT[t]/ATW #concentration in other tissue (ug/g)

#### FECES EXCRETION

Ft = C11*vec_AB1[t] + C15*vec_AL[t] #rate of excretion (ug/day)
vec_AF[t]=vec_AF[t-1]+Ft #amount excreted (ug)

#### URINARY EXCRETION

Ut = vec_AB3[t]*(1-vec_C17[t]) +vec_AK[t]*C19t #rate of urinary excret (ug/day)
vec_AU[t] =vec_AU[t-1]+ Ut      #cumulated amount in urine (ug)
vec_CU[t]=Ut/ADUA               # conc. in daily urine (ug/g cr)
#print(t)
}

return(vec_CU[TSTOP])
}

#-------------------------------------------------------------------------------
# calcul of body burden
calcul_bb=function(delta,tab_conta,vec_age,expo_tab,vec_age_deb,vec_age_fin,vec_G,vec_BW,vec_SEXE){
nb=round(365/delta)
vec_bb_ind=rep()
for (i in 1:length(vec_age)){
vec_bb_ind[i]=pbpk(delta,tab_conta,TSTOP=vec_age[i]*nb,EXPOAIR=expo_tab*delta,STARTEXPO=vec_age_deb[i]*nb,ENDE
XPO=vec_age_fin[i]*nb,G=vec_G[i]*delta,BW=vec_BW[i],SEXE=vec_SEXE[i])
}
return(vec_bb_ind)
}

#-------------------------------------------------------------------------------------------
library(compiler)
calcul_bb_compiled <- cmpfun(calcul_bb)

#-------------------------------------------------------------------------------
# calcul_lp
calcul_lp=function(vec_bb_simu,vec_bb_obs,vec_classe){

vec_l=rep()

## non fumeur
for (c in 1:nlevels(vec_class)){
a=vec_bb_simu[which(vec_classe==c&vec_fum==0)]
b=vec_bb_obs[which(vec_classe==c&vec_fum==0)]

N=length(a)
## connaitre le rang des valeurs du vecteur a
vec_combined=c(a,b)
vec_combined2=vec_combined[order(vec_combined)]

a=a[order(a)]
b=b[order(b)]


r=rep()
s=rep()
for (i in 1:N){
        r=c(r,which(vec_combined2==a[i]))
```

```
        s=c(s,which(vec_combined2==b[i]))
}


somme=0
for (i in round(N/2):N){
        somme=somme+((r[i]-2*i)^2+(s[i]-2*i)^2)
}

I=4/(N^3)*somme

vec_I[c]=I
}

### smokers
for (c in 1:nlevels(vec_class)){
a=vec_bb_simu[which(vec_classe==c&vec_fum==1)]
b=vec_bb_obs[which(vec_classe==c&vec_fum==1)]

N=length(a)
## connaitre le rang des valeurs du vecteur a
vec_combined=c(a,b)
vec_combined2=vec_combined[order(vec_combined)]

a=a[order(a)]
b=b[order(b)]


r=rep()
s=rep()
for (i in 1:N){
        r=c(r,which(vec_combined2==a[i]))
        s=c(s,which(vec_combined2==b[i]))
}


somme=0
for (i in round(N/2):N){
        somme=somme+((r[i]-2*i)^2+(s[i]-2*i)^2)
}

I=4/(N^3)*somme

vec_I[c]=I
}

return(vec_I)
}
```

6.3 Supplementary material related to the work "Toward an integrative approach for cumulative risk assessment of persistent chemicals: a case study on dioxins, furans and dioxin-like PCBs in France"

BMD analysis of NTP data (2006)

The chronic p.o. 2-year rat carcinogenicity assay as performed by the National Toxicology Program[1,2] provided a source for dose-response information for hepatic and thyroid toxicity. Similarly (in particular) reproductive toxicity which was used as the starting point for the derivation of the TDI of 2,3,7,8-TCDD[3,4] (JECFA, 2002; 2005) was selected to represent 2,3,7,8-TCDD's most sensitive toxic effects.

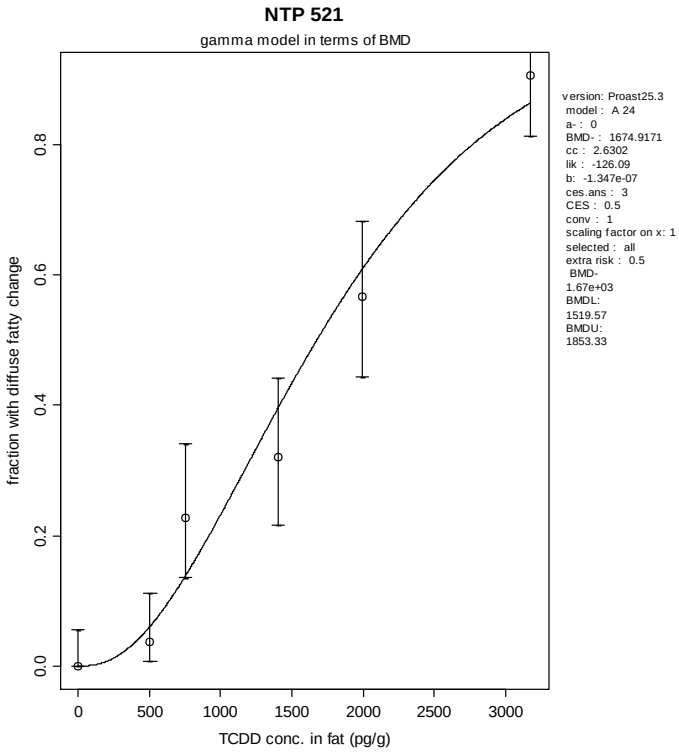
Regarding hepatic toxicity NTP reports various pathological effects. Effects with a high (>0.25) background incidence were disregarded. By fitting a dose-response model the dose may be obtained at which the average animal is affected by the specific effect. In Table S-1 the approximate order of magnitude at which the various hepatic effects occur are given. Table-S1 illustrates the increasing severity of hepatic toxicity at increasing hepatic accumulation of TCDD. In this context pigmentation, toxic hepatopathy, multinucleated hepatocytes, hepatocyte hypertrophy, hyperplasia and diffuse fatty change of liver cells are considered to reflect reversible, non-cancer, toxicity. Diffuse fatty change was assumed to reflect the liver concentration at the transition from reversible, non-cancer toxicity, to irreversible, non-cancer toxicity (fibrosis/necrosis) and pre-cancer toxicity (nodular hyperplasia, pre-neoplastic nodules).

Supplemental Material, table 1: Liver concentration at which the average rat is affected by specific hepatoxicity.

Effect	Internal concentration liver (pg/g liver (ww))
Pigmentation	2100
Toxic hepatopathy	4200
Multinucleated hepatocyte	4700
Hepatocyte hypertrophy	3100
Oval cell hyperplasia	5000

Diffuse fatty change	5600
Nodular hyperplasia	8400
Cholangiofibrosis	9000

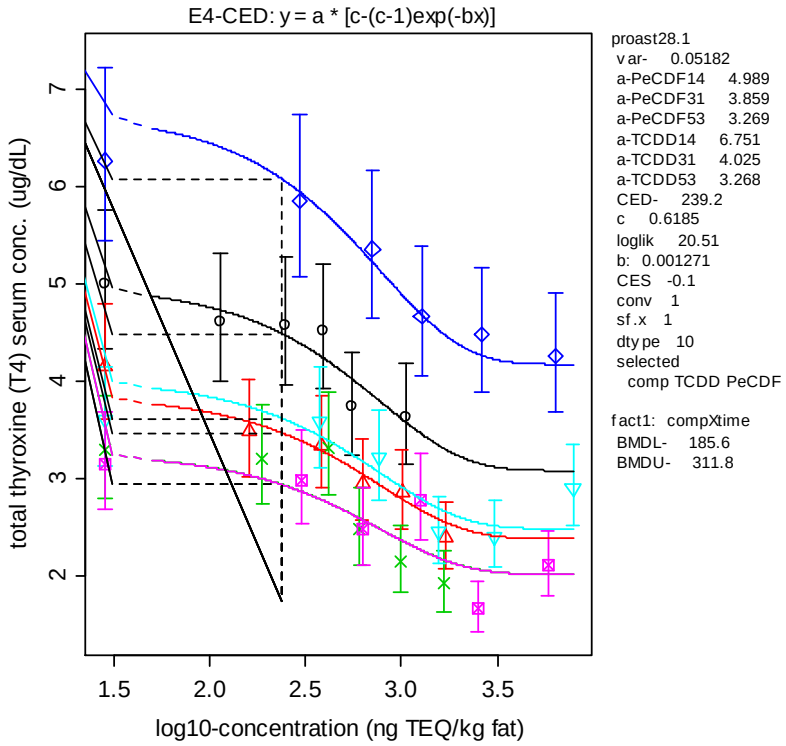
Figure 1 shows the dose-response curve for diffuse fatty change in the liver of the exposed rats.



Supplemental Material, figure 1: The dose-response curve of diffuse fatty change in the liver of female rats which have been exposed for 2 years to 2,3,7,8-TCDD. The TCDD concentration is given in pg/g adipose tissue fat. In this context the adipose tissue concentration was preferred over the liver concentration because of its easy conversion to the Body Burden ("Whole Body Concentration"). Modelling software: PROAST www.PROAST.nl, gamma model for dichotomous data.

Next to hepatic toxicity thyroid toxicity was reported as a decrease in Total Thyroid hormone (TT4) levels. These data are available for TCDD and PeCDF. The concentrations were transformed into TEQs[5] to enable simultaneous analysis.

In Figure S-2 the dose-responses are shown of TT4 against the adipose tissue TEQ concentration. The dose-responses have different background concentrations depending on the (TCDD and PeCDF) study and the week of sampling (week 14, 31 or 53). The dose-responses are parallel indicating that the female rats are equally sensitive after different exposure durations and for exposure to TCDD or PeCDF.



Supplemental Material, figure 2: The dose-response curves of thyroid toxicity as revealed by a lowering of the TT4 level in rats which have been exposed for 2 years to 2,3,7,8-TCDD or PeCDF. Curves represent TCDD, wk 14; PeCDF, wk 14; TCDD, wk 31; PeCDF, wk 31 downwards. Curves of week 53 of both compounds are on top of each other (bottom curve). Modelling software: www.PROAST.nl, exponential model family, model 4 for continuous data).

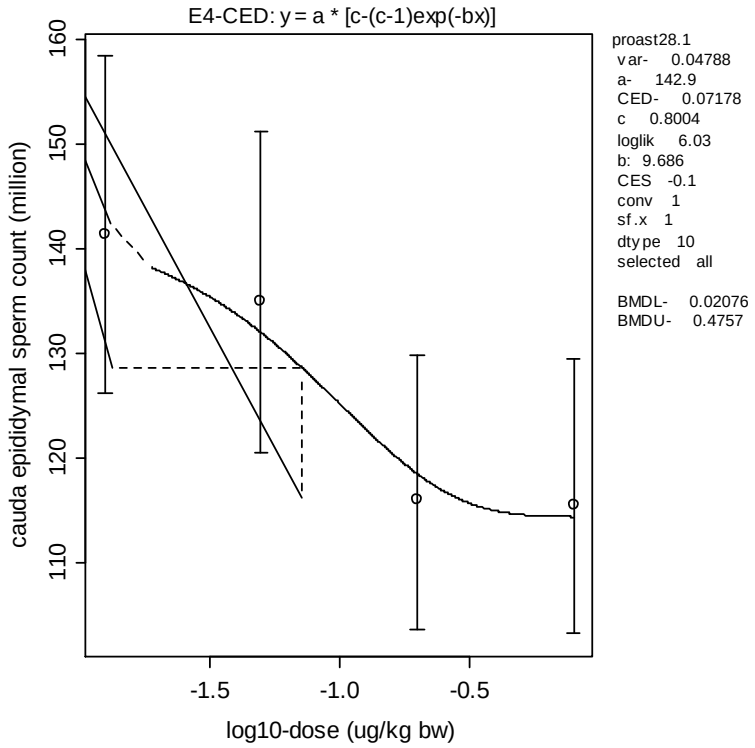
Analysis of reproductive toxicity

Table 2 presents the studies in which the lowest NOAELs and LOAELs were identified for the most sensitive adverse effects of TCDD, i.e. reproductive endpoints in rats. In the studies by Gehrs, Ohsako, and Mably[6,7,8] no data were reported that could be reanalyzed, i.e. no individual or summary data was reported. In the study of Faqi[9] sufficient (summary) data were reported to perform dose-response modelling. However, the quality of the data is poor, the lowest dose tested already results in the maximum response for all the related endpoints. Several sperm related endpoints of Gray[10] exhibited the same flaws as described for Faqi's data. However, one single sperm endpoint supplied

informative dose-response data, namely affected sperm count in the cauda epididimis. For this reason this endpoint was further analyzed (see Figure S-3).

Supplemental Material, table 2: Reproductive toxicity as revealed by single exposure to 2,3,7,8-TCDD

Effect	NOAEL/LOAEL (ng/kg bw)	Species	Reference
Immune suppression offspring	50 (LOAEL)	Rat	(Gehrs *et al.* 1997; Gehrs and Smialowicz 1999)
Reduced entral prostate weight; decreased anogenital distance in male offspring	13 (NOAEL) 51 (LOAEL	Rat	(Ohsako *et al.* 2001)
Decreased sperm count in adult offspring	28 (LOAEL)	Rat	(Mably *et al.* 1992)
Accelerated eye opening and decreased sperm count in offspring	51 (LOAEL)	Rat	(Gray *et al.* 1997)
Decreased sperm production and altered sexual behaviour in male offspring	25 (LOAEL)	Rat	(Faqi *et al.* 1998)



Supplemental Material, figure 3: The dose-response curve of decreased sperm production in offspring of dams which have been exposed to a single dose of 2,3,7,8-TCDD at gestational day 15. Data obtained from Gray et al. (1997). Modelling software: www.PROAST.nl, exponential model family, model 4 for continuous data).

References

1 NTP (2006a). Toxicology and Carcinogenesis Studies of 2,3,4,7,8-pentachlorodibenzofuran (PeCDF) (CAS No. 57117-31-4) in Female Harlan Sprague-Dawley Rats (Gavage Studies), TR 525, Research Triangle Park, NC, USA.

2 NTP (2006b). Toxicology and Carcinogenesis Studies of 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) (CAS No. 1746-01-6) in Female Harlan Sprague-Dawley Rats (Gavage Studies), TR 521, Research Triangle Park, NC, USA.

3 JECFA (2002). Joint FAO/WHO Expert Committee on Food Additives (JECFA), Safety evaluation of certain food additives and contaminants, WHO Food Additives Series 48, WHO, Geneva, 2002.

4 JECFA (2005). Joint FAO/WHO Expert Committee on Food Additives (JECFA), Summary and Conclusions of the sixty-fourth meeting, Rome, 2005.

5 Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M et al. 2006. The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds. Toxicol. Sci. 93(2): 223-241.

6 Gehrs, B. C., Riddle, M. M., Williams, W. C., and Smialowicz, R. J. (1997). Alterations in the developing immune system of the F344 rat after perinatal exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin: II. Effects on the pup and the adult. *Toxicology* **122**, 229-40.

7 Mably, T. A., Bjerke, D. L., Moore, R. W., Gendron-Fitzpatrick, A., and Peterson, R. E. (1992). In utero and lactational exposure of male rats to 2,3,7,8-tetrachlorodibenzo-p-dioxin. 3. Effects on spermatogenesis and reproductive capability. *Toxicol Appl Pharmacol* **114**, 118-26.

8 Ohsako, S., Miyabara, Y., Nishimura, N., Kurosawa, S., Sakaue, M., Ishimura, R., Sato, M., Takeda, K., Aoki, Y., Sone, H., Tohyama, C., and Yonemoto, J. (2001). Maternal exposure to a low dose of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) suppressed the development of reproductive organs of male rats: dose-dependent increase of mRNA levels of 5alpha-reductase type 2 in contrast to decrease of androgen receptor in the pubertal ventral prostate. *Toxicol Sci* **60**, 132-43.

9 Faqi, A. S., Dalsenter, P. R., Merker, H. J., and Chahoud, I. (1998). Reproductive toxicity and tissue concentrations of low doses of 2,3,7,8-tetrachlorodibenzo-p-dioxin in male offspring rats exposed throughout pregnancy and lactation. *Toxicol Appl Pharmacol* **150**, 383-92.

10 Gray, L. E., Ostby, J. S., and Kelce, W. R. (1997). A dose-response analysis of the reproductive effects of a single gestational dose of 2,3,7,8-tetrachlorodibenzo-p-dioxin in male Long Evans Hooded rat offspring. *Toxicol Appl Pharmacol* **146**, 11-20.

6.4 Bayesian hierarchic model for meta-analysis in JAGS language

```
model
{
for(k in 1:Nst) # k=study
{
            delta.st[k] ~ dnorm(0,tau.st)
}
for(l in 1:Nlines) # l=lines
{
              lgm[l] ~ dnorm(mu[l],tau[l])
              mu[l] <- mu0[sub[l]] + delta.st[s[l]]
             tau[l] <- n[l]*tau.0[sub[l]]


            # individual level
             lv[l] ~ dgamma(a[l],b[l])

              a[l] <- (n[l]-1)/2
              b[l] <- n[l]*(tau.0[sub[l]])/2
  }

   for(j in 1:Nsub) # j=substrate
   {
              mu0[j] ~ dnorm(mu.sub,tau.sub)
            tau.0[j] ~ dnorm(mu.tau.0,tau.tau.0)
   }



# priors
   mu.sub ~ dnorm(0,0.0001)
   tau.sub ~ dgamma(0.01,0.01)
   mu.tau.0~ dnorm(0,0.0001)
   tau.tau.0 ~ dgamma(0.01,0.01)
   tau.st ~ dgamma(0.01,0.01)
}
```

Résumé du manuscrit

Introduction

A travers leur alimentation et leur environnement, les individus sont exposés à un large éventail de produits chimiques naturels et synthétiques. Par conséquent, un certain nombre de questions doivent être abordées pour déterminer si une telle exposition pourrait présenter des effets néfastes sur la santé, comment les identifier, les caractériser et prendre des mesures pour réduire ces risques. Au cours des dernières décennies, de nombreuses organisations ont vu le jour à travers le monde, à différentes échelles géographiques et/ou politiques, afin de répondre à ces questions (par exemple, l'Agence américaine de protection de l'environnement: US-EPA, l'Autorité européenne de sécurité des aliments: L'EFSA, l'Agence française pour la sécurité sanitaire: l'ANSES). Elles visent à effectuer une analyse des risques afin d'assurer la sécurité sanitaire des aliments et/ou de l'environnement et à protéger la santé publique, mais aussi dans un contexte de commerce marqué par la mondialisation, de se conformer aux normes internationales et à la réglementation du marché. Dans ce contexte, un cadre formel définissant l'évaluation du risque a été proposé par la FAO et l'OMS dans le cadre du Codex Alimentarius créé en 1963.

Dans ce cadre formel, l'analyse des risques est définie comme «un processus comportant trois volets: la gestion des risques, l'évaluation des risques, et la communication des risques» (OMS, 1995). Traditionnellement, le processus d'évaluation des risques a ensuite été divisé en quatre étapes successives: l'identification des dangers, leur caractérisation, l'évaluation de l'exposition et caractérisation des risques. Cependant aujourd'hui, le développement de nouvelles méthodes telles que le mode d'action (MoA) et les modèles doses-réponses basés sur la physiologie (BBDR) offrent la possibilité d'évoluer vers une approche plus intégrée de l'évaluation des risques. En effet, même si les quatre étapes de l'évaluation des risques sont toujours nécessaires, elles peuvent désormais être regroupées dans un continuum entre l'exposition et la réponse. Considérer un tel continuum permet notamment une compréhension plus mécanistique de l'effet d'un produit chimique sur la santé humaine. Cela permet également de combiner plusieurs types de données qui sont souvent considérés de façon indépendante dans l'approche traditionnelle ainsi que de prendre en compte la variabilité et l'incertitude de façon homogène tout au long du continuum.

Ce travail de thèse s'inscrit dans ce contexte de continuum exposition-réponse. L'objectif de ce travail est de proposer des méthodes qui permettent de relier l'exposition externe à l'exposition interne en intégrant la cinétique de la substance ainsi que la variabilité intra et inter individuelle. Cette méthode vise à utiliser toutes les données disponibles pour l'évaluation de l'exposition afin d'améliorer l'évaluation des risques liés aux expositions à plusieurs substances et/ou par différentes voies d'exposition.

Présentation des principaux concepts et améliorations à apporter

Exposition externe

L'exposition externe fait référence à une mesure de la concentration des produits chimiques de l'environnement qui entre en contact avec le corps (Lyons et al., 2008). Dans la majorité des cas, des modèles d'exposition combinant les données sur la concentration de la substance chimique mesurée dans les médias (sol, eau, nourriture ...) et des données sur le comportement individuel (par exemple, la consommation alimentaire, le temps passé dans un milieu intérieur, extérieur ...) sont utilisés pour estimer l'exposition externe. Cependant, l'évaluation de l'exposition à travers l'exposition externe présente des limites. En effet, tout d'abord, puisque l'évaluation de l'exposition externe n'intègre pas la notion d'absorption, elle ne représente donc pas de façon réaliste la dose à laquelle l'individu est vraiment exposé depuis son environnement. Ainsi faire l'hypothèse d'un taux d'absorption de 100% pour chaque voie d'exposition peut conduire à une forte surestimation de l'exposition. Ensuite, l'exposition externe ne tient pas compte de la cinétique de la substance chimique dans l'organisme. Par conséquent, la potentielle accumulation de la substance dans l'organisme due à une succession d'expositions externes n'est pas considérée. En outre, pour les produits chimiques avec des demi-vies très longues, les expositions externes ayant eu lieu les années précédentes participent encore à la charge corporelle actuelle de l'individu. Ainsi, même si elle est rarement connue, l'évolution antérieure de l'exposition est importante pour l'évaluation des risques à ces substances qualifiées de persistantes.

Pour être en mesure de prendre en compte ces variations d'exposition dans le temps, il est nécessaire de passer d'une modélisation statique de l'exposition à une modélisation dynamique. Bertail et al. (2008) ont proposé une modélisation dynamique de l'exposition qui prend précisément en compte l'accumulation à travers le modèle KDEM (Kinetic Dietary Exposure Model). Cependant, bien que proposant une approche dynamique, ce modèle présente certaines limites. Tout d'abord, le modèle a été proposé uniquement à l'échelle de la population et ne prend donc pas en compte la variabilité inter et intra individus. Deuxièmement, la modélisation proposée ne permet pas de considérer une potentielle évolution dans le temps de l'exposition externe. Troisièmement, la toxicocinétique est décrite à partir d'un modèle de premier ordre à un compartiment. Un tel modèle peut être considéré comme suffisant pour décrire la cinétique de certaines substances (par exemple les substances qui s'accumulent principalement dans les lipides comme les dioxines ou le méthylmercure), mais pour d'autres substances présentant des cinétiques plus complexes, il n'est pas adapté. Enfin, le modèle a été construit pour l'exposition alimentaire uniquement, et ne considère qu'un seul produit chimique. Il a donc besoin d'être modifié pour être appliqué à des expositions agrégées et/ou combinées. Enfin, les charges corporelles estimées à partir du modèle KDEM n'ont jamais été comparées à des charges corporelles mesurées par les données de biosurveillance. Or une telle comparaison permettrait de valider ou d'informer le modèle.

Suite à l'observation de ces limites, l'un des objectifs de ce travail est de proposer une modélisation dynamique de l'exposition qui intègrerait la cinétique de la substance, la variabilité intra et inter individus, l'évolution potentielle de l'exposition tout au long de la vie de l'individu, ainsi que les informations provenant des études de biosurveillance.

Données de biosurveillance, une mesure de l'exposition interne

La biosurveillance est définie comme une méthode pour évaluer l'exposition humaine aux substances chimiques en mesurant les produits chimiques ou leurs métabolites dans les tissus ou des échantillons humains, tels que le sang ou l'urine (NRC, 2006). Le terme données de biosurveillance se réfère ainsi à des mesures analytiques de biomarqueurs dans ces tissus (Albertini et al., 2006) où les biomarqueurs peuvent être n'importe quelle substance, structure ou processus qui indique une exposition ou une prédisposition à une maladie. En conséquence, deux types de données de biosurveillance peuvent être définies: les biomarqueurs de l'exposition interne (par exemple mesurée en pg/g de lipides dans le sang ou en ug /g de créatinine dans les urines) et les biomarqueurs d'effet biologique (par exemple d'inhibition d'AChE). Cette partie sur l'exposition se concentre sur les biomarqueurs de l'exposition interne, leur interprétation et leurs avantages.

Les données de biosurveillance peuvent être interprétées en terme de risque soit directement en utilisant une valeur critique interne lorsqu'elle est disponible, soit en dérivant une valeur critique interne à partir d'une valeur externe et d'un modèle cinétique. La principale faiblesse des données de biosurveillance réside donc davantage dans l'incapacité à identifier les différentes sources d'exposition, leur contribution relative à l'exposition totale ou à interpréter les niveaux d'exposition observés en termes de comportements individuels. Afin de pouvoir les interpréter, il est donc nécessaire de relier ces données de biosurveillance à des niveaux d'exposition externe. Certaines approches de dosimétrie inverse ont déjà été proposées dans le but d'estimer les niveaux d'exposition externe qui pourraient expliquer une exposition interne mesurée. Cependant, ces approches se basant uniquement sur les données d'exposition interne n'intègrent pas d'information sur l'exposition externe précédemment évaluée. En effet, le plus souvent, les expositions externes et internes sont considérées séparément bien qu'elles représentent en fait un seul événement le long du continuum exposition-réponse. De plus, la plupart du temps, ces approches de dosimétrie inverse se limitent à une seule source d'exposition.

Par conséquent, l'un des objectifs de ce travail est de proposer une méthode qui permette de combiner toutes les informations disponibles pour évaluer l'exposition et de relier l'exposition externe et interne. Souvent, les données de biosurveillance sont disponibles à l'échelle individuelle et sont accompagnées d'informations sur le comportement de l'individu (par exemple, les habitudes de consommation, le statut tabagique) et des mesures anthropométriques (par exemple le poids). Ces données peuvent permettre d'évaluer une exposition externe individuelle en les combinant avec des données de contamination. Un autre objectif de ce travail est donc d'étudier dans quelle mesure il est possible de modéliser le lien entre l'exposition externe et interne au niveau individuel lorsque l'on intègre ces informations.

Modèles toxicocinétiques

Les modèles toxicocinétiques sont définis comme «des descriptions mathématiques simulant la relation entre le niveau d'exposition externe et la concentration de produits chimiques dans des matrices biologiques au cours du temps» (OMS, 2010). Ces modèles prennent en compte l'absorption, la distribution dans le corps, le métabolisme et l'excrétion (ADME) du produit chimique et de ses métabolites. De nombreux modèles sont maintenant disponibles et varient selon leur degré de complexité. Ils peuvent être classés en deux catégories principales: les modèles toxicocinétiques dits « data-based » ou modèles classiques et les modèles toxicocinétiques dits « physiologically-based » (PBTK) (US EPA, 2014).

La toxicocinétique étant le lien qui permet de relier l'exposition externe à l'exposition interne, dans ce sens (dosimétrie), ou dans l'autre (dosimétrie inverse), c'est donc un outil essentiel pour considérer l'exposition comme un continuum et pour utiliser toutes les informations et les données disponibles. Les modèles les plus simples (par exemple les modèles à un compartiment) ont l'avantage de dépendre de peu de paramètres et donc d'être plus faciles à développer et à implémenter. Les plus complexes (par exemple, les modèles PBTK avec de nombreux compartiments) proposent une description plus précise des mécanismes qui expliquent les différences entre les individus, les espèces et les produits chimiques et permettent donc une meilleure intégration de la variabilité inter et intra individuelle. Ils permettent également de se concentrer sur un organe ou un tissu spécifique qui présente un intérêt particulier pour l'évaluation des risques (par exemple les reins pour le cadmium) ainsi que de combiner plusieurs voies d'exposition.

Grâce aux modèles toxicocinétiques (TK) il est possible d'intégrer la cinétique de la substance et de lier l'exposition externe à la dose interne. Toutefois, l'intégration de la cinétique nécessite de passer d'une modélisation classique et statique de l'exposition externe à une modélisation dynamique i.e qui évolue dans le temps. En effet, la dose interne dépend de la succession d'expositions ou de la durée d'exposition et non pas d'une exposition isolée, en particulier pour les produits chimiques ayant une longue demi-vie. Par conséquent, les modèles (TK) doivent être utilisés avec des scénarii d'exposition sur une période de temps qui varie avec les propriétés de bioaccumulation de la substance chimique d'intérêt (un jour, un mois, la vie entière). Puisque certains modèles cinétiques permettent de prendre en compte la variabilité intra et interindividuelle dans la cinétique, il devient essentiel d'inclure cette variabilité également dans la modélisation dynamique de l'exposition. Par exemple, l'exposition externe est très souvent reconnue comme variant avec l'âge de l'individu et constitue un changement qu'il est important d'inclure dans la modélisation d'une exposition vie entière. La modélisation de l'exposition externe sur une longue période de temps conduit aussi à étudier l'évolution de l'exposition au cours de la vie de l'individu et donc l'évolution de la contamination dans son environnement.

Par conséquent, l'objectif de ce travail est d'élaborer un modèle dynamique pour l'évaluation de l'exposition sur une longue période de temps qui permet d'intégrer la variabilité intra et inter individuelle tout au long de la vie et ainsi de mieux correspondre à la modélisation cinétique.

Approches bayésiennes et Approximate bayesian computation (ABC)

Les grands principes de l'approche bayésienne et son utilisation dans l'évaluation des risques sont expliqués dans cette partie. En effet, l'approche bayésienne présente certaines caractéristiques spécifiques par rapport à l'approche fréquentiste qui peuvent expliquer son utilisation de plus en plus fréquente dans le contexte de l'évaluation des risques. Tout d'abord, les méthodes bayésiennes considèrent naturellement l'incertitude associée aux paramètres d'un modèle puisqu'une distribution est déterminée pour chaque paramètre plutôt qu'une seule valeur. La modélisation hiérarchique bayésienne fournit également un cadre naturel pour prendre en compte les différentes sources de variabilité et d'incertitude présentes dans un contexte d'évaluation des risques. En effet, dans ces modèles, la distribution des paramètres peut être interprétée en terme de variabilité, tandis que la distribution des hyper-paramètres peut être interprétée en termes d'incertitude (Pouillot et al., 2003). Enfin, grâce à la distribution a priori, les méthodes bayésiennes sont souvent recommandées pour utiliser l'information d'études antérieures ou l'opinion d'experts (Albert et al., 2012).

En ce qui concerne plus particulièrement la modélisation de la cinétique et le lien entre l'exposition externe et interne, l'approche bayésienne a été proposée dans de nombreux travaux afin de (re) estimer les demi-vies pour des modèles TK à un compartiment (Amzal et al., 2009, Albert et al., 2010), de calibrer des modèles PBTK (Loizou et Hoggs, 2011; Bois et al 2010) ou pour réaliser des approches de dosimétrie inverse (Lyon et al 2008, Bernillon et Bois 2010). Comme indiqué dans Bois et al. (2010), il existe de nombreuses sources de variabilité inter-individuelle au niveau de la toxicocinétique qui doivent être prises en compte. Afin de les modéliser correctement, mais aussi d'améliorer notre compréhension des déterminants de l'hétérogénéité au sein de la population, les modèles hiérarchiques Bayésiens sont très bien adaptés. En outre, comme nous l'avons vu, les données de biosurveillance sont très souvent disponibles au niveau individuel alors que certains paramètres sont disponibles à un niveau agrégé (population ou groupes d'individus). Par conséquent, lorsque l'on considère le lien entre les données de biosurveillance et l'exposition externe, la modélisation hiérarchique semble bien adaptée. Ce type de modélisation a d'ailleurs été proposé à plusieurs reprises dans le contexte spécifique de reconstruction de la dose ou dosimétrie inverse (Gelman et al, 1996; Bernillon et Bois, 2000; Lyons et al 2008). Un autre avantage de l'approche bayésienne dans ce contexte repose sur la distribution a priori. En effet, dans les modèles PBTK, les paramètres ont une interprétation physique ou biologique. Par conséquent, comme l'ont montré Lyon et al. (2008), la distribution a priori peut s'avérer très utile pour intégrer des connaissances préalables sur ces paramètres provenant d'études précédentes.

De nombreux travaux ayant montré la pertinence de l'approche bayésienne pour la modélisation de la cinétique dans l'évaluation du risque, ce travail se place également dans un cadre bayésien. Cependant, la volonté de prendre en compte la variabilité intra-individuelle (en plus de la variabilité inter-individuelle) qui peut être liée par exemple à des changements dans les comportements individuels avec l'âge et/ou à l'évolution de la contamination de l'environnement à travers le temps, a conduit à considérer des trajectoires individuelles d'exposition. Ces trajectoires évoluent donc dans le temps en fonction de la période écoulée entre deux prises successives comme proposé dans Verger et al. (2007), mais également en fonction de la date et de l'âge auquel l'individu est exposé. Or la vraisemblance d'un tel processus est très complexe et difficile à écrire, ce qui rend

difficile l'implémentation des algorithmes pour consuire des inférences bayésiennes classiques. C'est pourquoi il a été proposé dans ce travail, d'utiliser une approche ABC (Approximate Bayesian Computation, Beaumont et al., 2002). Il a en effet été démontré que ces méthodes étaient très appropriées en cas de problème d'inférence avec des vraisemblances complexes. Cependant, ces méthodes nécessitent de trouver «les bonnes» statistiques résumées pour le jeu de données ainsi qu'une distance appropriée. Or, ce choix peut s'avérer très délicat puisque des statistiques résumées et/ou une distance mal choisie peuvent conduire à l'échec de l'ajustement du modèle (Marjolaine et al., 2003). Des statistiques et des distances ont déjà été proposées dans plusieurs travaux antérieurs. Cependant, elles sont spécifiques à la question posée et à la nature des données et, d'après nos connaissances, les méthodes ABC n'ont jamais été appliquées dans le contexte spécifique de l'évaluation des risques ou à des données d'exposition ou de biosurveillance. Ainsi, l'un des objectifs de ce travail consiste à développer des méthodes qui permettent d'utiliser une approche ABC pour inférer sur les modèles dynamiques qui sont proposés.

Combiner la modélisation cinétique, l'évaluation de l'exposition externe et les données de biosurveillance pour améliorer l'évaluation des risques liés à des substances persistantes

Cette partie regroupe trois travaux différents qui ont pour but d'étudier comment les données d'exposition externes et de biosurveillance peuvent être combinées en utilisant la modélisation dynamique de l'exposition et des modèles toxicocinétiques. Ces travaux visent d'abord à fournir une interprétation des données de biosurveillance en termes de niveau d'exposition à plusieurs produits chimiques et/ou à partir de différentes sources d'exposition. Ils visent également à améliorer l'évaluation de l'exposition et donc l'ensemble de l'évaluation des risques liés à des produits chimiques persistants. Les méthodes développées sont appliquées à l'exposition cumulée aux dioxines, furanes et PCB-DL ainsi qu'à l'exposition agrégée au cadmium. Les modèles développés dans ce travail sont ensuite combinés avec des modèles toxicodynamiques fournis par d'autres études afin de procéder à une évaluation complète des risques.

Intégration des données de biosurveillance pour améliorer les modèles d'exposition alimentaire à l'aide d'une méthode ABC

L'objectif principal de ce travail qui fait l'objet de la publication Béchaux et al. (2014a) est de proposer une méthode permettant d'intégrer les informations fournies par les données de biosurveillance pour informer et améliorer les modèles d'exposition.

Plusieurs développements méthodologiques sont proposés. Tout d'abord, le modèle dynamique d'exposition alimentaire (KDEM) initialement proposé par Bertail et al. (2008) a été modifié afin de prendre en compte une potentielle variation historique de l'exposition à travers le

temps. KDEM a également été modifié afin de mieux intégrer la variabilité inter-individuelle et intra-individuelle grâce à la simulation des trajectoires individuelles qui utilisent des consommations et des demi-vies à l'échelle de l'individu ainsi que l'évolution de l'exposition et des demi-vies avec l'âge. Afin d'intégrer l'information fournie par les données de biosurveillance pour inférer sur les paramètres du modèle dynamique d'exposition, une méthode ABC a été proposée. La mise en œuvre d'une telle méthode a nécessité de fournir des statistiques et des distances analytiques pertinentes pour l'évaluation des risques. Compte tenu de la complexité des simulations, un algorithme efficace était nécessaire et un algorithme ABC-MCMC a donc été implémenté dans le logiciel R. En outre, un second niveau a été ajouté à l'algorithme dans le but d'effectuer la sélection de modèle en même temps que l'estimation des paramètres.

La méthode a été appliquée à des données de biosurvaillance mesurant la concentration sanguine en dioxines dans un groupe de familles de pêcheurs français. Une validation croisée a montré que le modèle prédit correctement les charges corporelles de cette population à partir de leur exposition alimentaire. Ce modèle a également permis de reconstituer l'évolution de l'exposition alimentaire aux dioxines en France entre 1930 et aujourd'hui.

Cependant, ce travail présente encore certaines limites. En effet, malgré l'implémentation d'un algorithme ABC-MCMC, le temps de calcul est encore long. Par conséquent, le développement d'un nouvel algorithme plus performant pourrait être utile, en particulier pour faire face à des modèles plus complexes ou à une population d'effectif plus important. La modélisation de la cinétique par un modèle d'ordre un à un compartiment unique était pertinente pour les dioxines, mais ne pourrait pas s'adapter à tous les produits chimiques. En outre, une telle modélisation, ne permet pas d'intégrer certains aspects anatomiques ou physiologiques. Par conséquent, la généralisation à des modèles cinétiques plus complexes comme les modèles PBTK pourrait être intéressante. En ce qui concerne l'application sur les dioxines, les furanes et les PCB-DL, il a été montré que certains congénères présentent des cinétiques d'élimination chez l'homme très différentes. Dans ce travail, cette différence de cinétique n'a pas été prise en compte et donc l'intégrer dans le modèle pourrait améliorer la modélisation. Enfin, ce travail pourrait être poursuivi afin d'aller jusqu'à l'évaluation des risques sur la base de la charge corporelle.

Vers une approche intégrée de l'évaluation des risques liés à l'exposition cumulée à des substances chimiques persistantes : estimations des risques actuels et futurs liés à l'exposition aux dioxines, furanes et PCB-DL

L'objectif de ce travail qui fait l'objet de la publication Béchaux et al. (2014b) est de développer une approche intégrée pour évaluer le risque lié à des produits chimiques persistants à partir de la charge corporelle en tenant compte de l'évolution de l'exposition au cours du temps et de l'accumulation dans le corps humain. Cette méthodologie reprend les travaux précédents qui proposaient d'intégrer l'information issue des données de biosurveillance dans la modélisation de l'exposition (Béchaux et al., 2014a) et propose de les étendre à l'évaluation du risque. L'exposition aux dioxines, furanes et PCB-DL en France reste l'exemple d'application.

Le modèle dynamique de l'exposition développé dans le travail précédent est donc utilisé et les paramètres sont également estimés par inférence bayésienne en utilisant les mêmes données de biosurveillance (c'est à dire de la concentration sanguine en dioxines dans un groupe de familles de pêcheurs en français). Cependant, contrairement aux travaux précédents, une évolution historique de l'exposition et une cinétique spécifique ont été considérées pour chaque congénère (ou groupe de congénères). Une fois ajustés, les modèles sont ensuite utilisés pour prédire la charge corporelle de la population française en combinant les données de consommation de cette population et des données de contamination des aliments en France. Le risque associé à ces charges corporelles prédites est ensuite évalué à partir de charges corporelles relatives à différents types de toxicité dans une approche probabiliste.

Ce travail fournit une estimation de l'évolution historique de l'exposition aux dioxines et furanes en France entre 1930 et aujourd'hui et montre que les niveaux d'exposition aux PCB-DL n'auraient pas évolué. Les risques liés à des modifications du fonctionnement de l'appareil thyroïdien, reproductif ou hépatique chez 4 groupes d'individus d'âge différents sont également fournis. Afin de donner un exemple d'application de cette approche pour tester des scénarii de gestion des risques, la charge corporelle de la population française en 2030 a été estimée en faisant l'hypothèse que l'exposition resterait stable à l'avenir. Ainsi, dans le cas de la toxicité thyroïdienne, la probabilité que la charge corporelle dépasse sa référence en 2009 a été estimée entre 2,8% (IC à 95%: 1.5 à 4.9%) et 3,9% (IC à 95%: 2.7 à 7.1%) (18-29 vs 60-79 ans). Malgré la diminution de l'exposition aux dioxines en France depuis les années 1970, cette probabilité devrait rester autour de 1,5% (IC 95%: 0,3 à 2,5%) en 2030 chez les 60-79 ans. En s'intéressant à la toxicité pour l'appareil reproductif, la probabilité de dépassement en 2009 se situe entre 3,1% (IC 95%: 1,4 à 5,0%) (18-29 ans) et 3,5% (IC à 95%: 2.2 à 5.2%) (30-44 ans). En 2030, cette probabilité serait négligeable chez les 18-29 ans, et faible chez les 30 à 44 ans (0,7%, IC 95%: de 0 à 1,6%). Dans le cas de la toxicité hépatique, la probabilité de dépassement en 2009 et en 2030, est négligeable y compris pour les 60-79 ans. En conclusion, cette approche indique qu'en France, bien que les niveaux de contamination des aliments en dioxines diminuent, les risques pour la santé ne peuvent être écartés.

Cette approche met en évidence que la combinaison de plusieurs méthodes et données peut améliorer l'évaluation des risques liés à des produits chimiques persistants. Cependant, la simplicité de la modélisation cinétique utilisé dans ce travail, a, d'un côté permis de prendre en compte facilement la différence de cinétique entre les congénères et d'un autre côté ne permet pas d'intégrer des informations physiologiques plus précises dans la modélisation. De plus, elle ne permet pas de prendre en compte, dans un même modèle, plusieurs voies d'exposition.

Interprétation du niveau de cadmium urinaire mesuré dans la population française en termes d'exposition alimentaire et d'exposition par le tabac.

Comme pour les deux précédents travaux, l'objectif principal de ce travail qui fait l'objet de la publication Béchaux et al. (2014c) est de lier les données portant sur l'exposition externe et les données de biosurveillance afin d'améliorer leur interprétation ainsi que l'évaluation des risques liés

à des produits chimiques persistants. Ce travail se focalise sur les concentrations de cadmium dans l'urine et les différentes sources d'exposition considérées sont l'alimentation et la fumée de cigarette.

Une modélisation dynamique de l'exposition est également utilisée pour tenir compte de l'évolution de l'exposition au cours de la vie des individus. L'approche ABC proposée dans le premier article est mise en œuvre en reprenant les mêmes statistiques résumées et distances. Cependant un algorithme ABC-SMC a été utilisé à la place de l'algorithme ABC-MCMC afin d'accélérer le temps de calcul. Contrairement aux travaux précédents, un modèle PBTK multi-compartiments est utilisé. Ce modèle permet de prendre en compte les deux voies d'exposition différentes, mais aussi de mieux décrire le processus ADME (Absorption, Distribution, Métabolisme, Elimination). Un modèle hiérarchique a été proposé afin de modéliser la variabilité inter-individuelle des paramètres physiologiques, de l'exposition alimentaire et des habitudes tabagiques.

Cette approche a permis d'expliquer les concentrations de cadmium urinaire actuelles à partir de l'exposition alimentaire passée et actuelle et de l'exposition par la fumée de cigarettes. Une partie toxicodynamique qui relie l'exposition interne mesurée par les concentrations urinaires de cadmium à des effets rénaux a été ajoutée et les risques associés ont été évalués pour les hommes et les femmes français en fonction de leur âge. Une comparaison de l'évaluation des risques à partir de l'exposition externe ou de l'exposition interne est également proposée.

La modélisation proposée présente encore certaines limites. En effet, toutes les sources potentielles de variabilité inter-individuelle n'ont pas été modélisées. Enfin, les résultats concernant l'estimation de la variation historique de l'exposition alimentaire pourraient être renforcés ou invalidés à l'aide de données de biosurveillance recueillies sur une longue période de temps.

Variabilité inter-individus dans le métabolisme par une enzyme polymorphique et lien avec les facteurs d'incertitude: focus sur les différences inter-ethniques et inter-phénotypiques dans le métabolisme par CYP2D6.

Conventionnellement, un facteur de 3,16 est utilisé par défaut pour couvrir la variabilité humaine dans la toxicocinétique dans le cadre de l'évaluation des risques liés à des substances non cancérigènes. L'objectif de ce travail est d'étudier la variabilité inter-individus dans la cinétique afin de tester la pertinence de ce facteur dans le cas d'un métabolisme par une enzyme polymorphique. Ce travail se concentre sur les adultes sains caucasiens et asiatiques dont le phénotype a été identifié.

Une revue des études pharmacocinétiques publiées fournissant des données de pharmacocinétiques recueillies chez des métaboliseurs lents (PM), intermédiaires (IM), rapides (EM) ou ultra-rapides (UM) pour des substances métabolisée par CYP2D6 a été réalisée. Les paramètres relatifs à l'exposition orale chronique (clairance orale : CL et l'aire sous la courbe de la concentration

plasmatique en fonction du temps : AUC) ont été extraits de ces publications et homogénéisés. Une méta-analyse est ensuite réalisée afin de regrouper les données issues de ces études différentes pour estimer le plus justement possible la variabilité inter-individus. Pour cela, un modèle Bayésien hiérarchique a été proposé. Ce modèle permet de décomposer la variabilité observée en variabilité inter-substrats, inter-études et inter-individus.

Ce travail met en évidence de grandes différences individuelles dans la cinétique des Caucasiens métaboliseurs rapides (coefficients de variation supérieur à 50%). Une variabilité inter-individuelle légèrement plus faible a été constatée chez les métaboliseurs lents (coefficients de variation autour de 35%) et dans la population asiatique (coefficients de variation inférieur à 40% pour les métaboliseurs rapides et intermédiaires). Les comparaisons entre les individus caucasiens métaboliseurs lents et rapides, et asiatiques, intermédiaires et rapides, ont révélé une dose interne plus importante pour les métaboliseurs lents et intermédiaires. En effet, le ratio par rapport aux métaboliseurs rapides varie de 1 à 34 et de 1 à 5. Les facteurs d'incertitude correspondants ont été calculés pour chaque sous-groupe, ainsi que pour les deux populations, en prenant en compte la prévalence de chacun des phénotypes par des simulations de Monte-Carlo. Il en résulte qu'aucun des sous-groupes ne serait couvert par le facteur par défaut de 3,16 fois et qu'un facteur de 11 serait nécessaire pour couvrir 95% des populations caucasienne et asiatique. Une relation exponentielle a été trouvée entre le pourcentage de métabolisme par CYP2D6 chez les métaboliseurs rapides et les facteurs d'incertitude pour métaboliseurs lents. Cette relation, combinée avec des résultats d'études in vitro, pourrait être utilisée pour établir des facteurs d'incertitude spécifiques aux produits chimiques ou comme entrée pour les modèles cinétiques utilisés dans l'évaluation des risques.

Discussion

Ces travaux ont montré qu'il est possible de lier l'évaluation de l'exposition externe aux doses internes mesurées y compris dans le cas de substances persistantes. En effet, l'effet de la cinétique sur l'exposition interne est d'autant plus critique lorsque la demi-vie est longue. Par conséquent, une modélisation qui considère l'exposition tout au long de la vie est d'autant plus appropriée dans le cas de substances persistantes. Ce travail ne vise pas à convaincre que toutes ces données sont nécessaires pour évaluer l'exposition, mais plutôt à investiguer ce qui peut être appris en combinant toutes les données. Ce travail amène également d'autres réflexions, en particulier, l'intégration de la variabilité inter et intra-individuelle est discutée en traitant particulièrement du cas du polymorphisme dans le métabolisme.

Le travail proposé dans ce manuscrit a été réalisé dans le cadre d'une exposition unique ou d'exposition cumulée. Cependant, les individus sont exposés à une combinaison très complexe et variable de produits chimiques provenant de différentes sources. Ainsi, l'évaluation des risques liés à l'exposition combinée à plusieurs substances formant alors un mélange, constitue un nouveau défi pour l'évaluation des risques. Il est donc proposé d'étudier comment ce travail pourrait être étendu à un contexte d'exposition à des mélanges de substances. Dans un premier temps, une méthode permettant d'identifier les principaux mélanges auxquels une population est exposée à de

nombreuses substances, à partir de son exposition alimentaire, est proposée. Cette méthode permet aussi de relier l'exposition à des mélanges au régime alimentaire des individus, ce qui facilite l'interprétation de l'exposition combinée et fait l'objet de la publication Béchaux et al. (2013). Enfin, différentes méthodes existent désormais pour modéliser des interactions toxicocinétiques entre plusieurs substances, notamment par des modèles PBTK multi-substances. Par conséquent, l'application de modèles PBTK à des mélanges de substances qui présentent une forte probabilité de co-exposition dans la population permettrait de prédire des doses internes plus proches de la réalité.

Abstract:

Traditionally in chemical risk assessment, external and internal exposures are considered separately and toxicokinetic is not taken into account. However, both exposures can actually provide complementary information when considered together in an exposure-response continuum. The objective of this work is to provide a dynamic modeling of exposure that integrates toxicokinetic and makes it possible to link external and internal exposure in order to improve risk assessment. In a first step, a dynamic model, based on a simple single compartment toxicokinetic model is proposed. Such approach allows to take into account the accumulation of the chemical in the body according to its kinetic and the successive dietary exposures which occur during the whole life of the individual. An ABC method (Approximate Bayesian Computation) is proposed to input biomonitoring data in order to inform this model. This makes it possible to reconstruct the past exposure of the individuals. In a second step, a PBTK model (physiologically-based toxicokinetic) is integrated into this dynamic model of exposure in order to consider several pathways of exposure and to improve the description of the kinetic. Biomonitoring data are also included in a Bayesian hierarchical model using the ABC method. Application of this approach to cumulative exposure to dioxins, furans and DL-PCBs as well as aggregate exposure to cadmium, highlight that such approach improves not only the interpretation of biomonitoring data but also the risk assessment of these persistent substances. Moreover, it allows to better take into account the variability within and between individuals. In particular, among the sources of inter-individual variability that can be integrated into such a model, the variability associated with polymorphism in the human metabolism is well known. A meta-analysis was thus conducted with data from pharmaceutical studies and a Bayesian hierarchical model to investigate this variability for substrates of CYP2D6. It highlights a high inter-individual variability related to polymorphism of CYP2D6 which would be necessary to integrate in chemical risk assessment of CYP2D6 substrates. Finally, the extrapolation of this work to chemical mixtures is investigated. Indeed, the application of PBTK models to mixtures of chemicals that have a high probability of co-exposure in the population would make it possible to take interactions into account and thus to predict more accurate internal doses.

Key words: Toxicokinetic, aggregated exposure, cumulative exposure, biomonitoring, baysian inferences, ABC methods, chemical risk assessment, cadmium, dioxins.

Résumé:

A l'heure actuelle l'évaluation du risque à des produits chimiques considère séparément les expositions externe et interne et ne prend pas en compte la toxicocinétique. Cependant, ces deux types d'exposition peuvent apporter des informations complémentaires lorsqu'elles sont intégrées dans un même continuum exposition-réponse. L'objectif de ce travail est de proposer une modélisation dynamique de l'exposition qui intègre la toxicocinétique et permet de relier l'exposition externe et interne dans le but d'améliorer l'évaluation des risques. Dans un premier temps, un modèle dynamique d'exposition basé sur un modèle toxicocinétique simple, à un compartiment, est proposé. Celui-ci permet de prendre en compte l'accumulation de la substance dans l'organisme en fonction de sa cinétique d'élimination mais aussi des expositions alimentaires successives au cours de la vie de l'individu. Une méthode ABC (Approximate Bayesian Computation) est proposée afin d'intégrer les données de biosurveillance pour informer ce modèle. Celles-ci permettent notamment de reconstituer l'exposition passée des individus. Dans un second temps, un modèle PBTK (physiologically-based toxicokinetic) est intégré dans ce modèle dynamique d'exposition afin de considérer plusieurs voies d'exposition et d'améliorer la description de la cinétique des substances. Les données de biosurveillance sont également utilisées dans un modèle bayésien hiérarchique à partir de la méthode ABC. A travers les applications à l'exposition combinée aux dioxines, furanes et PCB-DL ainsi qu'à l'exposition agrégée au cadmium, il est montré que cette approche permet d'améliorer à la fois l'interprétation des données de biosurveillance mais aussi l'évaluation du risque lié à ces substances persistantes. Elle permet également de mieux prendre en compte la variabilité intra et inter individus dans l'évaluation du risque. Parmi les sources de variabilité inter-individus qui peuvent être intégrées dans une telle modélisation, la variabilité liée au polymorphisme dans le métabolisme humain est peu renseignée. Une méta-analyse est proposée à partir de données pharmaceutiques issues de la littérature et d'un modèle bayésien hiérarchique afin d'investiguer cette variabilité pour des substrats de CYP2D6. Cette étude met en évidence une forte variabilité inter-individus liée au polymorphisme de CYP2D6 qu'il serait nécessaire d'intégrer dans l'évaluation du risque à des substances substrats de CYP2D6. Enfin, l'extrapolation de ce travail à des mélanges de substances est étudiée. En effet, l'application de modèles PBTK à des mélanges de substances qui présentent une forte probabilité de co-exposition dans la population permettrait de prendre en compte les interactions et ainsi de prédire des doses internes plus réalistes.

Mots-clés : Toxicocinétique, exposition agrégée, exposition cumulées, biosurveillance, inférences bayesiennes, méthodes ABC, évaluation du risque, cadmium, dioxines.