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.

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 $S_E^{-1} C S_Q$ and

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

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

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 $S_E^{-1} C S_Q W$ is introduced. Therefore $W^* = S_E^{-1} C S_Q W R^{-1}$ 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.