Estimated dietary pesticide exposure from plant-based foods using NMF-derived profiles in a large sample of French adults

Pauline Rebouillat, Rodolphe Vidal, Jean Pierre Cravedi, Bruno Taupier-Letage, Laurent Debrauwer, Laurence Gamet-Payrastre, Mathilde Touvier, Serge Hercberg, Denis Lairon, Julia Baudry, et al.

Estimated dietary pesticide exposure from plant-based foods using NMF-derived profiles in a large sample of French adults

Pauline Rebouillat[1], Rodolphe Vidal[2], Jean-Pierre Cravedi[3], Bruno Taupier-Letage[2], Laurent Debrauwer[3], Laurence Gamet-Payrastre[3], Mathilde Touvier[1], Serge Hercberg[1-5], Denis Lairon[4], Julia Baudry[1], Emmanuelle Kesse-Guyot[1].

[1] Sorbonne Paris Nord University, Inserm, INRAE, Cnam, Nutritional Epidemiology Research Team (EREN), Epidemiology and Statistics Research Center – University of Paris (CRESS), 93017 Bobigny, France

[2] Institut de l'Agriculture et de l'Alimentation Biologiques (ITAB), 75595 Paris, France

[3] Toxalim (Research Centre in Food Toxicology), Université de Toulouse, INRAE, ENVT, INP-Purpan, UPS, 31027 Toulouse, France

[4] Aix Marseille Université, INSERM, INRAE, C2VN, 13005 Marseille, France

[5] Département de Santé Publique, Hôpital Avicenne, 93017 Bobigny, France

Names for PubMed indexing: Rebouillat, Vidal, Cravedi, Taupier-Letage, Debrauwer, Gamet-Payrastre, Touvier, Hercberg, Lairon, Baudry, Kesse-Guyot

Disclaimers: The authors declare no conflict of interest.

***Corresponding author**:

Pauline REBOUILLAT, Equipe de Recherche en Epidémiologie Nutritionnelle (EREN)

SMBH Université Sorbonne Paris Nord, 74 rue Marcel Cachin, 93017 Bobigny, France.

p.rebouillat@eren.smbh.univ-paris13.fr

+33 1 48 38 89 79

ORCID Identifiers :

Pauline Rebouillat : 0000-0002-7270-6032

Mathilde Touvier : 0000-0002-8322-8857

Emmanuelle Kesse-Guyot : 0000-0002-9715-3534

Source of support: The NutriNet-Santé study is supported by the following public institutions: French Ministry of Health (DGS), Santé Publique France, the National Institute for Health and Medical Research (INSERM), the National Institute for Agricultural Research (INRA), the National Conservatory of Arts and Crafts (CNAM) and the Sorbonne Paris Nord University. The BioNutriNet project was supported by the French National Research Agency (ANR) in the context of the 2013 Programme de Recherche Systèmes Alimentaires Durables (ANR-13-ALID-0001). Pauline Rebouillat is supported by a doctoral fellowship from Sorbonne Paris Nord University.

Short running head: Dietary pesticides exposure

Abbreviations:

ADI: Acceptable Daily Intake; ANSES: Agence Nationale de sécurité sanitaire de l’alimentation, de l’environnement et du travail ; CNIL: Commission Nationale de l’Informatique et des Libertés ; CVUA: Chemisches und Veterinäruntersuchungsamt; EAT: Etude de l'Alimentation Totale; EDI: Estimated Daily Intake; EFSA: European Food and Safety Authority; EudraCT: European Union Drug Regulating Authorities Clinical Trials; FFQ: Food Frequency Questionnaire; IRB INSERM: Institutional Review Board of the French Institute for Health and Medical Research; MESA: Multi-Ethnic Study of Atherosclerosis; MUFA: Mono Unsaturated Fatty Acids; NMF: Non-negative Matrix Factorization; PUFA: Poly Unsaturated Fatty Acids; SFA: Saturated Fatty Acids.

Clinical Trial Registry: NCT03335644

Abstract:

Purpose: This study, conducted in participants of the NutriNet-Santé cohort, aims to identify dietary pesticide exposure profiles (derived from Non-negative Matrix Factorization) from conventional and organic foods among a large sample of general population French adults.

Methods: Organic and conventional dietary intakes were assessed using a self-administered semi-quantitative food frequency questionnaire. Exposure to 25 commonly used pesticides was evaluated using food contamination data from Chemisches und Veterinäruntersuchungsamt Stuttgart accounting for farming system (organic or conventional). Dietary pesticide exposure profiles were identified using Non-Negative Matrix factorization (NMF), especially adapted for non-negative data with excess zeros. The NMF scores were introduced in a hierarchical clustering process.

Results: Overall, the identified clusters (n=34,193) seemed to be exposed to the same compounds with gradual intensity. Cluster 1 displayed the lowest energy-intake and estimated dietary pesticide exposure, high organic food (OF) consumption (23.3%) and a higher proportion of male participants than other groups. Clusters 2 and 5 presented intermediate energy intake, lower OF consumption and intermediate estimated pesticide exposure. Cluster 3 showed high conventional fruits and vegetable (FV) intake, high estimated pesticide exposure, and fewer smokers. Cluster 4 estimated pesticide exposure varied more across compounds than for other clusters, with highest estimated exposures for acetamiprid, azadirachtin, cypermethrin, pyrethrins, spinosad. OF proportion in the diet was the highest (31.5%).

Conclusion: Estimated dietary pesticide exposures appeared to vary across the clusters and to be related to OF proportion in the diet.

Keywords: dietary exposure; pesticides; organic farming; epidemiology.

Introduction:

Plant protection products also referred hereafter as pesticides are used in large quantities in current agriculture all over Europe [1]. France is one of the countries presenting the highest use of pesticides in Europe, with between 60 000 and 80 000 tons sold in 2016 [2, 3] and the fifth highest European country, in terms of average use of pesticides by surface area, in 2001-2003 [4].

Current use of pesticides raises environmental and health concerns. On the environmental side, pesticides can have detrimental effects on biodiversity, soil quality, and water quality [5]. In toxicological studies, a wide range effects have been reported for pesticide active substances and/or metabolites [6] such as alteration of genetic material, neurotransmission interference, endocrine disrupting effects, deregulation of cellular apoptosis, and oxidative stress induction, causing neurological, respiratory, dermatological, reproductive or developmental disorders, and cancers [4, 7, 8]. Considering human health, several epidemiological studies have found associations between occupational exposure to pesticides in agricultural settings and diverse diseases including Parkinson disease [9], lymphomas and Alzheimer's disease [4, 10]. However, the general population is also exposed to cocktail of these compounds or their metabolites at very low doses, mainly through the dietary pathway [11], and data on this type of exposure in the general population is lacking .

More than 400 active substances are currently authorized as plant protection products within the European Union (EU) according to the EU Pesticides database [12], to which must be added corresponding metabolites, and pesticide formulants., Thus, the assessment of the dietary exposure to pesticide is complex and has a high financial burden because of the range of food commodities to sample and test, in addition to the multiresidues analyses to carry out.

The French Agency for Food, Environmental and Occupational Health Safety (*Agence Nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail*, ANSES) has described dietary exposures, by merging contamination data and consumption levels, in the general population in the Total Diet Studies (*Etude de l'Alimentation Totale)* published for the periods 2000-2004 and 2006. From these data, exposures to some pesticides were largely related to fruits and vegetables consumption [13, 14]. However, in this study, farming system (i.e. conventional versus organic) to produce foods was not taken into account while non-use of synthetic fertilizers and pesticides in organic farming leads to less (in frequency and concentration) or absence of pesticides residues in organic foods compared to conventional ones by regulation (less than 100 active substances compared to 400 in conventional farming) [15, 16].

Given the increase in Europe of organic food consumption over the last years and the differences between farming practices (synthetic pesticides and fertilizers use) used to produce organic food versus "conventional" food, it is important to take this parameter into account in order to discriminate more effectively human exposure profiles.

The purpose of this study was to identify dietary pesticide exposure profiles (derived from Non-negative Matrix Factorization)from conventional and organic foods among a large sample of the French web-based cohort NutriNet-Santé.

Material and Methods:

Study population:

NutriNet-Santé study is an on-going web-based prospective cohort of French adult volunteers launched in May 2009 [17]. Participants completed at baseline a set of self-administered questionnaires repeated every year, regarding socio-economic status, anthropometrics, lifestyle, physical activity, and dietary intakes. This set of web-based questionnaires has been validated against traditional methods [18–20]. Complementary questionnaires on dietary behaviors, nutritional and health status were also proposed to participants during follow-up. The NutriNet-Santé study is conducted in accordance with the Declaration of Helsinki, and all procedures were approved by the Institutional Review Board of the French Institute for Health and Medical Research (IRB Inserm 0000388FWA00005831) and the Commission Nationale de l'Informatique et des Libertés (CNIL 908,450 and 909,216). All participants provided their informed consent with an electronic signature, and this study is registered in EudraCT (European Union Drug Regulating Authorities Clinical Trials; no. 2013-000929-31) and in ClinicalTrials.gov (NCT03335644).

Dietary intake assessment:

Dietary intake of organic and conventional foods was assessed using a 264-item web-based self-administered semi-quantitative organic food frequency questionnaire (Org-FFQ) between June and December 2014. The Org-FFQ was constructed according to an existing validated FFQ [21] to which a 5-point ordinal scale was added to measure the frequency of organic food consumption. For each of the 264 questionnaire items, participants provided their frequency of consumption and the quantity consumed for the previous year. For organic-labelled food and beverages, participants were asked "How often was the product of organic origin?" by selecting 1 of the 5 following frequency modalities: never, rarely, half-of-time, often, or always. The organic food consumption parameter was then obtained by attributing the respective percentages, 0, 25, 50, 75, and 100, to the modalities. Full details regarding the

Org-FFQ and sensitivity analyses as regards the weighting have been published elsewhere [22].

All food and beverage items were aggregated into 16 food groups. Nutritional values were obtained from a published food composition database [23]. The global proportion of organic food in the diet was calculated by computing the ratio between the sum of all organic items (in g/d) and the sum of all items (organic and conventional) excluding water (in g/d). The proportion of organic food for each of the 16 food groups (organic food ratios) was calculated using the same method.

Dietary scores :

PANDiet Score was initially developed by Verger et al.[24], and updated by De Gavelle et al. 2018 [25]. It measures the probability of adequate nutrient intake based on current nutrient reference values. It is composed of two subscales, one referring to adequacy and the other to moderation. The following nutrients were included in the score: protein, total carbohydrate, total fat, added simple carbohydrates, saturated fatty acids, poly-unsaturated fatty acids, n-3 fatty acids, n-6 fatty acids, cholesterol, fiber, vitamins A, B1, B2, B3, B5, B6, B9, B12, C, D and E, Ca, Mg, Zn, P, K, Fe, I, Se and Na. The final score (/100) is the arithmetic mean of the two subscores. The provegetarian score, developed by Martinez et al. was computed as follows [26] : 7 vegetable food groups and 5 animal food groups (22) were defined and sex-specific quintiles adjusted for total energy intake were calculated. For each plant component, 1 to 5 points were allocated to quintile 1 to 5 and for animal food groups the scoring was reversed. The provegetarian score was obtained by summing each quintile value of vegetable food group and each reverse quintile value of animal food group thus ranging from 12 (low plant food consumption) to 60 (high plant food consumption).

Pesticide exposure assessment:

Dietary pesticide exposure was evaluated by combining concentration values with dietary intakes of each adult using contamination data from the Chemisches und Veterinäruntersuchungsamt Stuttgart (CVUAS) database on pesticides, designated as a European Union reference laboratory for pesticide measurements requiring single residue methods for analysis [27]. CVUAS aims to analyze plant products from more than 88 countries (in the European Union and outside) and the database includes a wide variety of contaminants (pesticides, hormones etc.). It contains more than 6.7 million datapoints (i.e. a result for a pesticide residue and a product), with 1 million referring to organic data. In the present work, four year data (2012-2015) were pooled. Twenty-five commonly used pesticides were selected, given the data availability in the CVUAS database (some pesticides widely used were excluded from the analysis due to lack of data, *e.g.* dithiocarbamates) and on data presented in the 2015 EFSA Report on pesticide residues in food (most frequently quantified or most frequently exceeding ARfD/ADI compounds)[28]. Three pesticides commonly used in organic agricultural systems (azadirachtin, pyrethrins and spinosad) were included. The 264 Org-FFQ items were decomposed into ingredients and all ingredients present in a proportion of at least 5% in at least one food item (i.e. 442 ingredients) were considered in this study. Animal-based ingredients were excluded, as CVUAS database only encompassed plant-based ingredients. Indeed, plant-based foods have markedly frequent and higher pesticides residues levels than foods of animal origin [29]. The resulting 180 plant ingredients were matched to CVUAS database and then were attributed a contamination value in organic and conventional farming modes (as the mean of corresponding data point). A flowchart of the different steps for the decomposition and matching is shown in Supplemental Material 2.

For each ingredient/pesticide couple in conventional and organic agriculture we determined a frequency of detection and a frequency of quantification using the formula as follows:

$$Frequency\ of detection = 100 \times \frac{Number\ of analyses - Number\ of\ undetected}{Number\ of\ analyses}$$

$$Frequency\ of quantification = 100 \times \frac{Number\ of\ analyses - Number\ of\ unquantified}{Number\ of\ analyses}$$

Treatment of undetected data has been extensively described elsewhere [30].

As food consumption data from NutriNet-Santé referred to edible foods (bone-free, peeled or cooked products), edibility and cooking factors were allocated to each ingredient when necessary [31, 32]. However, potential concentration or dilution effects during washing or cooking on pesticide residue levels were not accounted for. The same conversion factors were used for both conventional and organic products.

For each pesticide, the estimated daily intake (EDI) (in µg/kg bw/d) under both lower- and upper-bound scenarios was calculated using methods recommended by EFSA and WHO workshop [33, 34].

Statistical Analyses:

For the present study, participants who completed the Org-FFQ between June and December 2014 (N = 37,685), with no missing covariates (N =37,305), who were not detected as under- or over-reporters (N = 35,196), who lived in mainland France (N=34,193) to permit the computation of a weighting procedure described below were selected. The detection method for under and overreporters was based on the comparison between energy intake and energy requirement and is extensively described in a previous article by Baudry et al.[35]. Moreover, in order to make the study sample more representative of the French population, for each gender, weighting was calculated using the iterative proportional fitting procedure of the %CALMAR macro [36] using 2009 national census reports [37] on age, occupational category, educational level, area of residence and marital status and all analyses were weighted.

Dietary pesticide exposure profiles were identified using Non-Negative Matrix factorization (NMF) (see Supplemental material 1), a specially adapted method for non-negative data with excess zeros developed by Lee et al [38]. NMF was performed on the 25 pesticide exposure variables. Optimal number of NMF components, *K*, was chosen after running NMF for different values of *K*, and evaluating criteria such as the residual sum of squares, as suggested by Zetlaoui et al.[39]. Four components were retained. The NMF scores for these four components were then introduced into a two-step clustering process : Ward's hierarchical clustering using the four dimensions was performed, followed by a non-hierarchical, K-means clustering procedure based on the earlier hierarchical clustering. The number of clusters was determined using standard criteria (Cubic Clutering Criterion, PseudoF, Pseudo T^2) [40]. NMF was performed using R's NMF package [41], and the clustering procedure was performed using SAS® PROC CLUSTER and PROC FASTCLUS. Identified clusters were described in terms of sociodemographics, dietary patterns and exposure to pesticides.

Participant characteristics across clusters are reported as means (SD). P values refer to Chi-square tests and ANCOVA models for food groups are adjusted for energy intake.

Relative mean differences of estimated dietary pesticide exposure, compared to the mean of the whole sample were calculated for each cluster.

For the 16 food groups, adjusted means for energy intake were computed and corrected for multiple comparisons using Tukey procedure. Energy adjustment was conducted using the residual method for nutrient intake [42]. For information, crude average of food consumption across clusters are shown in **Supplemental Table 3**.

Two-sided tests were used, and a P value of <0.05 was considered significant. Data management and statistical analyses were performed using SAS (version 9.4; SAS Institute, Inc.).

Results:

Sociodemographic characteristics of the weighted sample are presented in **Table 1**. All the descriptive statistics presented in this section are weighted. Given the small size of cluster 6, statistical tests were performed only between the 5 first clusters for sociodemographic and nutritional characteristics.

The total analyzed population was constituted of 34,193 individuals, 76% of them were women and the average age was 50 years (SD=16). When weighting procedure was applied, most represented monthly income (per household unit, obtained using the income by month in the household and the composition of the household) categories were 1,200 to 1,800€ (29%) and 1,800 to 2,700€ (26%) categories. One third of the population was retired. Most participants had a high-school diploma (49%) and lived in an urban setting of more than 200,000 inhabitants (41%).

After the hierarchical clustering process, six clusters were identified. Relative mean differences (compared to whole sample mean) for estimated dietary pesticide exposure for each cluster are presented in **Table 2**. For information purpose, absolute values are presented in **Supplemental Table 1**. Estimated pesticide exposure under the lower-bound scenario were all inferior to ADI except for two individuals for chlorpyrifos and imazalil pesticides (data not shown). Globally, in our estimation, all clusters were exposed to the same pesticide moieties, with gradual intensity. Nutritional characteristics are presented in **Table 3** and dietary intakes for food groups in **Table 4**.

The cluster 1 constituted 51% of our sample. Regarding estimated dietary pesticide exposure, this cluster was overall the less exposed (except for anthraquinone and pyrethrins pesticides). Participants in this cluster were younger and more often men than in other clusters. Predominantly represented monthly income category was 1200 to 1800€ per household unit (29%). This cluster comprised the most often sedentary participants (23%). The proportion of overweight in this cluster was the highest among all clusters (40%). This population group displayed the lowest energy intake of all clusters, and intermediate proportion of pescetarians, vegetarians and vegans (see **Table 3**). They also exhibited high consumptions of alcoholic beverages, fast food, sweetened foods and meat/poultry/processed meat (see **Table 4**). They presented an intermediate global proportion of organic food in their diet (23%), a relatively low Provegetarian score but organic proportions for plant-based food groups were high (41%).

The cluster 2 included 28% of the sample. Estimated dietary pesticide exposure was lower than for clusters 3 to 6. However, they showed higher exposure to methamidophos. This cluster was constituted of 76% of women, 36% of retired persons, and a high proportion had obtained a high school diploma (51%). They had intermediate energy intake and a low organic food proportion in their diet (14%). Organic food proportion for the 16 food groups were lower than cluster 1. Similarly to the 1st cluster, they showed high consumptions of alcoholic beverages, fast food, sweetened foods and meat/poultry/processed meat.

The highest estimated pesticide exposures for anthraquinone, chlorpropham, fenhexamid, methamidophos were observed in the 3rd cluster (representing 2% of the sample). This cluster was the second most highly exposed of the 6 clusters for other pesticides. Cluster 3 was composed of the highest proportion of unemployed subjects (12%), of employees (25%) and of participants with a high school diploma (57%). This group had the highest proportion of subjects living in a city with more than 200,000 inhabitants (44%). The proportion of overweight individuals was the lowest compared to other clusters (19%). The proportion of never-smokers was the lowest (48%). Individuals from this group had intermediate energy intake and the lowest average organic food proportion in the diet (12%). Relatively high proportions of pescetarians, vegetarians and vegans were observed in this group (2 to 4%). They showed the highest consumption of seafood and extra food groups and the lowest consumption of whole-grain products. Organic food proportion for food groups were low, eggs being the most consumed food group as organic (31%).

In cluster 4 (8% of the sample), estimated pesticide exposure varied more across compounds than for other clusters. Participants showed the highest exposures across clusters for acetamiprid, azadirachtin, cypermethrin, pyrethrins, spinosad pesticides. Exposures to anthraquinone, carbendazim, dimethoate were intermediate. This group was composed of 87% women, had the highest proportion of individuals with a monthly income superior to 2700€ per household unit (21%) and of self-employed/farmer occupational category (4%). The highest proportion of subjects living in a rural setting (29%) and of former smokers (40%) was found in this group. This cluster had relatively low energy intake, the highest average proportion of organic food in the diet (32%) and high proportions of pescetarians, vegetarians and vegans. The provegetarian score was high. Highest consumptions of non-alcoholic drinks, whole grain products, oil and soy, and lowest consumption of dairy products were observed in this group. Organic food proportion for vegetable food groups were the highest compared to other clusters: more than 40% organic for whole grain products, fruits and vegetables and starchy foods.

Cluster 5 (11% of sample) was intermediate in terms of estimated pesticide exposure; higher than clusters 1 and 2 and lower than clusters 3 and 6, apart from exposure to methamidophos. This group had the highest proportion of never-smokers (55%). Cluster 5 dietary intakes were characterized by intermediate energy intake and low organic food proportion in the diet (12%). Low consumption can be underlined for whole-grain products group and soy-based products. Organic food proportion for food groups were low (less than 18%) except for Oil and Eggs.

Cluster 6 was constituted of 16 individuals only (less than 1% of the sample). Participants exhibited very distinctive characteristics : 97% were women, 66% were retired. Given the small number of individuals in this cluster, we have chosen not to comment further on their characteristics and exposure, as estimation for such small clusters might be unreliable [43]. Descriptive statistics of upper-bound scenario data are presented in **Supplemental Table 2**. The gradient observed for lower-bound scenario has been also observed for upper-bound scenario.

Discussion:

This study identified different dietary pesticide exposure profiles (derived from Non-negative Matrix Factorization) among a large sample from the French web-based cohort NutriNet-Santé. Using a specially adapted method for this type of left-censored data, our analysis established 6 clusters as regards estimated dietary exposure to 25 commonly used pesticides. It is the first study to estimate exposure on a large population sample accounting for different farming practices using or not synthetic pesticides (conventional versus organic farming). Our analysis showed that the less exposed group was the one with the most individuals (51% of the whole sample for cluster 1). On the other side, the most exposed cluster (cluster 3) had high energy intake, high intakes of conventional fruits and vegetablesand low organic proportion for plant-based food groups. It was constituted of much less individuals (2% of whole sample). The global proportion of organic food in the diet was inversely correlated with estimated levels of exposure to pesticides.

Few studies have examined dietary exposures to pesticides in general population.

The Etude de l'Alimentation Totale (EAT) [13, 14, 44] studies, conducted by the French Agency for Food, Environmental and Occupational Health Safety (ANSES) have analyzed exposure to pesticides and chemicals from food but did not consider the potential role of farming practices to produce foods. Indeed, the production system can influence the pesticide residue levels in food, as shown in EFSA's last report on pesticide residues in food [28] but also in the MESA Study, in United States [45] and in a review of scientific articles [46]. Not accounting for farming practice thus makes difficult to compare the EAT study findings with our study. In addition, the EAT study used data from 2006, eight years prior to the BioNutriNet questionnaire (2014) and considered pesticides were different, due to changes in legislation on the use of pesticides (prohibitions, limitations of use). However, high estimated levels of Imazalil (fungicides used, for instance, in citrus and potatoes production) were found in our study, especially for clusters with high fruit and vegetables intakes, in accordance with the EAT Study analyses of 2013 [44].

Herein, we used a factorization method specially adapted for this type of data which allowed us to describe data using a "*profile approach"* to capture the exposure towards all 25 pesticides concomitantly rather than individual compounds as it is classically done in toxicological studies. Our data can be helpful to design future studies investigating the possible synergistic effects of mixtures of pesticide residues, based on realistic and comprehensive exposure scenarios.

Our analysis showed that the less exposed cluster (cluster 1) had a high proportion of organic food in the diet and conversely the most exposed cluster (cluster 6) had the lowest proportion of organic food in the diet. This is consistent with other studies showing that organic eaters are usually less exposed to synthetic pesticides [30, 45, 47].

For example, in another study conducted in the NutriNet-santé cohort we found that regular organic eaters had lower urinary levels of organophosphorus and pyrethrynoïd pesticides metabolites and lower dietary pesticide exposure, which is consistent with EFSA's surveillance differentiated (organic vs conventional) data [30, 47].

However, some nuances to this gradient can be drawn: cluster 1 was the less exposed cluster but it was also the cluster with the lowest consumption of fruit and vegetables. This quantity was nevertheless higher than the World Health Organization and official French Dietary Guidelines recommending more than 400 grams of fruit and vegetables per day.

Cluster 4 individuals consumed the highest proportion of organic food in their diet (31.5%) but exhibited high range of exposure levels depending on the compounds. This could be explained by relatively high intakes of non-alcoholic drinks (fruit juices), fruits and vegetables and whole foods, as exposure levels to pesticides used for fruit and vegetables (acetamiprid, azadirachtin, cypermethrin) were particularly high. Pyrethrins and spinosad exposure levels were also high in this group where proportion of organic food was important.

This is consistent with agricultural practices, as these two pesticides are authorized in organic farming.

Overall, it seems like all clusters were exposed to all pesticides but with gradual intensity. This is not surprising as diet is constituted from very varied foods with different level of consumption constituting a continuum that can be observed in the pesticide exposure derived from foods with variable proportions from conventional and organic agriculture.

We elected to primarily consider the lower-bound scenario, as upper-bound scenario imputes data even when values are null or almost null. As organic food contain less and less frequently pesticide residues [28], use of upper-bound scenario would have artificially overestimated pesticide exposure. However, similar trends were found in both scenarios in terms of exposure.

Some limitations related to our sample should be mentioned. The NutriNet-Santé cohort includes volunteers, probably more interested in food and health issues than the general population, with a majority of women and highly educated individuals [48]. The weighting process was done in order to limit selection bias but it cannot be completely dismissed. Dietary intakes were self-reported and the use of the organic food consumption scale and food frequency questionnaire may have led to an overestimation of intakes of organic food. However, the FFQ was derived from another validated FFQ [21] and used the same structure. Indeed, proportion of organic food in the diet was higher than that mentioned in the Kantar purchase data published in 2017 [49]. This can be due to our specific population and/or the FFQ but it should also be mentioned that the Kantar study does not include some purchase places such as fresh produce markets. However, when looking at the percentage of non-consumers of organic foods, statistics from our study are consistent with the survey conducted by the French Organic Agency, and the survey question “Have you consumed any organic products over the past year?” where 12% of the 506 respondents reported never consuming any organic foods in 2015 [50]. In our study, the percentage of non-consumers was 10% in 2014. The dietary exposure in the general French population might be higher for synthetic pesticides, as organic consumption is lower, but this can be balanced by the fact that NutriNet population has higher intakes for fruits and vegetables therefore potentially a higher exposure. This type of data, linked to health outcomes could be integrated to benefit/risk analyses, which are currently lacking on this topic. Available literature to date, is leaning more in the direction of numerous benefits of fruit and vegetable consumption compared to dietary pesticide risks [51, 52] . However, scarcity of data does not make it possible to firmly draw conclusion.

Other limitations can be mentioned.

Firstly, it is noteworthy that the descriptive aim of this study limits the possibility to establish associations between different characteristics of the diet and dietary pesticide exposure. Cluster analysis allowed us to explore the exposure of different groups within the population. However, the clusters that we identified are complex making it difficult to describe and characterize them. Only global trends and relative comparison could be commented on. Secondly, data was not available for animal-based products, which could have led to an underestimation of the exposure levels and some disparities in the coverage of plant food. However, residues of currently authorized pesticides are usually found in plant-based food and less frequently in food of animal origin [28]. In the EAT 2 Study, the clusters with the highest exposure to pesticides were those with elevated intakes of fruit and vegetables or plant-based foods. This is concordant with our findings, namely exposures observed in cluster 3 and 6, which in addition have low organic fruits and vegetables rates.

It is possible that organic producers would use larger quantities of some pesticides authorized in organic production systems in order to compensate the reduced number of unauthorized pesticides although this phenomenon is not yet documented. This could be the case for mineral-based pesticides such as copper or sulfur (authorized both in organic and conventional system), that we could not include as no data were available. However, no convincing adverse health effects for the general population have been reported until now while issues on the environmental level have been highlighted in recent years [53].

Pesticide data were from Germany but covered foods of all European Union's countries. It is noticeable that pesticide regulations are the same for France and Germany and should not have modified results. Analyzed products were products marketed in the whole European Union.

Finally, potential concentration or dilution effects during washing or cooking on pesticide residue levels were not taken into account., These aspects would be very interesting to be accounted for given their influences on residue concentration [54]. However, such factors are not available yet for a sufficient number of food/pesticide couples. In addition, this would require precise information on participants' peeling and washing practices.

Conclusion:

This study reports the characterization of dietary exposure to pesticides of French adults with variable proportion of organic food in their diet. In our sample, we have observed an inverse correlation between exposure to pesticides from diet and the proportion of organic food in the diet. It should be kept in mind that the less exposed individuals constitute the largest cluster, which is reassuring on a public health perspective.

Our study provides information on dietary pesticide exposure in a sample of the general population, which is not well-documented. It is a necessary first step before studying the specific role, independently of the nutritional quality of diets, of pesticide exposure on health. It would be also interesting to integrate the distinction between organically versus conventionally produced food in national and European pesticide exposure surveillance studies in order to have more accurate exposure measurement in representative populations.

Contributions

The authors' contributions are as follows:

EK-G, MT, and SH conducted the study.

RV, DL, JB and EK-G conducted the research and implemented databases.

PR performed statistical analyses and drafted the manuscript.

All authors critically helped in the interpretation of results, revised the manuscript and provided relevant intellectual input. They all read and approved the final manuscript. EK-G supervised the study, had primary responsibility for the final content, she is the guarantor.

Acknowledgements

We especially thank Cédric Agaesse, Vristi Desan and Cynthia Perlin (dietitians); Thi Hong Van Duong, Younes Esseddik (IT manager), Paul Flanzy, Régis Gatibelza, Jagatjit Mohinder and Aladi Timera (computer scientists); Julien Allegre, Nathalie Arnault, Laurent Bourhis and Fabien Szabo de Edelenyi, PhD (supervisor) (data-manager/statisticians) for their technical contribution to the NutriNet-Santé study and Nathalie Druesne-Pecollo, PhD (operational manager).

We also thank the CVUAS for the pesticide residue database and Noémie Soton for her contribution to the data management of the CVUAS database. We warmly thank all of the dedicated and conscientious volunteers involved in the NutriNet-Santé cohort.

Disclaimers: The authors declare no conflict of interest.

References

1. Popp J, Pető K, Nagy J (2013) Pesticide productivity and food security. A review. Agron Sustain Dev 33:243–255. https://doi.org/10.1007/s13593-012-0105-x

2. Eurostat - Tables, Graphs and Maps Interface (TGM) map. https://ec.europa.eu/eurostat/tgm/mapToolClosed.do?tab=map&init=1&plugin=1&language=en&pcode=tai02&toolbox=types. Accessed 28 Jun 2019

3. L'utilisation des pesticides en France : état des lieux et perspectives de réduction. https://agriculture.gouv.fr/lutilisation-des-pesticides-en-france-etat-des-lieux-et-perspectives-de-reduction. Accessed 9 Oct 2019

4. Collectif INSERM. Pesticides : Effets sur la santé, une expertise collective de l'Inserm. Salle Presse Inserm. 2013. http://presse. inserm.fr/pesticides-effets-sur-la-sante-une-expertise-collective- de-linserm/8463/

5. Garnett T (2013) Food sustainability: problems, perspectives and solutions. Proc Nutr Soc 72:29–39. https://doi.org/10.1017/S0029665112002947

6. Lukowicz C, Ellero-Simatos S, Régnier M, et al (2018) Metabolic Effects of a Chronic Dietary Exposure to a Low-Dose Pesticide Cocktail in Mice: Sexual Dimorphism and Role of the Constitutive Androstane Receptor. Environ Health Perspect 126:067007. https://doi.org/10.1289/EHP2877

7. Mnif W, Hassine AIH, Bouaziz A, et al (2011) Effect of Endocrine Disruptor Pesticides: A Review. Int J Environ Res Public Health 8:2265–2303. https://doi.org/10.3390/ijerph8062265

8. Mostafalou S, Abdollahi M (2013) Pesticides and human chronic diseases: evidences, mechanisms, and perspectives. Toxicol Appl Pharmacol 268:157–177. https://doi.org/10.1016/j.taap.2013.01.025

9. van der Mark Marianne, Brouwer Maartje, Kromhout Hans, et al (2012) Is Pesticide Use Related to Parkinson Disease? Some Clues to Heterogeneity in Study Results. Environ Health Perspect 120:340–347. https://doi.org/10.1289/ehp.1103881

10. Ntzani EE, G CMN, Evangelou E, Tzoulaki I (2013) Literature review on epidemiological studies linking exposure to pesticides and health effects. EFSA Support Publ 10:497E. https://doi.org/10.2903/sp.efsa.2013.EN-497

11. Nougadère A, Sirot V, Kadar A, et al (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. https://doi.org/10.1016/j.envint.2012.02.001

12. EU Pesticides database - European Commission. https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=activesubstance.selection&language=EN. Accessed 27 Sep 2019

13. Crépet A, Tressou J, Graillot V, et al (2013) Identification of the main pesticide residue mixtures to which the French population is exposed. Environ Res 126:125–133. https://doi.org/10.1016/j.envres.2013.03.008

14. Traoré T, Béchaux C, Sirot V, Crépet A (2016) To which chemical mixtures is the French population exposed? Mixture identification from the second French Total Diet Study. Food Chem Toxicol 98:179–188. https://doi.org/10.1016/j.fct.2016.10.028

15. (2017) The 2015 European Union report on pesticide residues in food. In: Eur. Food Saf. Auth. https://www.efsa.europa.eu/en/efsajournal/pub/4791. Accessed 28 Jun 2019

16. INAO, ITAB (2018) Guide des produits de protection des cultures utilisables en France en Agriculture Biologique

17. Hercberg S, Castetbon K, Czernichow S, et al (2010) The Nutrinet-Santé Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. BMC Public Health 10:242. https://doi.org/10.1186/1471-2458-10-242

18. Touvier M, Méjean C, Kesse-Guyot E, et al (2010) Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. Eur J Epidemiol 25:287–296. https://doi.org/10.1007/s10654-010-9433-9

19. Touvier M, Kesse-Guyot E, Méjean C, et al (2011) Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. Br J Nutr 105:1055–1064. https://doi.org/10.1017/S0007114510004617

20. Vergnaud A-C, Touvier M, Méjean C, et al (2011) Agreement between web-based and paper versions of a socio-demographic questionnaire in the NutriNet-Santé study. Int J Public Health 56:407–417. https://doi.org/10.1007/s00038-011-0257-5

21. Kesse-Guyot E, Castetbon K, Touvier M, et al (2010) Relative validity and reproducibility of a food frequency questionnaire designed for French adults. Ann Nutr Metab 57:153–162. https://doi.org/10.1159/000321680

22. Baudry J, Méjean C, Allès B, et al (2015) Contribution of Organic Food to the Diet in a Large Sample of French Adults (the NutriNet-Santé Cohort Study). Nutrients 7:8615–8632. https://doi.org/10.3390/nu7105417

23. Etude Nutrinet-Santé (2013) Table de Composition des Aliments de l'étude Nutrinet-Santé (Nutrinet-Santé Study Food Composition Database). Paris: Economica.

24. Verger EO, Mariotti F, Holmes BA, et al (2012) Evaluation of a Diet Quality Index Based on the Probability of Adequate Nutrient Intake (PANDiet) Using National French and US Dietary Surveys. PLoS ONE 7:. https://doi.org/10.1371/journal.pone.0042155

25. De Gavelle E, Huneau J-F, Mariotti F (2018) Patterns of Protein Food Intake Are Associated with Nutrient Adequacy in the General French Adult Population. Nutrients 10:226. https://doi.org/10.3390/nu10020226

26. Martínez-González MA, Sánchez-Tainta A, Corella D, et al (2014) A provegetarian food pattern and reduction in total mortality in the Prevención con Dieta Mediterránea (PREDIMED) study. Am J Clin Nutr 100 Suppl 1:320S–8S. https://doi.org/10.3945/ajcn.113.071431

27. UA-BW | CVUA Stuttgart (Startseite). http://www.cvuas.de/pub/default.asp?subid=1. Accessed 29 May 2019

28. (2017) The 2015 European Union report on pesticide residues in food. EFSA J 15:e04791. https://doi.org/10.2903/j.efsa.2017.4791

29. (2016) The 2014 European Union Report on Pesticide Residues in Food. EFSA J 14:e04611. https://doi.org/10.2903/j.efsa.2016.4611

30. Baudry J, Pointereau P, Seconda L, et al (2019) Improvement of diet sustainability with increased level of organic food in the diet: findings from the BioNutriNet cohort. Am J Clin Nutr 109:1173–1188. https://doi.org/10.1093/ajcn/nqy361

31. Bognár A (2002) Tables on weight yield of food and retention factors of food constituents for the calculation of nutrient composition of cooked foods (dishes). BFE Karlsruhe

32. Bergström L (1994) Nutrient losses and gains in the preparation of foods. Rapp-Livsmedelsverket Swed

33. GEMS/Food-Euro. Second workshop on reliable evaluation of low-level contamination of food. Report on a workshop in the frame of GEMS/Food-Euro, EUR/HFA target 22. 26-27 May 1995. Kulmbach, German. Geneva: WHO. 1995.

34. European Food Safety Authority. Management of left-censored data in dietary exposure assessment of chemical substances: Management of left-censored data in dietary exposure assessment of chemical substances. EFSA J. 2010;8:1557.

35. Baudry J, Méjean C, Allès B, et al (2015) Contribution of Organic Food to the Diet in a Large Sample of French Adults (the NutriNet-Santé Cohort Study). Nutrients 7:8615–8632. https://doi.org/10.3390/nu7105417

36. La macro SAS CALMAR | Insee. https://www.insee.fr/fr/information/2021902. Accessed 28 Jun 2019

37. INSEE. French national census data. [Internet]. 2009. Available from: http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/popul ations-legales/default.asp?annee=2009.

38. Lee DD, Seung HS Algorithms for Non-negative Matrix Factorization. 7

39. 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. https://doi.org/10.1111/j.1541-0420.2011.01588.x

40. Khattree R, Naik DN (2000) Multivariate Data Reduction and Discrimination with SAS Software, 1 edition. Wiley-SAS, Cary, N.C

41. Gaujoux R, Seoighe C (2010) A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11:367. https://doi.org/10.1186/1471-2105-11-367

42. Willett W, Stampfer MJ (1986) TOTAL ENERGY INTAKE: IMPLICATIONS FOR EPIDEMIOLOGIC ANALYSES. Am J Epidemiol 124:17–27. https://doi.org/10.1093/oxfordjournals.aje.a114366

43. Dolnicar S (2002) A review of unquestioned standards in using cluster analysis for data-driven market segmentation

44. Béchaux C, Zetlaoui M, Tressou J, et al (2013) Identification of pesticide mixtures and connection between combined exposure and diet. Food Chem Toxicol 59:191–198. https://doi.org/10.1016/j.fct.2013.06.006

45. Curl CL, Beresford SAA, Fenske RA, et al (2015) Estimating Pesticide Exposure from Dietary Intake and Organic Food Choices: The Multi-Ethnic Study of Atherosclerosis (MESA). Environ Health Perspect 123:475–483. https://doi.org/10.1289/ehp.1408197

46. Barański M, Średnicka-Tober D, Volakakis N, et al (2014) Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta-analyses. Br J Nutr 112:794–811. https://doi.org/10.1017/S0007114514001366

47. Baudry J, Debrauwer L, Durand G, et al (2018) Urinary pesticide concentrations in French adults with low and high organic food consumption: results from the general population-based NutriNet-Santé. J Expo Sci Environ Epidemiol. https://doi.org/10.1038/s41370-018-0062-9

48. Andreeva VA, Salanave B, Castetbon K, et al (2015) Comparison of the sociodemographic characteristics of the large NutriNet-Santé e-cohort with French Census data: the issue of volunteer bias revisited. J Epidemiol Community Health 69:893–898. https://doi.org/10.1136/jech-2014-205263

49. Boizot-Szantai C, Hamza O, Soler L-G (2017) Organic consumption and diet choice: An analysis based on food purchase data in France. Appetite 117:17–28. https://doi.org/10.1016/j.appet.2017.06.003

50. Agence Bio (2015) Baromètre de Consommation et de Perception des Produits Biologiques en France.

51. Valcke M, Bourgault M-H, Rochette L, et al (2017) Human health risk assessment on the consumption of fruits and vegetables containing residual pesticides: A cancer and non-cancer risk/benefit perspective. Environ Int 108:63–74. https://doi.org/10.1016/j.envint.2017.07.023

52. Reiss R, Johnston J, Tucker K, et al (2012) Estimation of cancer risks and benefits associated with a potential increased consumption of fruits and vegetables. Food Chem Toxicol 50:4421–4427. https://doi.org/10.1016/j.fct.2012.08.055

53. Arena M, Auteri D, Barmaz S, et al (2018) Peer review of the pesticide risk assessment of the active substance copper compounds copper(I), copper(II) variants namely copper hydroxide, copper oxychloride, tribasic copper sulfate, copper(I) oxide, Bordeaux mixture. EFSA J 16:e05152. https://doi.org/10.2903/j.efsa.2018.5152

54. Yigit N, Velioglu YS (2019) Effects of processing and storage on pesticide residues in foods. Crit Rev Food Sci Nutr

Tables :

Table 1: Characteristics of the participants (weighted), NutriNet Santé Study, 2014 (N=34,193)

	Whole Population	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6[1]	P-value[2]
Unweighted, ***n***	34,193	16,922	9,328	684	3,471	3,772	*16*	
Weighted, **%** of whole sample		50.70	27.58	2.20	8.27	11.20	*0.05*	
Age, years	50.1 (16.1)	48.6 (16.0)	51.0 (16.5)	52.8 (17.0)	52.4 (13.1)	52.1 (17.0)	*56.94 (17.2)*	<0.0001[2]
Sex								<0.0001
Women, %	75.5	71.40	75.74	86.71	87.08	82.74	*96.48*	
Monthly income per household unit, %								<0.0001
<€1200	18.80	19.30	18.04	19.59	17.27	19.21	*2.84*	
€1200-1800	28.70	28.7	29.6	33.52	26.65	27.34	*51.03*	
€1800-2700	26.10	26.48	25.71	20.77	26.81	25.63	*25.21*	
>€2700	18.70	18.21	19.15	17.31	20.66	18.26	*11.77*	
Unwilling to answer	7.80	7.31	7.51	8.80	8.61	9.56	*9.16*	
Occupational status, %								<0.0001
Retired	32.10	28.25	35.77	38.06	33.31	38.47	*65.84*	
Employee, manual worker	24.10	24.8	23.36	24.85	22.58	23.5	*2.03*	
Intermediate profession	15.60	17.12	14.86	9.82	14.62	12.72	*5.65*	
Managerial staff, intellectual profession	11.10	12.58	9.93	6.41	11.22	6.97	*7.99*	
Unemployed	9.80	10.03	8.83	12.47	11.61	9.02	*7.01*	
Never employed (students)	4.70	4.1	5.16	6.47	2.87	7.23	*11.48*	
Self-employed, farmer	2.80	3.11	2.1	1.91	3.79	2.09	*0*	
Educational level, %								<0.0001
Less than high-school diploma	17.80	17.72	17.95	19.56	16.46	18.68	*0*	
High school diploma	48.80	46.34	50.75	56.87	49.11	53.04	*61.11*	
Post Graduate	33.40	35.94	31.3	23.56	34.43	28.28	*38.89*	
Physical activity, %								<0.0001
High	33.10	30.44	33.16	43.81	41.39	36.32	*59.76*	
Moderate	34.20	34.96	34.40	28.86	31.04	33.48	*9.60*	
Low	19.90	22.56	18.49	14.06	15.35	16.11	*9.16*	
Missing data	12.80	12.04	13.95	13.27	12.21	14.08	*21.48*	
Body Mass Index (kg/m^2)	24.40 (4.90)	24.95 (5.29)	24.45 (4.63)	22.33 (3.95)	22.98 (3.52)	23.46 (4.09)	*22.45 (3.00)*	
≥25, %	36.10	40.27	36 .91	18.97	23.03	27.84	*38.76*	<0.0001
Place of residence, %								<0.0001
Rural community	24.70	25.25	23.71	22.59	29.07	22.24	*2.84*	
Urban unit with a population <20,000 inhabitants	16.20	16.23	16.52	13.44	17.37	14.46	*65.95*	
Urban unit with a population between 20,000 and 200,000 inhabitants	17.90	18.12	17.32	19.75	16.22	19.41	*3.55*	
Urban unit with a population >200,000 inhabitants	41.20	40.40	42.45	44.22	37.33	43.89	*27.66*	
Smoking habits, %								<0.0001
Never smoker	50.40	49.39	51.08	48.27	49.18	54.77	*50.58*	
Former smoker	37.50	37.12	37.93	39.31	40.38	36.14	*12.68*	
Current smoker	12.10	13.49	10.99	12.41	10.44	9.09	*36.74*	
Ethanol, grams/day	7.50 (12.10)	7.54 (12.05)	7.81 (11.77)	6.19 (14.65)	6.97 (8.74)	6.77 (14.75)	*2.82 (5.23)*	<0.0001[2]

[1]:The results for this group should be treated with caution as the small sample size can lead to unreliable estimates [2]: p-values for comparisons between five first clusters using Chi-square tests [3] : p-values for comparisons between five first clusters using linear regression

Table 2: Relative mean differences[1] (compared to whole sample mean) for pesticide dietary exposure, lower-bound scenario, NutriNet Santé Study, 2014 (N=34,193)

Pesticide active substances	**Cluster 1** (50.70%)[2]	**Cluster 2** (27.58%)[2]	**Cluster 3** (2.20%)[2]	**Cluster 4** (8.27%)[2]	**Cluster 5** (11.20%)[2]	**Cluster 6[3]** (0.05%)[2]
Acetamiprid	-0.45	-0.03	1.00	1.95	0.47	*1.92*
Anthraquinone	-0.06	0.05	0.20	0.03	0.09	*-0.21*
Azadirachtin	-0.09	-0.11	0.17	0.90	-0.03	*0.54*
Azoxystrobin	-0.43	0.33	2.33	-0.25	0.83	*5.41*
Boscalid	-0.36	0.29	1.53	-0.11	0.67	*3.30*
Carbendazim	-0.38	-0.01	0.91	1.49	0.43	*1.76*
Chlorpropham	-0.16	0.22	0.60	-0.31	0.30	*0.35*
Chlorpyrifos	-0.50	0.12	2.69	0.48	1.06	*10.50*
Lambda Cyhalothrin	-0.39	0.27	1.80	-0.05	0.76	*5.72*
Cypermethrin	-0.43	-0.07	0.97	1.99	0.45	*1.94*
Cyprodinil	-0.38	0.31	1.62	-0.14	0.73	*3.10*
Difenoconazole	-0.37	0.21	1.43	0.31	0.61	*3.30*
Dimethoate Ometoate	-0.46	0.05	1.43	1.49	0.57	*2.73*
Fenhexamid	-0.44	0.38	2.01	-0.16	0.79	*1.35*
Glyphosate	-0.21	0.16	1.03	-0.09	0.38	*1.85*
Imazalil	-0.68	0.23	4.74	-0.46	1.82	*21.15*
Imidacloprid	-0.43	0.27	1.38	0.38	0.74	*1.50*
Iprodione	-0.43	0.34	1.89	-0.15	0.83	*7.06*
Malathion	-0.41	0.18	2.79	-0.14	0.95	*10.15*
Methamidophos	-0.12	0.14	0.31	-0.15	0.25	*-0.12*
Profenofos	-0.66	0.33	3.97	-0.45	1.67	*14.92*
Pyrethrins	-0.07	0.01	0.10	0.24	0.08	*-0.48*
Spinosad	-0.27	-0.27	0.24	2.51	-0.04	*1.12*
Tebuconazole	-0.43	0.33	1.99	-0.15	0.84	*6.23*
Thiabendazole	-0.64	0.30	4.14	-0.43	1.60	*15.78*

[1]: Relative mean differences were calculated as follows : $(mean_{cluster} - mean_{whole\ sample})/mean_{whole\ sample}$

[2] : Weighted percent of whole sample

[3]:The results for this group should be treated with caution as the small sample size can lead to unreliable estimates

Table 3: Nutritional characteristics of participants' diets (weighted), NutriNet Santé study, 2014 (n=34,193)

	Cluster 1 (50.70%)[1] **Mean (SD)**	**Cluster 2** (27.58%)[1] **Mean (SD)**	**Cluster 3** (2.20%)[1] **Mean (SD)**	**Cluster 4** (8.27%)[1] **Mean (SD)**	**Cluster 5** (11.20%)[1] **Mean (SD)**	**Cluster 6[2]** (0.05%)[1] **Mean (SD)**
Energy intake without alcohol (kcal/day)	1845.8 (621.3)	2053.4 (642.2)	2297.1 (716.5)	1972.2 (549.4)	2125.3 (648.3)	*2833.5 (610.1)*
PANDiet Score(/100)	63.7 (8.0)	64.6 (8.1)	66.3 (8.7)	66.6 (7.2)	65.0 (8.3)	*70.5 (6.5)*
Provegetarian Score	35.5 (6.2)	35.5 (5.8)	37.4 (6.3)	37.9 (5.9)	35.7 (5.9)	*37.2 (7.1)*
Proportion of organic food in the diet,%	23.3 (22.5)	14.4 (15.2)	11.7 (14.4)	31.5 (21.6)	12.1 (14.8)	*9.2 (12.9)*
Proportion of organic food for plant-based food groups, %	41.0 (31.6)	23.0 (21.5)	15.0 (17.7)	49.9 (31.4)	18.6 (19.6)	*14.3 (17.2)*
Proportion of individuals with organic food in the diet ≥ 50%	14.7	3.2	1.8	23.1	2.2	*2.3*
Special diet, %						
Omnivorous	94.4	96.5	91.3	90.1	95.0	*87.7*
Pesco-vegetarian diet	1.8	1.4	2.6	3.9	1.7	*0.00*
Vegan diet	1.5	0.8	3.2	3.3	1.7	*11.5*
Vegetarian diet	2.3	1.3	2.9	2.7	1.6	*0.8*
Protein types						
Animal / Vegetal protein ratio	2.6 (1.7)	2.6 (1.4)	2.0 (1.2)	2.0 (1.2)	2.4 (1.3)	*1.3 (1.0)*
Vegetal / total protein ratio	0.3 (0.2)	0.3 (0.1)	0.4 (0.2)	0.4 (0.2)	0.3 (0.2)	*0.5 (0.2)*
Lipids (% of alcohol-free energy intake)	41.8 (7.3)	39.9 (9.6)	37.0 (7.8)	41.9 (5.1)	38.9 (1.7)	*27.0 (8.6)*
Carbohydrates (% of alcohol free energy intake)	39.1 (7.8)	40.9 (7.5)	45.2 (9.5)	39.6 (7.6)	42.2 (7.7)	*57.2 (11.7)*
Protein (% of alcohol-free energy intake)	18.9 (4.0)	18.8 (3.7)	17.2 (3.8)	18.1 (3.6)	18.3 (3.6)	*14.6 (4.1)*

[1] : Weighted percent of whole sample
[2]:The results for this group should be treated with caution as the small sample size can lead to unreliable estimates
SD : Standard Deviation
All p-values for comparisons between 5 first clusters using linear regression were <0.0001

Table 4: Daily intakes of sixteen food groups (weighted), NutriNet-Santé study, 2014 (N=34,193)

Food groups (grams/day)	Cluster 1 (50.70%)[1] Mean[2] (CI)	Cluster 2 (27.58%)[1] Mean[2] (CI)	Cluster 3 (2.20%)[1] Mean[2] (CI)	Cluster 4 (8.27%)[1] Mean[2] (CI)	Cluster 5 (11.20%)[1] Mean[2] (CI)	Cluster 6[3] (0.05%)[1] Mean[2] (CI)
Alcoholic beverages	94.5 (92.4, 96.8)	90.3 (87.4, 93.3)	63.1 (52.6,73.5)	82.1 (76.7, 87.5)	76.3 (71.7, 80.9)	*8.7 (-63.8;81.2)*
Other fats (including mayonnaise, fresh cream, vegetable fresh cream)	3.7 (3.7, 3.8)	3.5 (3.4, 3.6)	3.0 (2.7,3.4)	3.6 (3.4, 3.8)	3.1 (2.9, 3.2)	*0.4 (-2.2; 3.0)*
Non-alcoholic drinks	1675.6 (1663.9, 1687.3)	1706.2 (1690.3, 1722.1)	1838.1 (1781.9, 1894.3)	2082.6 (2053.7, 2111.5)	1758.4 (1733.5, 1783.4)	*1794.1 (1405.0; 2183.1)*
Butter	6.9 (6.7, 7.1)	6.6 (6.4, 6.7)	4.8 (4.3, 5.3)	5.7 (5.5, 6.0)	6.1 (5.9, 6.3)	*1.7 (-1.7; 5.1)*
Whole-grain products	54.4 (53.4,55.4)	45.2 (43.8, 46.5)	42.9 (38.1,47.7)	68.3 (65.8,70.8)	45.5 (43.3,47.6)	*63.2 (29.9;96.5)*
Extra food (including snacks, chips, salted biscuits, dressing and sauces)	17.9 (17.7,18.1)	16.1 (15.7,16.4)	18.2 (17.0,19.4)	17.8 (17.2,18.4)	15.8 (15.3,16.4)	*7.0 (-1.0;15.0)*
Fruit and vegetables (including juices and soups)	571.1 (565.7, 576.6)	784.9 (777.5, 792.4)	1307.8 (1281.5,1334.0)	895.9 (882.4, 909.4)	962.3 (950.6, 974.0)	*2955.9 (2774.0;3137.8)*
Starchy foods	186.3 (184.8, 187.8)	174.4 (172.4, 176.4)	152.7 (145.6, 159.7)	163.8 (160.2, 167.5)	158.1 (155.0, 161.3)	*52.9 (4.0; 101.7)*
Oil	18.8 (18.6,19.1)	18.0 (17.7,18.3)	19.8 (18.7,20.9)	24.4 (23.8, 24.9)	18.5 (18.0,18.9)	*10.1 (2.7; 17.5)*
Sweetened foods	74.5 (73.7,75.2)	70.7 (69.7, 71.8)	54.0 (50.2,57.7)	60.9 (59.0,62.8)	65.3 (63.7, 67.0)	*35.6 (9.7; 61.5)*
Fast food	36.7 (36.1,37.3)	33.0 (32.2,33.8)	21.2 (18.4,24.0)	27.2 (25.7,28.6)	29.7 (28.5, 30.9)	*0.5 (-18.7; 19.7)*
Seafood	40.9 (40.3,41.6)	44.0 (43.1,44.8)	47.3 (44.3,50.3)	45.3 (43.7,46.8)	44.0 (42.6, 45.3)	*10.5 (-10.6; 31.5)*
Dairy products	251.7 (248.5, 254.8)	280.1 (275.9, 284.4)	265.1 (250.0, 280.3)	233.5 (225.8, 241.3)	287.4 (280.7, 294.1)	*384.5 (279.7; 489.3)*
Eggs	11.1 (10.8,11.2)	11.5 (11.3,11.8)	11.8 (10.9,12.7)	11.5 (11.1, 12.0)	11.9 (11.5,12.3)	*11.8 (5.5; 18.1)*
Soy-based products	36.0 (34.5,37.6)	21.9 (19.8,24.0)	32.1 (24.9,39.4)	53.7 (49.9,57.4)	30.9 (27.7,34.1)	*45.9 (-4.5; 96.4)*
Meat, poultry, processed meats	121.0 (119.8, 122.1)	119.8 (118.2, 121.4)	75.9 (70.4, 81.5)	93.5 (90.6,96.3)	108.2 (105.7, 110.6)	*2.1 (-36.4; 40.5)*
Percentage of organic food for 16 food groups (%)						
Alcoholic beverages	17 (17, 17)	10 (9, 10)	6 (4, 7)	20 (20, 21)	7 (6, 8)	*-1 (-10; 10)*
Other fats (including mayonnaise, fresh cream, vegetable fresh cream)	28 (28, 29)	16 (16, 17)	12 (9, 14)	35 (34, 36)	13 (12, 14)	*10 (-10; 30)*
Non-alcoholic drinks	13 (13, 14)	8 (8, 8)	8 (7, 9)	17 (16, 18)	7 (6, 7)	*2 (-10; 10)*
Butter	29 (29, 30)	17 (17, 18)	10 (7, 12)	32 (31, 34)	14 (13, 15)	*12 (-10; 30)*
Whole-grain products	34 (33, 34)	21 (21, 22)	16 (14, 19)	42 (41, 44)	18 (17, 19)	*14 (0; 30)*
Extra food (including snacks, chips, salted biscuits, dressing and sauces)	25 (25, 26)	13 (13, 14)	10 (8, 12)	33 (32, 34)	11 (10, 12)	*6 (-10; 20)*
Fruit and vegetables (including juices and soups)	39 (39, 39)	21 (20, 21)	12 (10, 14)	51 (50, 53)	15 (14, 16)	*7 (-10; 20)*
Starchy foods	32 (31, 32)	19 (18, 19)	16 (14, 18)	42 (41, 43)	15 (14, 16)	*38 (20; 50)*
Oil	41 (40, 41)	26 (26, 27)	21 (18, 24)	49 (48, 51)	22 (20, 23)	*7 (-10; 30)*
Sweetened foods	27 (27, 28)	16 (16, 17)	14 (12, 16)	36 (35, 37)	15 (14, 16)	*16 (10;30)*
Fast food	20 (20, 21)	11 (10, 11)	6 (4, 7)	25 (24, 26)	9 (8, 10)	*7 (-10;20)*
Seafood	17 (17, 17)	11 (11, 12)	6 (5, 8)	19 (18, 20)	9 (8, 9)	*0 (-10;10)*
Dairy products	29 (28, 29)	17 (17, 18)	11 (9, 13)	33 (32, 34)	14 (14, 15)	*4 (-10; 20)*
Eggs	53 (52, 53)	43 (42,43)	31 (28, 34)	62 (61, 64)	36 (35,38)	*43 (20; 60)*
Soy	30 (29, 30)	19 (18, 20)	21 (18, 24)	38 (37, 40)	18 (16,19)	*53 (30; 70)*
Meat, poultry, processed meats	23 (23, 23)	15 (14,15)	8 (6, 10)	28 (27,28)	11 (10,12)	*2 (-10; 10)*

[1] : Weighted percent of whole sample / [2]: Adjusted means for energy-intake / [3] :The results for this group should be treated with caution as the small sample size can lead to unreliable estimates / Abbreviations: CI: Confidence Interval; MUFA: Mono-Unsaturated Fatty Acids; PUFA: Poly-Unsaturated Fatty Acids; SFA: Saturated Fatty Acids. All p-values for comparisons between 5 first clusters (using ANCOVA) were <0.0001.

Appendix A: Detailed Explanation for the Non-Negative Matrix Factorization procedure

Non-negative Matrix Factorization is a non-supervised data decomposition method, proposed by Lee to deal with non-negative data using non-negativity constraints.

This method is relevant for non-negative data with excess zeros and measurement error such as exposure to pesticides constrained by the detection limits of dosing techniques.

The purpose of NMF is to explain observed data through a limited number of components approximating the original data as accurately as possible.

The matrix representing the basis components and the matrix of mixture coefficients are constrained to have non-negative values, and no orthogonality or independence constraints are imposed on the basis components.

Let X be a matrix ($n \times p$) containing only non-negative values and without a row or column containing only 0 and r a relatively small integer $< n$ and $< p$.

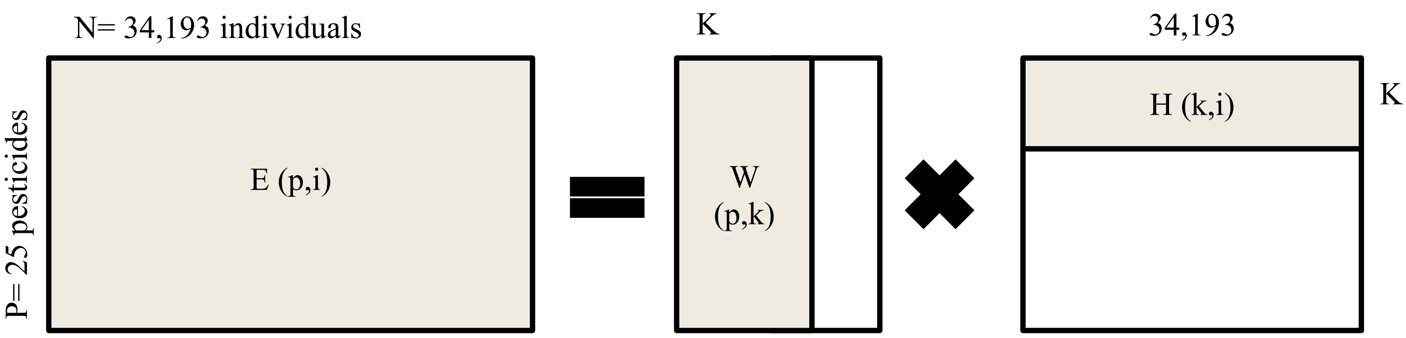
The non-negative factorization of matrix X is the search for two matrices W ($n\times r$) and H ($r \times p$) containing only positive or zero values and whose product approaches X so that $X \approx WH$.

The factorization is solved by searching for a local optimum of the optimization problem:

$\min W , H \geq 0[L (X , WH)]$

L is a loss function measuring approximation quality. Since the objective is usually to reduce the dimension of the original data, the factorization rank r is in practice often chosen such that $r << \min(n, p)$. This equation is solved by a multiplicative algorithm based on a gradient descent approach.

In this study of dietary pesticide profiles, W would be the total dietary exposure to the 25 selected pesticides (previously obtained after combining contamination values for each food and foods consumed by each participant) and H the number of individuals.



K, the number of NMF Components

adapted from Zetlaoui et al., 2011

Appendix B: Flowchart for the decomposition of ingredients and matching

Selected Ingredients included in the pesticides database :