



HAL
open science

Combining data from heterogeneous surveys for aggregate exposure Application to children exposure to lead in France

Marie Vanacker, Jessica Tressou, Guillaume Perouel, Philippe Glorennec,
Amelie Crépet

► **To cite this version:**

Marie Vanacker, Jessica Tressou, Guillaume Perouel, Philippe Glorennec, Amelie Crépet. Combining data from heterogeneous surveys for aggregate exposure Application to children exposure to lead in France. *Environmental Research*, 2020, 182, pp.109069. 10.1016/j.envres.2019.109069 . hal-02441217

HAL Id: hal-02441217

<https://univ-rennes.hal.science/hal-02441217>

Submitted on 7 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Combining data from heterogeneous surveys for aggregate exposure: application to**
2 **children exposure to lead in France**

3
4

5 Marie Vanacker¹, Jessica Tressou², Guillaume Perouel¹, Philippe Glorennec³, Amélie Crépet^{1*}

6

7 ¹ ANSES, French Agency for Food, Environmental and Occupational Health & Safety, Risk
8 Assessment Department, 94701, Maisons-Alfort, France

9 ² UMR MIA-Paris, AgroParisTech, INRA, Paris-Saclay University, 75005 Paris

10 ³ Rennes University, EHESP, INSERM, IRSET (Research Institute for Environmental and
11 Occupational Health) - UMR_S 1085, F-35000 Rennes, France

12

13 *Corresponding author. tel: (+33) 1 49 77 26 78, email: amelie.crepet@anses.fr. Address:
14 ANSES, French Agency for Food, Environmental and Occupational Health & Safety, Risk
15 Assessment Department, 14 rue Pierre et Marie Curie, 94701, Maisons-Alfort, France

16 **Abstract**

17 To assess human health risks related to the environment, it is necessary to aggregate exposure
18 from multiple sources. The objective of this paper was to propose a relevant approach to
19 combine data from heterogeneous populations and methodologies. Five different methods
20 based on Monte-Carlo simulations were tested and compared. Differences were: taking into
21 account or not stratification variable, timeline to assign exposure factors and concentration
22 and way to account for concentration correlations. The methods were applied to estimate lead
23 exposure from food, dust, soil, air, and tap water of French children aged between six months
24 and three years old.

25 Comparing results' uncertainty, it is recommended to 1) select a reference population
26 representative of the target population, 2) select stratification variables to combine surveys,
27 and 3) simulate a new population by randomly sampling individuals in the reference
28 population and simultaneously assigning human exposure factors and environmental
29 concentrations from other surveys in integrating correlations (MC1S). No difference was
30 observed when taking into account correlations using vectors of deterministic data from one
31 survey or rank of correlations with the Iman-Conover method. Regardless the methods used to
32 combine data, dust was the main exposure source, followed by soil and in a less extent by
33 food. Exposures from air and tap water were found to be insignificant for most children.

34

35 **Highlights**

- 36 • Five calculation processes were tested to combine dietary and environmental surveys
- 37 • Resampling individuals and variables decrease the uncertainty
- 38 • Use stratification variables to combine surveys limits risk of error
- 39 • The tested methods to account for correlations between exposure factors gave similar
40 results
- 41 • Dust and soil were main exposure sources of children in France

42

43 **Keywords**

44 Environmental health; lead; public health; Monte Carlo simulations; risk assessment.

45

46 **Abbreviations**

47 BDQA: French database for air quality

48 BEBE-SFAE: French database of children dietary
49 BW: Body weight
50 DL: Dust load
51 MC: Monte Carlo
52 MC1: Monte Carlo in first step
53 MC1S: Monte Carlo in first step with the use of stratification variables
54 MC2: Monte Carlo in second step
55 MC2S: Monte Carlo in second step with the use of stratification variables
56 MCIC: Monte Carlo with the method of Iman and Conover
57 PH: French database for lead concentration in dwellings
58 UI: Uncertainty intervals

59

60 **Fundings**

61 This work took in account studies on human subjects, especially children. These studies were
62 approval by an appropriately constituted committee for human subjects.
63 This research did not receive any specific grant from funding agencies in the public,
64 commercial, or not-for-profit-sectors.

65 **1 Introduction**

66 Aggregate exposure can be defined as the sum of several sources of exposure (air, dust, food,
67 water, etc.) via different exposure routes (ingestion, inhalation and dermal absorption). To
68 perform human health risk assessments based on total exposure, it is important to consider
69 aggregate exposure. By nature, biomonitoring data considers aggregate exposure, but this data
70 does not enable the contribution of exposure sources to be assessed. In order to determine
71 management options to mitigate exposure and the associated risks, the identification of
72 exposure sources and factors contributing to total exposure is currently essential. However,
73 the modelling of aggregate exposure is complex from several standpoints. Except in specific
74 cases (Cao et al., 2016), complete surveys seldom collect all exposure sources for the same
75 individual. Therefore, performing aggregate exposure assessments often requires considering
76 data from different databases with different populations and methodologies. In addition,
77 exposure can be expressed at the individual level, which is generally the case for food, or at
78 the population level, which is more common for environmental exposure.

79 Even though the definition of “aggregate exposure” is soon to be standardised, some authors
80 have recently proposed “aggregate exposure” when considering only one source and one route
81 of exposure. This was the case of Cowan-Ellsberry and Robison (2009), Delmaar et al. (2015)
82 and Gosens et al. (2014) who studied dermal exposure to parabens and phthalates from
83 cosmetics. Other authors have tried to be as exhaustive as possible in the calculation of
84 aggregate exposure. For example, Beko et al. (2013) considered two sources of exposure (dust
85 and air) for phthalates and extrapolated dietary exposure from internal measurements.
86 Pelletier et al. (2017) took into account three routes of indoor residential exposure to semi-
87 volatile organic compounds. Furthermore, Glorennec et al. (2016) assessed aggregate
88 exposure to metals and metalloids in children between three and six years of age in France,
89 considering several sources of exposure in a Monte Carlo (MC) simulation via various
90 surveys. However, in these studies, the strategy used to combine data from different surveys
91 with different levels (individual vs population) to assess aggregate exposure was not
92 discussed. Some tools have been developed in the last few years to assess exposure via
93 different routes and sources. The European projects EuroMix, HBM4EU or SOLUTIONS
94 developed methodologies and guidance for assessing risks in mixture from combined
95 exposure to multiple chemicals for different regulatory sectors. In these projects, mixture risk
96 assessment is limited by the difficulties in considering aggregate exposures from different
97 sources with dietary source combined with non-dietary sources (Bopp et al., 2018). Exposure

98 tools addressing multiple exposure routes were embedded in the EuroMix toolbox, Monte
99 Carlo Risk Assessment software (van der Voet et al., 2015) by adding to diet exposure results
100 from MCRA to non-dietary exposure sources from other software as PACEM for personal
101 care products (Karrer et al., Submitted). Kennedy et al. (2012) developed the Bystander and
102 Residential Exposure Assessment Model (BREAM) to evaluate non-dietary exposure. They
103 then proposed options to develop an aggregate exposure model combining BREAM with the
104 MCRA platform (Kennedy et al., 2015a; Kennedy et al., 2015b). However, this work was
105 restricted to pesticide exposure via agricultural activities and thus only involved workers,
106 bystanders and residents living near agricultural areas. Two software applications for
107 Stochastic Human Exposure and Dose Simulation (SHEDS-Multimedia and SHEDS-
108 Residential) developed by the US. EPA only take into account specific scenarios of non-
109 dietary exposure with no link to dietary exposure. A new tool, SHEDS-High-Throughput
110 (SHEDS-HT), based on SHEDS-Multimedia, combines direct dermal exposure, inhalation
111 and accidental ingestion with the ingestion of food and drinking water by MC simulation
112 (Isaacs et al., 2014). However, SHEDS-HT mainly focused on aggregate exposure from diet
113 and consumer product sources, and the scenario of exposure sources by dust and soil is not
114 clearly developed. Moreover, individual intakes are specified for the American population
115 only. Thus, a harmonised consistent approach for aggregate exposure in case of different
116 sources of exposure is still lacking.

117 The aim of this work was to set out general principles for assessing aggregate exposure of a
118 target population from various sources (diet, dust, air, soil and tap water) when data come
119 from heterogeneous surveys. Five different calculation processes were tested and compared.
120 The different methods were based on the general principle which consists in creating a
121 simulated population from the individuals of the different surveys via MC simulations. MC
122 simulations are often used to combine risk assessment data (Kennedy et al., 2012; Kennedy et
123 al., 2015a; Kennedy et al., 2015b; Paustenbach, 2000; Safford et al., 2015; Zartarian et al.,
124 2017). They make it possible to draw random samples from distributions of datasets in order
125 to reconstruct the sources of exposure for each individual. These methods, which take into
126 account inter-individual variability as well as uncertainty, provide a more realistic estimate of
127 aggregate exposure for individuals across a population (Paustenbach, 2000).

128 Lead exposure for the target population of French children between the ages of six months
129 and three years was chosen as a case study to test these methods.

131 **2 Materials and Methods**

132 **2.1 Exposure factors**

133 **2.1.1 Food consumption (Q_{Food}) and quantities of ingested tap water (Q_{Water})**

134 Food consumption (Q_{Food}) and quantities of ingested tap water (Q_{Water}) for children were
135 evaluated in the national cross sectional survey named BEBE-SFAE (Fantino and Gourmet,
136 2008) which was conducted in France from January to March 2005 in the population of
137 children aged 15 days to three years. Individual consecutive three-day weighed food was
138 recorded in non-breastfed infants. More than 1260 food products specifically made for
139 toddlers and young children were notified in the database with 850 “specific baby foods”
140 (Fantino and Gourmet, 2008). To be representative of the child population, sampling weights
141 were assigned to each infant. A total of 706 children were recorded in BEBE-SFAE using
142 proportionate quota sampling based on the child’s age, the mother’s occupation and the
143 family’s socioeconomic strategy.

144

145 **2.1.2 Inhalation rates (IRs)**

146 Inhalation rates were evaluated based on the U.S. EPA recommendations in the Exposure
147 Factors Handbook (2011). For each age group (zero- to one-year old, one- to two-years old,
148 two- to three-years old), mean, 95th percentile and maximum inhalation rate values were
149 available. From these statistics, the mean and standard deviation of a normal truncated
150 distribution were adjusted for each age group.

151

152 **2.1.3 Dust loads (DLs)**

153 Dust loads were estimated from the publication by Giovannangelo et al. (2007) who studied
154 the distribution of dust loads collected from the floor in 46 German homes, 42 Dutch homes
155 and 34 Swedish homes. Since the data from Sweden were only collected from rugs, they were
156 excluded. The parameters of a truncated lognormal distribution were determined from the
157 logarithms of the geometric means and geometric standard deviations calculated from the
158 German and Dutch results weighted by the number of samples per country.

159

160 **2.1.4 Quantities of ingested soil (Q_{Soil}) and dust (Q_{Dust})**

161 Quantities of ingested soil (Q_{Soil}) and dust (Q_{Dust}) were derived from the Exposure Factors
162 Handbook (U.S. EPA, 2011) for children under the age of one year and between the ages of
163 one and three years. A truncated lognormal distribution of Q_{Soil} and Q_{Dust} , as proposed by
164 Özkaynak et al. (2011), was fitted.

165

166 **2.2 Lead contamination surveys**

167 Table 1 summarises the available data from the different surveys, the distributions used, and
168 descriptive statistics for lead concentrations for the various investigated exposure sources and
169 factors. A middle-bound scenario that consists in replacing values below the limit of detection
170 (LOD) or the limit of quantification (LOQ) with LOD/2 or LOQ/2 (EFSA, 2012) was used in
171 the case of censored data for concentrations in food, soil, dust and tap water.

172 *Table 1. Summary of input variables used for the calculation of aggregate exposure to lead for children aged six months to three years.*

| Input variables | | Age | References | Distribution | Mean | SD | Median | P95 | Min | Max |
|---|---------------------------|--------------------|--------------------------------|--------------|-------|-------|--------|-------|-------|-------|
| Concentration in food ($\mu\text{g}_{\text{Pb}}\cdot\text{kg}^{-1}$) | C_{Food} | 0 months – 3 years | Guerin et al. (2017) | Empirical* | | | | | | |
| Consumption of food ($\text{g}\cdot\text{d}^{-1}$) | Q_{Food} | 6 months – 3 years | Fantino and Gourmet (2008) | Empirical* | | | | | | |
| Dietary exposure ($\mu\text{g}_{\text{Pb}}\cdot\text{kg}_{\text{bw}}^{-1}\cdot\text{d}^{-1}$) | E_{Dietary} | 6 months – 3 years | BEBE-SFAE survey in this study | Empirical** | 0.208 | 0.095 | 0.194 | 0.381 | 0.020 | 0.632 |
| Body weight (kg) | BW | 6 months – 3 years | BEBE-SFAE survey in this study | Empirical | 10.4 | 2.6 | 10.0 | 15.0 | 3.4 | 20.0 |
| Consumed quantity of tap water ($\text{mL}\cdot\text{d}^{-1}$) | Q_{Water} | 6 months – 3 years | BEBE-SFAE survey in this study | Empirical | 65.3 | 159.8 | 0 | 250- | 0 | 47.0 |
| Inhalation rate ($\text{m}^3\cdot\text{d}^{-1}$) | IR | 6 - 12 months | U.S. EPA (2011) | Normal | 5.4 | 1.6 | - | 8.0 | 0 | 26.25 |
| | | 1 - 2 years | | Normal | 8.0 | 2.9 | - | 12.8 | 0 | 24.77 |
| | | 2 - 3 years | | Normal | 8.9 | 2.9 | - | 13.7 | 0 | 28.17 |
| Ingested soil ($\text{mg}\cdot\text{d}^{-1}$) | Q_{Soil} | 6 - 12 months | U.S. EPA (2011) | Lognormal | - | - | 30 | 200 | 0 | 1000 |
| | | 1 - 3 years | | Lognormal | - | - | 50 | 200 | 0 | 1000 |
| Ingested dust ($\text{mg}\cdot\text{d}^{-1}$) | Q_{Dust} | 6 - 12 months | U.S. EPA (2011) | Lognormal | - | - | 30 | 100 | 0 | 1000 |
| | | 1 - 3 years | | Lognormal | - | - | 60 | 100 | 0 | 1000 |
| Dust load ($\text{mg}\cdot\text{m}^{-2}$) | DL_{Germany} | - | Giovannangelo et al. (2007) | Lognormal*** | 194 | 4.1 | - | - | 0 | 2000 |
| | $DL_{\text{Netherlands}}$ | | | Lognormal*** | 151 | 5.5 | - | - | 0 | 2000 |
| Concentration in dust ($\mu\text{g}_{\text{Pb}}\cdot\text{m}^{-2}$) | C_{Dust} | - | PH survey in this study | Empirical | 25.4 | 61.8 | 9.0 | 109.6 | 1.00 | 694.8 |
| Concentration in tap water ($\mu\text{g}_{\text{Pb}}\cdot\text{L}^{-1}$) | C_{Water} | - | PH survey in this study | Empirical | 2.5 | 5.4 | 0.826 | 12.5 | 0 | 47.0 |
| Concentration in soil ($\mu\text{g}_{\text{Pb}}\cdot\text{g}^{-1}$) | C_{Soil} | - | PH survey in this study | Empirical | 70.7 | 106.8 | 34.8 | 273.4 | 2.4 | 830.9 |
| Concentration in air ($\mu\text{g}_{\text{Pb}}\cdot\text{m}^{-3}$) | C_{Air} | - | BDQA survey in this study | Empirical | 7.8 | 5.6 | 6.6 | 17.1 | 1.4 | 41.4 |

173 *Not given since C_{Food} and Q_{Food} mainly depend on the food.

174 **Individual dietary exposure, $E_{\text{i Dietary}}$, had previously been estimated in the report by ANSES (2014) and was included in the BEBE-SFAE

175 survey in this study by combining concentration data from the French infant Total Diet Study (Guerin et al., 2017) with the consumed quantities

176 and body weights available in the BEBE-SFAE survey (Fantino and Gourmet, 2008).

177 ***The geometric mean and geometric

178 **2.2.1 Food contamination by lead (C_{Food})**

179 Food contamination by lead (C_{Food}) was recorded in 2011, from the first infant Total Diet
180 Study (iTDS), conducted in non-breastfed children under three years of age (Hulin et al.,
181 2014). In iTDS, more than 500 chemical substances were analysed in foods. These included
182 substances naturally found in the environment and those originating in human activities (e.g.
183 industrial, agricultural, domestic, etc.). Food items were selected using the results of the
184 BEBE-SFAE survey, enabling home cooking practices to be considered. Overall, the iTDS
185 contained more than 5500 items as consumed foods, including foods such as vegetables, fruits
186 and cakes as well as specific children's food products. To limit censored data, a more
187 sensitive inductively coupled plasma mass spectrometry method was developed and validated
188 for lead in 291 samples (Guerin et al., 2017). With this method, the LOQ was 0.6 or 0.9
189 $\mu\text{g}_{\text{Pb}}\cdot\text{kg}^{-1}$ for solid and liquid samples respectively. Lead was detected in most samples, where
190 the highest concentrations were mainly found in foods containing chocolate, and a maximum
191 value of 16 $\mu\text{g}_{\text{Pb}}\cdot\text{kg}^{-1}$ was observed (Guerin et al., 2017).

192

193 **2.2.2 Child home and environmental contamination (C_{Dust} , C_{Soil} and C_{Water})**

194 The "Plomb-Habitat" (PH) survey recorded lead concentration data for tap water, soil and
195 dust in 472 homes of children aged from six months to six years in France between October
196 2008 and August 2009 (Glorennec et al., 2015; Lucas et al., 2012). Population sample weights
197 were available, to be representative of French dwellings. Lead concentrations in tap water
198 (C_{Water}) were measured in kitchens. The LOQ for lead in tap water was 1 $\mu\text{g}\cdot\text{L}^{-1}$. The average
199 lead load in dust (C_{Dust}) for each dwelling was evaluated in $\mu\text{g}\cdot\text{m}^{-2}$. The LOQ for lead in dust
200 was 2 $\mu\text{g}\cdot\text{m}^{-2}$ for total lead. For concentrations in soil (C_{Soil}), in cases of children playing
201 outside on soft ground (for 315 dwellings), samples were collected from the outdoor
202 playground. The LOQ for lead in soil was 1.3 $\mu\text{g}\cdot\text{g}^{-1}$ for total lead.

203

204 **2.2.3 Air contamination (C_{Air})**

205 Lead concentrations in outdoor air were collected from the regulatory monitoring network
206 (BDQA, the French database for air quality). Air quality monitoring has been implemented in
207 each France region in more than 650 rural, urban, suburban areas or linked to the traffic road
208 including more than 3 000 instruments. In this study, annual mean concentrations of lead in
209 outdoor air were considered from 2007 to 2011 no representative French survey exists on air

210 concentrations in inside dwellings, outdoor air lead concentration were used to estimate
 211 concentrations in indoor air of children's homes (C_{Air}). Data were not included when the
 212 measurements were too low to calculate the annual mean, specifically when annual coverage
 213 did not exceed 14% or when it exceeded 100%. Thus, a total of 176 measurements were
 214 considered in rural, urban and suburban areas.

215 2.3 Aggregate exposure model

216 Daily aggregate exposure to lead was calculated for each individual by combining exposure
 217 from the various sources (food, water, soil, dust and air) and via the various routes (ingestion
 218 and inhalation). Dermal lead exposure was considered as insignificant compared to the two
 219 other routes (EFSA, 2010). In the case of censored data, we applied a middle-bound scenario
 220 that consisted in replacing all values below the LOD and LOQ with either LOD/2 or LOQ/2.

$$221 \quad E_{i,Aggregate} = (E_{i,Dietary} + E_{i,Soil} + E_{i,Dust} + E_{i,Water}) \times \tau_{ingestion} + E_{i,Air} \times \tau_{inhalation} \quad (1)$$

222 To aggregate the various sources, absorption factors are commonly used for the two routes of
 223 exposure: ingestion ($\tau_{ingestion}$) and inhalation ($\tau_{inhalation}$). In this case study, the two absorption
 224 factors were equal to one.

$$225 \quad E_{i,Dietary} = \sum(Q_{i,Food} \times C_{Food}) / BW_i \quad (2)$$

226 $E_{i,Dietary}$ was the dietary exposure to lead of an individual i , expressed in $\mu\text{gPb.kg}_{bw}^{-1}.\text{d}^{-1}$ and
 227 was assessed by the sum of all products between $Q_{i,Food}$, the quantity of food consumed by
 228 individual i (g.day^{-1}), and C_{Food} , the associated level of lead contamination for the food
 229 ($\mu\text{gPb.g}^{-1}$ food). BW_i denoted the body weight of the individual i .

$$231 \quad E_{i,Water} = Q_{i,Water} \times C_{Water} / BW_i \quad (3)$$

232 $E_{i,Water}$ was the lead exposure via tap water of an individual i and was expressed in $\mu\text{gPb.kg}_{bw}^{-1}.\text{d}^{-1}$
 233 with the quantity consumed ($Q_{i,Water}$, in L.d^{-1}) and the level of tap-water contamination
 234 (C_{Water} , in $\mu\text{gPb.L}^{-1}$).

$$235 \quad E_{i,Soil} = Q_{i,Soil} \times C_{Soil} \times 10^3 / BW_i \quad (4)$$

236 $E_{i,Soil}$ was the lead exposure via soil of an individual i ($\mu\text{gPb.kg}_{bw}^{-1}.\text{d}^{-1}$) where $Q_{i,Soil}$ was the
 237 ingested quantity for the individual i ($\text{mg}_{Soil}.\text{d}^{-1}$) and C_{Soil} was the level of lead contamination
 238 in the soil ($\mu\text{gPb.g}^{-1}$ soil).

$$239 \quad E_{i,Dust} = (Q_{i,Dust} / DL) \times C_{Dust} / BW_i \quad (5)$$

240 $E_{i,Dust}$ was the lead exposure via dust of an individual i ($\mu\text{g}_{Pb} \cdot \text{kg}_{bw}^{-1} \cdot \text{d}^{-1}$) where $Q_{i,Dust}$ was the
241 quantity of dust ingested by the individual i ($\text{mg}_{Dust} \cdot \text{d}^{-1}$) and C_{Dust} was the level of lead
242 contamination in the dust ($\mu\text{g}_{Pb} \cdot \text{m}^{-2}$ dust). DL was a dust load factor expressed in $\text{mg} \cdot \text{m}^{-2}$.

$$243 \quad E_{i,Air} = IR_i \times C_{Air} / BW_i \quad (6)$$

244 $E_{i,Air}$ was the lead exposure via air of an individual i ($\mu\text{g}_{Pb} \cdot \text{kg}_{bw}^{-1} \cdot \text{d}^{-1}$) where IR_i was the
245 inhalation rate for the individual i ($\text{m}^3 \cdot \text{d}^{-1}$) and C_{Air} was the air concentration of lead ($\text{ng}_{Pb} \cdot \text{m}^{-3}$).
246

247 **2.4 Methods for combining surveys**

248 A three-step process was proposed to answer the following underlying questions (Fig. 1): (1)
249 How to choose a reference population? (2) Did the different surveys have common variables?
250 If common variables were observed, did they influence concentrations for the different
251 sources? Could these variables be considered as stratification variables and divided into
252 several classes to link the surveys to the reference population? (3) Did the surveys contain
253 sampling weights to correct for under- or over-represented individuals?

254 Regarding the answers to the above questions, five methods, namely MC1, MC2, MC1S,
255 MC2S and MCIC were tested and compared. These five different methods were based on the
256 general principle which consisted in creating a simulated population from the individuals of
257 the different surveys using second-order MC simulations (Fig. 1). Missing values, which are a
258 common issue in statistical analyses but are not specific to aggregate exposure, were treated
259 by replacement using the mean value.

260

261 **2.4.1 Step 1: reference population**

262 A reference population was defined as the most representative population of the target
263 population considering major characteristics (age, region, gender, etc.). Then, the
264 characteristics of the reference population had been reproduced in the simulated population.
265 The five tested methods had the same population of reference.

266

267 **2.4.2 Step 2: selection of stratification variables**

268 Stratification variables were defined as population characteristics (like age, sex, region, etc.)
269 that would be used to link the surveys. Stratification variables were selected in a two-step
270 process:

- 271 1. The first step was to identify sociodemographic variables shared between surveys that
272 could influence the input variables.
- 273 2. The second step was to test, via a statistical analysis, the significance of the
274 correlations between shared sociodemographic variables and the concentrations for
275 each survey. In case of significant correlations, it proves the importance to integrate
276 the stratification variables when sampling the concentration values in the different
277 surveys.

278 The impact of using or not stratification variables was tested in comparing the MC1S, MC2S
279 methods which included stratification variables with the MC1, MC2 methods which did not.
280 The MCIC did not include stratification variables.

281

282 **2.4.3 Steps 3&4: Monte Carlo simulation strategies**

283 The five methods used second-order MC simulations and integrated sampling weights. Values
284 of exposure factors (e.g. quantity of ingested soil/dust or inhalation rate) were randomly
285 selected based on the age of each individual from respective distributions presented in Table
286 1. Concentration values were assigned to each individual in the newly simulated population
287 using an observed value from the other surveys.

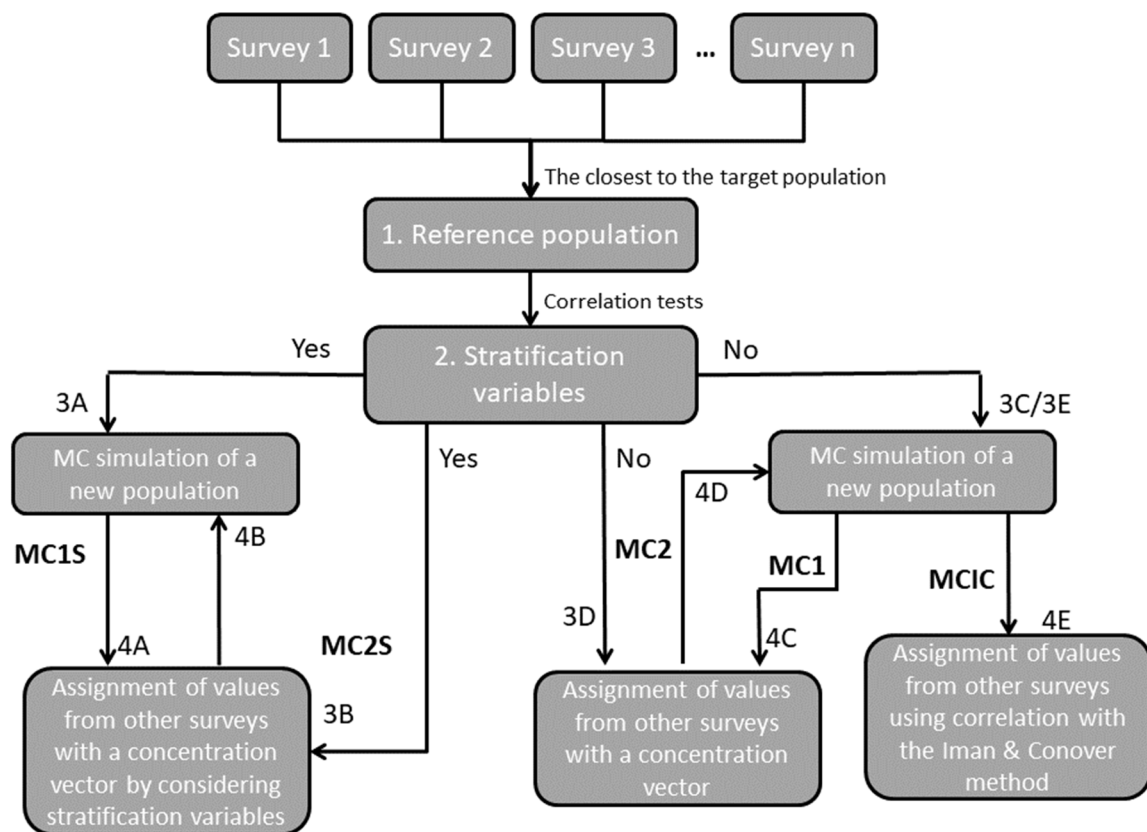
288 One major difference between the methods was the timeline of assigning exposure factors and
289 concentrations in the simulation process. The two MC1 and MC1S (MC in first step) methods
290 created a simulated population of 100,000 individuals taken randomly from the reference
291 population and assigned exposure factors and concentration values from the other surveys.
292 The MC2 and MC2S (MC in second step) methods first assigned exposure factors and
293 concentration values from the other surveys to each individual in the reference population.
294 Next, 100,000 individuals were randomly sampled from the combined data to create a
295 simulated population. The difference between MC1S/MC2S and MC1/MC2 was that MC1S
296 and MC2S used stratification variables to combine surveys, while this was not the case of
297 MC1 and MC2.

298 Furthermore, it is possible that surveys include several variables of interest, as for example
 299 concentrations of lead in tap water, in dust and in soil for one dwelling in the PH survey,
 300 which are highly correlated. In this case, the three concentrations need to be selected together
 301 to be assign to individuals in the simulated population. To keep correlations, one proposal
 302 applied in MC1S/MC1 and MC2S/MC2 methods was to select vectors of these three
 303 concentration variables.

304 Another method consisted in reproducing correlations between the concentration variables
 305 during simulation process This is proposed by the method named MCIC, based on the method
 306 of Iman and Conover (Iman and Conover, 1982) which used observed rank correlations and
 307 marginal distributions of concentrations.

308 To quantify the uncertainty associated with each method, the process was repeated 100 times.
 309 Thus, for each method, 100 simulated populations of 100,000 individuals were created, and an
 310 uncertainty interval was estimated.

311
 312



313
 314 Figure 1. Diagram of different methodologies for combining surveys in order to evaluate
 315 aggregate exposure. Step 1: Choice of reference population; Step 2: Choice of stratification
 316 variables; Step 3: Simulation of a new population (3A: MC1S, 3C: MC1, 3E: MCIC) or

317 assignment of values from other surveys with a concentration vector (3B: MC2, 3D: MC2S);
318 Step 4: Assignment of values from other surveys using a concentration vector (4A: MC1S,
319 4C: MC1) or reproducing correlation with the Iman and Conover method (4E: MCIC) or
320 simulation of a new population (4B: MC2S, 4D: MC2).

321

322 2.5 Computation with the R program

323 MC simulations, distribution fitting and statistical testing were performed in the R program (R
324 Core Team, 2017). The *msm* R package (Jackson, 2011) and the *EnvStat* R package (Millard,
325 2013) were used to adjust normal and lognormal truncated distributions, respectively. For the
326 selection of stratification variables, Wald tests were performed with univariate general linear
327 models by considering sampling weights using the *survey* R package (Lumley, 2004). The
328 MCIC method was implemented with the *mc2d* R package (Pouillot et al., 2016).

329 For each exposure source and each of the 100 new populations, descriptive statistics (mean,
330 median, standard deviation and percentiles) were estimated from population results. Mean
331 contributions to aggregate exposure for each exposure source were calculated using the mean
332 of the individual ratios for each exposure source and aggregate exposure multiplied by 100.
333 Contribution values were recorded for the 50%, 10% and 5% of individuals with the highest
334 aggregate exposure. From the 100 values calculated for each statistic, the median and the 2.5th
335 and 97.5th percentiles were displayed in result tables to give estimates with credible intervals
336 reflecting uncertainty. Significant differences were observed when confidence intervals did
337 not overlap.

338 Furthermore, to evaluate the sensitivity of the various parameters to the evaluation of
339 aggregate exposure, Spearman correlations were computed.

340

341 3 Results

342 3.1 Reference population

343 Five hundred eleven children aged six months to three years were selected from the BEBE-
344 SFAE survey, as well as 214 dwellings from the PH survey. Two hundred eleven samples for
345 lead contamination were recorded for tap water and dust, as well as 101 samples for soil. In
346 this case study, the BEBE-SFAE population was chosen as the reference population. Firstly, it

347 had the highest number of studied children. Secondly, children were the core issue in BEBE-
 348 SFAE survey, with physiological and sociodemographic data for each child. Thus, BEBE-
 349 SFAE was considered more representative than PH of the target population of children aged
 350 between 6 months and 3 years old in France. BDQA was not specific to a population group
 351 and thus could not be used as the reference population.
 352

353 3.2 Selection of stratification variables

354 The variables common to the BEBE-SFAE and PH surveys were age, gender and region. The
 355 only variable they had in common with BDQA was region. Two region classifications were
 356 studied: the first corresponded to the 22 administrative regions of France (before they were
 357 modified in 2016) and the second had five classes: Paris region, North-West, North-East,
 358 South-East, and South-West.

359 Table 2 shows p-value results for the weighted univariate general linear models with the
 360 *survey* package. Regarding the univariate analysis, a relationship clearly appeared between the
 361 22-class region variable and all input variables (p-value < 0.001***). A significant
 362 relationship with the five-class region variable was only observed with air concentrations.
 363 Age was significantly correlated with all factors except dust concentrations. The gender of the
 364 children did not appear to be related to any input variables.

365 Thus, it was decided to stratify the reference population according to the two common
 366 variables of the BEBE-SFAE and PH surveys, which were age and region (22-class regions).
 367 Stratification between BEBE-SFAE and BDQA was performed considering the 22-class
 368 regions. If a class of the stratification variable (age-region here) was present in the reference
 369 population but absent from PH or BDQA, it was decided to assign data from younger age
 370 within the same region or, if that was not possible, to take the mean of the whole population.
 371

372 Table 2. Results for the selection of the age, region (22-class and five-class) and gender
 373 stratification variables, for the input variables used in the various surveys. P-value results
 374 under 0.05 were considered significant and are notified in bold.

| Input variables | Survey | Sociodemographic parameters | | | |
|--------------------------------|-----------|-----------------------------|------------------|------------------|-------|
| | | Age | 22 Regions | 5 Regions | Sex |
| <i>Dietary exposure</i> | BEBE-SFAE | <0.001 | <0.001 | 0.218 | 0.204 |
| <i>Tap water concentration</i> | PH | 0.002 | <0.001 | 0.203 | 0.115 |
| <i>Dust concentration</i> | PH | 0.559 | <0.001 | 0.448 | 0.223 |
| <i>Soil concentration</i> | PH | 0.040 | <0.001 | <0.001 | 0.372 |

| | | | | | |
|--------------------------|------|---|--------|--------|---|
| <i>Air concentration</i> | BDQA | - | <0.001 | <0.001 | - |
|--------------------------|------|---|--------|--------|---|

375

376 3.3 Comparison of methods for combining surveys

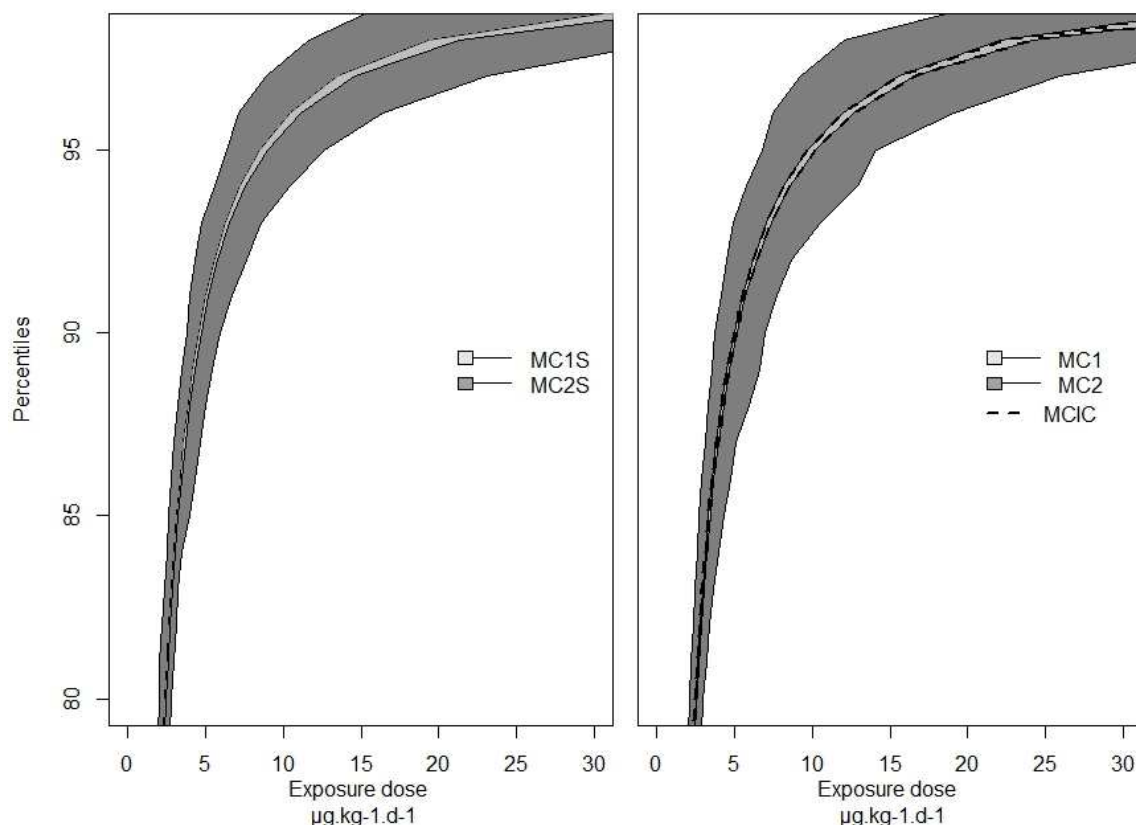
377 To simplify, details for the method comparisons were displayed only for aggregate
 378 exposure. Similar results were obtained for the various exposure sources with the exception of
 379 dietary exposure.

380

381 3.3.1 Monte Carlo in first step or in second step?

382 Regarding the median estimates at the P50 and P95 levels for aggregate exposure, the
 383 different methods produced values between 0.85 and 0.95 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ and between 8.7 and
 384 10 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$, respectively (Fig. 2, Table 3). The lowest standard deviation values were
 385 9.8 and 11 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$, meaning that variability was lower with method 2 (MC2S and
 386 MC2).

387 The results showed no significant difference between the two methods, as the uncertainty
 388 intervals (UIs) of the methods with MC in first step and MC in second step overlapped
 389 (MC1S vs MC2S, and MC1 vs MC2). However, it was observed that the MC2 methods had
 390 larger UIs, especially for highly exposed children (Fig. 2 and Fig. 3). At the 99th percentile,
 391 the upper bound of the UI was twice as high with MC in second step (greater than or equal to
 392 88 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for MC2S and MC2) than with MC in first step (around 40 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$
 393 for MC1S and MC1). This higher uncertainty could also be observed for the contribution of
 394 the various sources to aggregate exposure, especially for higher percentiles (Table 4). For
 395 example, for the 5% of children with the highest exposure levels from soil, the upper bound of
 396 contribution for MC2S reached 16% while it was 6% for the MC1S method.



397

398

399 Figure 2. Distributions of aggregate exposure ($\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$) to lead in children between the
 400 ages of six months and three years in France with methods using stratification variables (on
 401 the left) and methods without stratification (on the right). Ninety-five percent uncertainty
 402 intervals appear in grey.

403

404 3.3.2 Stratification or no stratification?

405 Significant differences were observed between results for aggregate exposure and the
 406 percentiles of other exposure taking into account stratification with MC in first step (MC1S vs
 407 MC1, Table 3). Methods without stratification produced around 10% higher significant values
 408 of aggregate exposure (e.g. P50 observed at $0.86 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ and $0.95 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for
 409 the MC1S and MC1 methods respectively). For tap water, the values were around 50% lower
 410 with MC1 than with MC1S ($0.07 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ vs $0.14 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for the 95th percentile).
 411 The same significant differences were observed for contributions (Table 4).

412 Table 3. Descriptive statistics for aggregate lead exposure as well as lead exposure via food, soil, dust, tap water and air ($\mu\text{g}_{\text{Pb}}\cdot\text{kg}_{\text{bw}}^{-1}\cdot\text{d}^{-1}$) in
 413 children aged six months to three years in France. Estimates are expressed as median values and 95% uncertainty intervals presented in square
 414 brackets.

| Exposure sources | Methods | Mean | SD | P25 | P50 | P75 | P90 | P95 | P99 |
|--------------------|---------|--------------------|--------------------|--------------------|--------------------|-----------------------|--------------------|--------------------|--------------------|
| Aggregate exposure | MC1S | 3.1 [3.0 - 3.3] | 22 [17 - 58] | 0.49 [0.48 - 0.49] | 0.86 [0.85 - 0.86] | 1.9 [1.9 - 1.9] | 4.6 [4.6 - 4.7] | 8.8 [8.6 - 9.0] | 37 [36 - 39] |
| | MC2S | 2.8 [2.0 - 5.6] | 9.8 [4.8 - 55] | 0.48 [0.45 - 0.52] | 0.85 [0.78 - 0.95] | 1.9 [1.6 - 2.2] | 4.6 [3.8 - 5.7] | 8.7 [5.9 - 12] | 36 [18 - 88] |
| | MC1 | 3.5 [3.3 - 3.8] | 27 [19 - 77] | 0.55 [0.54 - 0.55] | 0.95 [0.94 - 0.96] | 2.0 [2.0 - 2.0] | 5.1 [5.0 - 5.1] | 10 [9.7 - 10] | 42 [40 - 44] |
| | MC2 | 3.2 [2.4 - 8.2] | 11 [4.9 - 89] | 0.55 [0.50 - 0.59] | 0.95 [0.86 - 1.1] | 2.0 [1.8 - 2.4] | 5.2 [4.0 - 6.4] | 9.4 [6.5 - 15] | 42 [20 - 105] |
| | MCIC | 3.5 [3.3 - 3.7] | 26 [19 - 55] | 0.55 [0.54 - 0.56] | 0.95 [0.94 - 0.96] | 2.0 [2.0 - 2.0] | 5.1 [5.0 - 5.1] | 10 [9.7 - 10] | 42 [40 - 44] |
| Food | MC1S | 0.22 [0.22 - 0.22] | 0.09 [0.09 - 0.09] | 0.16 [0.16 - 0.16] | 0.21 [0.21 - 0.21] | 0.27 [0.27 - 0.27] | 0.34 [0.33 - 0.34] | 0.38 [0.38 - 0.39] | 0.52 [0.52 - 0.53] |
| | MC2S | 0.22 [0.22 - 0.22] | 0.09 [0.09 - 0.09] | 0.16 [0.16 - 0.16] | 0.21 [0.21 - 0.21] | 0.27 [0.27 - 0.27] | 0.34 [0.34 - 0.34] | 0.38 [0.38 - 0.39] | 0.52 [0.52 - 0.53] |
| | MC1 | 0.22 [0.22 - 0.22] | 0.09 [0.09 - 0.09] | 0.16 [0.16 - 0.16] | 0.21 [0.21 - 0.21] | 0.27 [0.27 - 0.27] | 0.34 [0.33 - 0.34] | 0.38 [0.38 - 0.39] | 0.52 [0.52 - 0.54] |
| | MC2 | 0.22 [0.22 - 0.22] | 0.09 [0.09 - 0.09] | 0.16 [0.16 - 0.16] | 0.21 [0.21 - 0.21] | 0.27 [0.27 - 0.27] | 0.34 [0.33 - 0.34] | 0.38 [0.38 - 0.39] | 0.52 [0.52 - 0.52] |
| | MCIC | 0.22 [0.22 - 0.22] | 0.09 [0.09 - 0.09] | 0.16 [0.16 - 0.16] | 0.21 [0.21 - 0.21] | 0.27 [0.27 - 0.27] | 0.34 [0.33 - 0.34] | 0.38 [0.38 - 0.39] | 0.52 [0.52 - 0.54] |
| Soil | MC1S | 0.39 [0.39 - 0.40] | 0.77 [0.75 - 0.80] | 0.06 [0.06 - 0.06] | 0.15 [0.15 - 0.15] | 0.39 [0.38 - 0.39] | 0.95 [0.93 - 0.96] | 1.6 [1.5 - 1.6] | 3.6 [3.5 - 3.7] |
| | MC2S | 0.39 [0.33 - 0.45] | 0.72 [0.53 - 1.0] | 0.06 [0.06 - 0.07] | 0.15 [0.13 - 0.17] | 0.39 [0.34 - 0.46] | 0.95 [0.80 - 1.2] | 1.5 [1.2 - 1.9] | 3.3 [2.3 - 5.3] |
| | MC1 | 0.43 [0.42 - 0.43] | 0.86 [0.80 - 0.96] | 0.09 [0.09 - 0.09] | 0.22 [0.21 - 0.22] | 0.46 [0.46 - 0.47] | 0.92 [0.91 - 0.94] | 1.4 [1.4 - 1.5] | 3.5 [3.4 - 3.6] |
| | MC2 | 0.43 [0.36 - 0.53] | 0.73 [0.49 - 1.7] | 0.09 [0.08 - 0.12] | 0.22 [0.19 - 0.25] | 0.46 [0.40 - 0.54] | 0.92 [0.77 - 1.2] | 1.4 [1.1 - 1.8] | 3.5 [2.2 - 5.4] |
| | MCIC | 0.43 [0.42 - 0.43] | 0.86 [0.81 - 0.93] | 0.09 [0.09 - 0.09] | 0.22 [0.21 - 0.22] | 0.46 [0.46 - 0.47] | 0.93 [0.91 - 0.93] | 1.4 [1.4 - 1.4] | 3.5 [3.4 - 3.6] |
| Dust | MC1S | 2.5 [2.3 - 2.7] | 22 [17 - 57] | 0.07 [0.07 - 0.07] | 0.26 [0.25 - 0.26] | 0.97 [0.95 - 0.99] | 3.5 [3.4 - 3.6] | 7.7 [7.6 - 7.9] | 36 [35 - 39] |
| | MC2S | 2.2 [1.4 - 4.9] | 9.8 [4.7 - 55] | 0.08 [0.06 - 0.09] | 0.26 [0.21 - 0.33] | 0.94 [0.77 - 1.2] | 3.4 [2.5 - 4.6] | 7.7 [4.9 - 11.4] | 35 [18 - 87] |
| | MC1 | 2.8 [2.6 - 3.1] | 27 [19 - 77] | 0.08 [0.08 - 0.08] | 0.28 [0.28 - 0.29] | 1.1 [1.1 - 1.1] | 4.0 [3.9 - 4.1] | 8.9 [8.6 - 9.1] | 41 [39 - 43] |
| | MC2 | 2.6 [1.7 - 7.6] | 11 [4.8 - 89] | 0.08 [0.06 - 0.10] | 0.29 [0.23 - 0.38] | 1.2 [0.78 - 1.5] | 4.1 [2.9 - 5.7] | 9.0 [6.0 - 15] | 41 [19 - 105] |
| | MCIC | 2.8 [2.6 - 3.1] | 26 [19 - 55] | 0.08 [0.08 - 0.08] | 0.28 [0.28 - 0.29] | 1.1 [1.1 - 1.1] | 4.0 [3.9 - 4.1] | 8.9 [8.6 - 9.1] | 41 [39 - 43] |
| Tap water | MC1S | 0.03 [0.03 - 0.03] | 0.16 [0.16 - 0.17] | 0 [0 - 0] | 0 [0 - 0] | 0.007 [0.006 - 0.007] | 0.04 [0.04 - 0.04] | 0.14 [0.13 - 0.15] | 0.64 [0.64 - 0.69] |
| | MC2S | 0.03 [0.03 - 0.04] | 0.16 [0.10 - 0.20] | 0 [0 - 0] | 0 [0 - 0] | 0.007 [0.006 - 0.009] | 0.04 [0.03 - 0.05] | 0.14 [0.07 - 0.22] | 0.64 [0.48 - 1.2] |
| | MC1 | 0.02 [0.02 - 0.02] | 0.09 [0.09 - 0.10] | 0 [0 - 0] | 0 [0 - 0] | 0.007 [0.007 - 0.007] | 0.03 [0.03 - 0.03] | 0.07 [0.07 - 0.07] | 0.34 [0.32 - 0.35] |
| | MC2 | 0.02 [0.01 - 0.03] | 0.08 [0.04 - 0.17] | 0 [0 - 0] | 0 [0 - 0] | 0.007 [0.006 - 0.008] | 0.03 [0.02 - 0.04] | 0.07 [0.05 - 0.11] | 0.34 [0.17 - 0.62] |

| | | | | | | | | | |
|------------|-------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | MCIC | 0.02 [0.02 - 0.02] | 0.09 [0.08 - 0.10] | 0 [0 - 0] | 0 [0 - 0] | 0.007 [0.007 - 0.007] | 0.03 [0.03 - 0.03] | 0.07 [0.07 - 0.07] | 0.34 [0.32 - 0.35] |
| Air | MC1S | 0.005 [0.005 - 0.005] | 0.005 [0.005 - 0.005] | 0.002 [0.002 - 0.002] | 0.003 [0.003 - 0.003] | 0.006 [0.006 - 0.006] | 0.009 [0.009 - 0.009] | 0.012 [0.012 - 0.012] | 0.025 [0.025 - 0.026] |
| | MC2S | 0.005 [0.004 - 0.005] | 0.005 [0.004 - 0.006] | 0.002 [0.002 - 0.002] | 0.003 [0.003 - 0.004] | 0.006 [0.005 - 0.006] | 0.009 [0.008 - 0.010] | 0.012 [0.010 - 0.015] | 0.025 [0.018 - 0.034] |
| | MC1 | 0.005 [0.005 - 0.005] | 0.005 [0.005 - 0.005] | 0.003 [0.003 - 0.003] | 0.004 [0.004 - 0.004] | 0.007 [0.007 - 0.007] | 0.010 [0.010 - 0.010] | 0.013 [0.013 - 0.013] | 0.024 [0.024 - 0.025] |
| | MC2 | 0.005 [0.005 - 0.006] | 0.005 [0.004 - 0.006] | 0.003 [0.002 - 0.003] | 0.004 [0.004 - 0.004] | 0.007 [0.006 - 0.007] | 0.010 [0.009 - 0.011] | 0.013 [0.011 - 0.016] | 0.024 [0.018 - 0.032] |
| | MCIC | 0.005 [0.005 - 0.005] | 0.005 [0.004 - 0.005] | 0.003 [0.003 - 0.003] | 0.004 [0.004 - 0.004] | 0.007 [0.007 - 0.007] | 0.010 [0.010 - 0.010] | 0.013 [0.013 - 0.013] | 0.024 [0.024 - 0.025] |

415 **3.3.3 Taking into account correlations by using a vector of observations or**
 416 **reproducing correlations**

417 The different methods for taking into account concentration correlations (MCIC vs MC1)
 418 produced similar results, as the UIs overlapped.

419

420 Table 4. Mean contributions (%) of the various sources of exposure for the 50%, 10% and 5%
 421 of children (aged six months to three years) with the highest aggregate lead exposure.
 422 Estimates are expressed as median values and 95% uncertainty intervals presented in square
 423 brackets.

| | | 50% most exposed | 10% most exposed | 5% most exposed |
|------------------|-------------|-------------------------|-------------------------|------------------------|
| Food | MC1S | 13 [13 - 13] | 2.5 [2.5 - 2.6] | 1.3 [1.3 - 1.4] |
| | MC2S | 13 [11 - 14] | 2.5 [1.5 - 3.5] | 1.3 [0.8 - 2.0] |
| | MC1 | 12 [12 - 12] | 2.3 [2.3 - 2.3] | 1.2 [1.2 - 1.2] |
| | MC2 | 12 [10 - 13] | 2.3 [1.5 - 3.2] | 1.2 [0.61 - 1.9] |
| | MCIC | 12 [12 - 12] | 2.3 [2.2 - 2.3] | 1.2 [1.1 - 1.2] |
| Soil | MC1S | 27 [27 - 27] | 12 [12 - 13] | 5.7 [5.2 - 6.1] |
| | MC2S | 27 [24 - 30] | 12 [3.8 - 20] | 5.3 [1.3 - 16] |
| | MC1 | 28 [28 - 28] | 10 [9.8 - 11] | 5.5 [5.1 - 6.0] |
| | MC2 | 28 [24 - 32] | 9.5 [5.1 - 15] | 5.0 [1.1 - 14] |
| | MCIC | 28 [27 - 28] | 10 [9.7 - 11] | 5.6 [5.1 - 6.0] |
| Dust | MC1S | 57 [57 - 58] | 84 [84 - 85] | 93 [92 - 93] |
| | MC2S | 57 [53 - 62] | 84 [75 - 94] | 93 [82 - 98] |
| | MC1 | 59 [58 - 59] | 87 [87 - 88] | 93 [93 - 94] |
| | MC2 | 59 [54 - 64] | 87 [82 - 93] | 94 [85 - 98] |
| | MCIC | 59 [58 - 59] | 87 [87 - 88] | 93 [93 - 94] |
| Tap water | MC1S | 2.3 [2.3 - 2.4] | 0.93 [0.85 - 1.0] | 0.29 [0.26 - 0.34] |
| | MC2S | 2.4 [1.6 - 3.2] | 0.61 [0.03 - 2.4] | 0.21 [0.01 - 1.3] |
| | MC1 | 1.2 [1.1 - 1.2] | 0.33 [0.28 - 0.37] | 0.14 [0.12 - 0.17] |
| | MC2 | 1.2 [0.66 - 1.9] | 0.24 [0.04 - 1.1] | 0.10 [0.005 - 0.64] |
| | MCIC | 1.3 [1.2 - 1.4] | 0.25 [0.17 - 0.34] | 0.11 [0.07 - 0.19] |
| Air | MC1S | 0.25 [0.25 - 0.26] | 0.06 [0.05 - 0.06] | 0.03 [0.03 - 0.03] |
| | MC2S | 0.25 [0.20 - 0.29] | 0.05 [0.03 - 0.09] | 0.03 [0.02 - 0.05] |
| | MC1 | 0.28 [0.27 - 0.28] | 0.05 [0.05 - 0.06] | 0.03 [0.03 - 0.03] |
| | MC2 | 0.28 [0.22 - 0.33] | 0.05 [0.03 - 0.08] | 0.03 [0.01 - 0.05] |
| | MCIC | 0.28 [0.27 - 0.28] | 0.06 [0.05 - 0.06] | 0.03 [0.03 - 0.03] |

424

425 3.4 Contribution of exposure sources to aggregate lead exposure

426 Results obtained with the MC1S method are presented in this section.

427 3.4.1 Aggregate exposure

428 The 50th percentile of aggregate exposure was $0.86 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$, the 95th percentile was 8.8
429 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$, and the 99th percentile was $37 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ (Table 3). Lead exposure levels
430 in children varied considerably between the different exposure sources with a low
431 contribution of air and water and a high contribution of soil and dust. More specifically,
432 exposure via dust contributed the most to aggregate exposure, in particular for the most
433 exposed children, as it reached up to 93% of aggregate exposure at the P95 level. It was
434 followed by exposure via soil, food, tap water and air (Table 4).

435 C_{Dust} and DL were the exposure factors most influencing aggregate exposure, followed by
436 C_{soil} , Q_{soil} and Q_{Dust} (Table 5). Aggregate exposure was weakly sensitive to body weight
437 (BW).

438

439 3.4.2 Exposure via dust

440 Dust exposure values varied from $0.26 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for the 50th percentile to 7.7 for the 95th
441 percentile (Table 3). The contribution of dust ingestion to total exposure was 57% for median
442 aggregate exposure (Table 4). Regarding the top 10% most exposed children, the contribution
443 to aggregate exposure was 84%. C_{Dust} was positively correlated, at 0.572, with total exposure
444 (Table 5). Conversely, DL was negatively correlated, at -0.531.

445

446 3.4.3 Exposure via soil

447 Exposure via soil was the second largest factor contributing to aggregate exposure (Table 4).
448 The observed soil exposure values ranged from $0.15 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for the 50th percentile to
449 $1.6 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for the 95th percentile, reaching $3.6 \mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for the 99th percentile
450 (Table 3). The contribution of exposure via soil decreased from 27% for the top 50% of the
451 population to 5.7% for the top 5% most exposed children (Table 4). Aggregate exposure was
452 moderately sensitive to C_{Soil} (0.387) and Q_{Soil} (0.287) (Table 5).

453 Table 5. Correlations between exposure factors and lead concentrations with aggregate lead exposure in children aged six months to three years.

| | | MC1S | MC2S | MC1 | MC2 | MCIC |
|--------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Body weight | BW | 0.048 [0.041 - 0.055] | 0.047 [-0.017 - 0.121] | -0.046 [-0.051 - -0.039] | -0.046 [-0.126 - 0.043] | -0.046 [-0.052 - -0.039] |
| Soil | Q_{Soil} | 0.287 [0.282 - 0.292] | 0.291 [0.180 - 0.361] | 0.324 [0.317 - 0.329] | 0.323 [0.226 - 0.399] | 0.325 [0.319 - 0.335] |
| | C_{Soil} | 0.387 [0.382 - 0.391] | 0.385 [0.305 - 0.457] | 0.331 [0.326 - 0.336] | 0.316 [0.230 - 0.407] | 0.341 [0.270 - 0.393] |
| Dust | Q_{Dust} | 0.190 [0.185 - 0.190] | 0.188 [0.098 - 0.188] | 0.169 [0.164 - 0.169] | 0.168 [0.089 - 0.168] | 0.170 [0.165 - 0.170] |
| | C_{Dust} | 0.572 [0.567 - 0.576] | 0.576 [0.502 - 0.633] | 0.559 [0.554 - 0.563] | 0.560 [0.494 - 0.622] | 0.546 [0.518 - 0.570] |
| | DL | -0.531 [-0.535 - -0.527] | -0.532 [-0.611 - -0.468] | -0.536 [-0.541 - -0.532] | -0.533 [-0.608 - -0.457] | -0.541 [-0.550 - -0.533] |
| Tap water | Q_{Water} | 0.037 [0.032 - 0.037] | 0.035 [-0.036 - 0.035] | 0.042 [0.037 - 0.042] | 0.048 [-0.048 - 0.048] | 0.046 [0.040 - 0.046] |
| | C_{Water} | 0.032 [0.037 - 0.043] | -0.036 [0.035 - 0.105] | 0.037 [0.042 - 0.049] | -0.048 [0.048 - 0.136] | 0.040 [0.046 - 0.054] |
| Air | IR | 0.136 [0.131 - 0.143] | 0.146 [0.057 - 0.223] | 0.087 [0.081 - 0.094] | 0.084 [-0.005 - 0.159] | 0.096 [0.013 - 0.157] |
| | C_{Air} | -0.039 [-0.045 - -0.032] | -0.037 [-0.111 - 0.058] | 0.004 [-0.001 - 0.011] | -0.004 [-0.088 - 0.100] | 0.004 [-0.002 - 0.011] |

454

455 **3.4.4 Exposure via food**

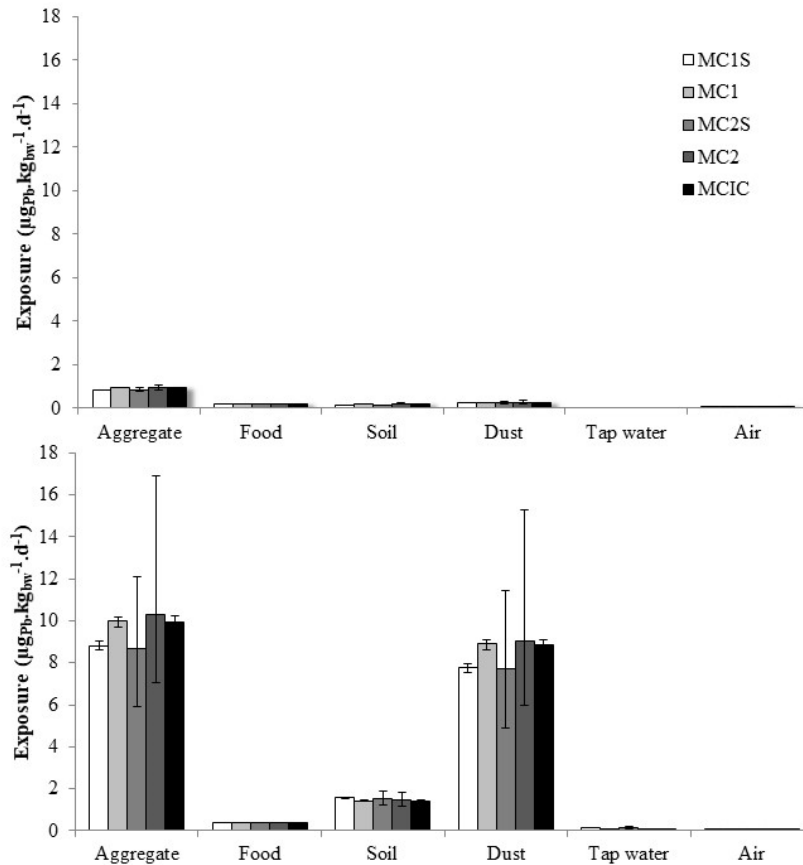
456 The 50th and 95th percentiles of exposure from food were observed respectively at 0.21 and
457 0.34 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$, reaching up to 0.52 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ for children in the top 1% for lead
458 exposure (Table 3). Dietary exposure moderately contributed to aggregate exposure, with its
459 contribution decreasing from 13% for the top half of the child population to 1.3% when
460 children were more heavily exposed to lead (Table 4).

461

462 **3.4.5 Exposure via air and tap water**

463 Exposure levels via air and tap water in children were very low compared to the other
464 exposure sources (Table 3). Exposure via air was observed at 0.012 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ at the 95th
465 percentile and accounted for 0.03% of aggregate exposure (Table 4). As with exposure via air,
466 exposure from tap water did not significantly contribute to the evaluation of aggregate
467 exposure, with values reaching 0.14 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ and with a contribution below 0.29% for
468 the top 5% of children most exposed to lead.

469



470

471 Figure 3. Histogram of the medians estimated for the 50th percentile (above) and the 95th
 472 percentile (below) of exposure ($\mu\text{g}_{\text{Pb}} \cdot \text{kg}_{\text{bw}}^{-1} \cdot \text{d}^{-1}$) to lead for children aged six months to three
 473 years, for aggregate exposure and for the various sources with the five implemented methods.
 474 Error bars show the P5 and P95 of the uncertainty intervals.

475

476 4 Discussion

477 4.1 Recommendations for aggregate exposure

478 To estimate aggregate exposure from different sources, it is ideal to have all exposure
 479 information for the same group of individuals with the same consistency regarding the units
 480 and methods, but this is rare in practice (Kennedy et al., 2015a). Thus, the main difficulty in
 481 the assessment of aggregate exposure involves combining data from several independent
 482 surveys and populations. Scientific articles related to aggregate exposure generally used
 483 different methodologies based on pre-existing models (Safford et al., 2015) or different
 484 softwares (MCRA, SHEDS, PACEM, BROWSE, etc.) with probabilistic simulations.

485 SHEDS-HT for example, combines dietary exposure from individual data with theoretical
486 values concentrations from distributions fitted on data for the other exposure sources. In the
487 case of MCRA, the software combines dietary exposure individual data with non-dietary
488 exposure data (empirical or theoretical) evaluated first by another software. Thus, providing a
489 harmonized method for combining different surveys with empirical data in case of aggregate
490 exposure was the main aim of this paper. It can also be applied to reconstruct the information
491 for the whole dataset for an existing study where different level of information was studied for
492 different sub-populations. This is the case for some biomonitoring surveys, where
493 measurements of chemicals are performed for sub-populations. It may also be applied for
494 researchers and practitioners to combine data as a unique database to have a unique survey for
495 different purposes.

496 To establish general steps to aggregate exposure from several surveys, five methods for
497 combining data were compared using children (6 month to 3 years old) lead exposure case
498 study in integrating uncertainties. They combined surveys focused on dietary exposure
499 (BEBE-SFAE combined with TDSi), contamination in homes (PH surveys) and in air (BDQA
500 survey). All methods provided consistent results and showed that dust was the main exposure
501 source although there were some significant differences between exposure levels due to the
502 difference in the methodology to combines these surveys. Moreover, uncertainties
503 considerably varied between methods. Bonnel et al. (2015) also concluded that combining
504 surveys with different populations and methodologies increased the probability to include
505 biases due to sampling or measurements. Thus, it is primordial to consider and follow the
506 recommended steps to decrease biases and uncertainties.

507 The first step consists in selecting a reference population between the three surveys to
508 simulate a new population of 100 000 individuals. It is an important step as its
509 sociodemographic characteristics and parameters will be duplicated in the simulated
510 population. Thus, the reference of the population needs to be the most representative of the
511 target population to considerably decrease these biases.

512 In the second step, it is recommended to consider stratification variables when
513 sociodemographic variables shared between the reference population and the other survey
514 variables are significantly correlated with concentration/exposure levels. Indeed, specific
515 parameters such as sociodemographic variables, especially when they influence the quality of
516 the other variables, may decrease the inclusion of biases when they are used to combine data
517 (Bayart and Bonnel, 2015). When stratification variables was not considered, as with MC1

518 and MC2 methods, aggregate exposure values tended to be overestimated (around 11% for
519 P50), compared to the methods considering stratification. The selection of stratification
520 variables is thus an important step in the calculation of aggregate exposure allowing to merge
521 appropriate exposure values for a specific sub-population. Stratification variables should be
522 carefully selected, with the presence of classes in sufficient but not too high numbers. Indeed,
523 the variables should not contain too many classes, otherwise it will be difficult to find
524 connections between surveys and the probability of drawing the same value will increase.
525 Conversely, when the number of classes is too low, there is a risk of not reproducing the
526 observed correlations. To define the number of classes, it is recommended to test their impact
527 when making statistical tests of the correlation significance between possible stratification
528 variables and the concentration levels. In this study, the age and the region variables were
529 used as stratification variables. Although gender is often used to combine data (Biesterbos et
530 al., 2013; Comiskey et al., 2017; Kennedy et al., 2015a; Kennedy et al., 2015b), this
531 parameter was not used as a stratification variable because it was not correlated with the
532 concentrations in soil, dust and tap water. Furthermore, combining different consumption
533 surveys with stratification variables are very current on the case of personal care products and
534 cosmetics (Biesterbos et al., 2013; Comiskey et al., 2017) due to the lack of data for product
535 usage for the subjects. Biesterbos et al. (2013) recommended for the combination of data to
536 use gender, age and the level of education as they are important factors in the case of
537 exposure assessment of personal care product. Comiskey et al. (2017) also merged data by
538 pairing subjects with similar demographics (age range, gender, and geography) assuming that
539 they will have similar habits and practices.

540 The last step is to combine data using MC simulations. Our results showed that the UIs of the
541 methods with MC in second step (MC2 and MC2S) were very large. These two methods
542 showed high uncertainty in exposure values through the 100 new populations, especially for
543 the most heavily exposed children. A difference of around 20% in the aggregate exposure
544 values was observed at the 50th percentile between one simulation and the next. This
545 difference could reach more than 50% for the 95th percentile. This meant that the exposure
546 values were not stable from one simulation to the next. Moreover, the variability within a
547 population simulated with MC in second step was lower than with MC in first step. Thus,
548 high exposure values were given less consideration with MC in second step. Consequently, it
549 is recommended to use the MC1S method which randomly samples individuals in the
550 reference population and simultaneously assigns exposure factors and concentration values

551 from other surveys. In this study, this resulted in consistent and stable exposure values
552 between the 100 simulations.

553 In the case of multiple and independent data sets, correlations are often observed for analysis
554 process. In this study, PH survey included several variables of interest, i.e. concentrations of
555 lead in tap water, in dust and in soil for one dwelling, which were highly correlated between
556 them. To keep these correlations, we selected vectors of these three concentration variables in
557 MC1S/MC1 and MC2S/MC2 methods, and reproduced correlations between the
558 concentration variables during simulation process by the Iman-Conover method (Iman and
559 Conover, 1982) with the MCIC method. In astrophysics field, methods treats correlations
560 between multiple data sets and give appropriate relevant weights of multiple data sets with
561 mutual correlations by the creation of a hyperparameter matrix. The marginalization can be
562 carried out with a brute-force grid evaluation of the hyperparameters, or it can be explored by
563 using MC methods which directly sample the posterior distribution (Ma and Berndsen, 2014).
564 Other methods such as copula methods (Haas, 1999), or principal component analysis
565 (Cowan-Ellsberry and Robison, 2009) can be used to take into account correlations during
566 simulations. Regarding ways of taking into account correlations during simulation processes,
567 there was no significant difference between MCIC method and the selection of a vector of
568 concentrations (MC1). MC1 has the advantage to avoid additional uncertainty related to the
569 choice of simulation correlations, while MCIC has the advantage of being easier to implement
570 with popular commercial add-ins to Excel® such as @risk, Cristal Ball.

571 4.2 **Aggregate children lead exposure in France**

572 In this paper, dust and soil were found to be the most significant sources of exposure to lead
573 in France for children under the age of three years. Similar results were observed in France
574 (Glorennec et al., 2016) and the USA (Zartarian et al., 2017) for the most heavily exposed
575 children between the ages of three and six years. Indeed, between soil and dust, dust is more
576 likely to accumulate trace metals (Acosta et al., 2015; Gabarron et al., 2017). The ingestion of
577 lead from soil and dust is very significant in children due to their more intense hand-mouth
578 behaviour (Gabarron et al., 2017; Glorennec et al., 2012). Contaminated soil and especially
579 dust have been identified as contributors to blood lead levels in children in France (Etchevers
580 et al., 2014; Etchevers et al., 2015; Glorennec et al., 2010; Oulhote et al., 2013). Exposure via
581 dust in this paper is six times higher than observed by Glorennec et al. (2016) in children aged
582 three to six years. This is due to some differences in exposure factors and in concentration

583 data used to evaluate exposure. Indeed, Glorennec et al. (2016) used for dust and soil
584 ingestion a lognormal distribution with a standard deviation of 3.2 and truncated to the highest
585 observed value whereas in this study, the lognormal distribution was simulated with a
586 standard deviation of around 26, allowing a larger range of possible values as shown in
587 original data. This difference emphasises the importance of carefully choosing exposure
588 parameters in order to increase the confidence level of the analysis. As exposure factors came
589 from studies with different periods and regions, error could appeared in exposure estimates.
590 For example, Q_{Soil} and Q_{Dust} which moderately impacted aggregate exposure have often been
591 discussed (Dor et al., 2012; Moya and Philips, 2014; Özkaynak et al., 2011; U.S. EPA, 2011;
592 von Lindern et al., 2016; Wilson et al., 2013). In the present study, Q_{Soil} and Q_{Dust} were taken
593 from the Exposure Factors Handbook (U.S. EPA, 2011) which derived distributions from 12
594 key studies mainly conducted in North America between the 1980s and 1990s. However,
595 since the 90's, activity patterns, micro-environments and hygiene practices have been
596 improved (Moya and Philips, 2014). Furthermore, these data came from studies conducted in
597 North America, while this case study focused on the French population. Soil ingestion
598 quantities can vary depending on the geographic location, climate, season, or soil
599 characteristics. To our knowledge, no data on quantities of ingested soil and dust and on
600 inhalation rate, are available for children in Europe. Thus, there is a need of further research
601 on exposure factors to improve data quality and exposure assessment in Europe.

602 After the ingestion of soil and dust, food ingestion was the source that most contributed to
603 total exposure to lead from the half of the most exposed children. For low exposure (25th
604 percentile), dietary exposure is higher than other sources of exposure. As dietary exposure
605 values were directly recorded in the reference population (i.e. BEBE-SFAE survey), it had
606 already been aggregated by construction with sample pooling, leading to low variations in
607 exposure in the simulated population. Consequently, no significant differences between
608 methods were observed for dietary exposure. Dietary exposure were relatively similar with
609 the median and 95th percentile values observed at 0.21 and 0.38 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$ in this study,
610 while values for older children were found at 0.37 and 0.42 $\mu\text{g}_{\text{Pb}}.\text{kg}_{\text{bw}}^{-1}.\text{d}^{-1}$. Similarly, dietary
611 exposure was the main contributor to aggregate lead exposure for older children for the half of
612 the population exposed to lead, with notable contributions by milk, fresh soft drinks,
613 vegetables and bread (Glorennec et al., 2016), and then was overtaken by the exposure of dust
614 and soil for the most exposed children. Lower exposure values from food and higher from
615 dust and soil were observed in the present work compared to results at European level from

616 (EFSA, 2010). The differences may come that dust and soil concentrations from the PH study
617 were higher (35mg/kg in mean to 831 in max for soil for ex.) than the single mean value used
618 by EFSA 2010 (23mg par kg, for soil). Moreover, EFSA used single mean value for ingestion
619 rate (100 mg per day) and body weight (12.5 kg) whereas we used probabilistic distributions.
620 In that way, the ingestion rate can reach 1000 mg per day in the present work.
621 The contribution of tap water and air to aggregate exposure was very low for children under
622 the age of three years. This is consistent with findings for older French children (Glorennec et
623 al., 2016) as well as for the exposure in air children in the USA (Zartarian et al., 2017). Lead
624 concentrations inside dwellings were extrapolated from outdoor measurements despite they
625 can be influenced by environmental tobacco smoke (Etchevers et al., 2014; Lucas et al.,
626 2014). Consequently, the extrapolated inhalation exposure was a weaker part due to lack of
627 relevant data.

628
629 EFSA (2010) estimated that the external dose corresponding to the BMDL₀₁ of 12 µg/L
630 (blood lead level) for developmental neurotoxicity was 0.50 µg.kg⁻¹ bw.d⁻¹. All values of the
631 margin of exposure from aggregate sources were low and below to 10, which meant that a risk
632 could not be excluded for all children. Furthermore, biomonitoring studies (Etchevers et al.,
633 2014) showed that most children under 6 years old exceed the BMD01. Consequently, current
634 levels of lead exposure, are a public health concern in France.

635

636 **5 Conclusion**

637 This work proposed a step-by-step approach for performing aggregate exposure assessments
638 by comparing different methods for combining heterogeneous surveys. The first step
639 consisted in selecting a representative reference population of the target population. In the
640 second step, it was recommended to consider stratification variables when combining surveys
641 to prevent exposure values from being overestimated. In the last step, it was recommended to
642 create a simulated population from the reference population and to simultaneously assign
643 exposure factors and concentration values from the other surveys using the stratification
644 variables. This timeline approach lowers the uncertainty of aggregate exposure. The
645 methodology was implemented to evaluate aggregate exposure to lead in the population of
646 children between the ages of six months and three years in France. Dust was the main
647 exposure source, followed by soil and food in a lesser extent.

648 This work is a first application of the proposed methodology which needs to be applied in
649 other case studies.

650

651 **Acknowledgements**

652 The authors thank the Plomb-Habitat and BEBE-SFAE surveys for providing the data. Plomb-
653 Habitat was supported by the French Directorate General for Health (DGS) and the French
654 Directorate for Housing (DHUP). The Nutri-Bébé SFAE 2013 survey dietary behaviour and
655 nutritional intakes from 0 to 3 years old was in partnerships with the surveys institutes
656 CREDOC and TNS SOFRES.

657

658 **Conflicts of Interest**

659 The authors declare no conflicts of interest.

660

661 **References**

- 662 Acosta, J. A., et al., 2015. Influence of population density on the concentration and speciation
663 of metals in the soil and street dust from urban areas. *Chemosphere*, 134, 328-37. doi:
664 10.1016/j.chemosphere.2015.04.038.
- 665 ANSES, Exposition agrégée au plomb : prise en compte des différents voies d'exposition.
666 Appui à la saisine 2013-SA-0092. Agence nationale de sécurité de l'alimentation, de
667 l'environnement et du travail, Maisons-Alfort, France, 2014, pp. 111.
- 668 Bayart, C., Bonnel, P., 2015. How to combine survey media (web, telephone, face-to-face):
669 Lyon and Rhône-Alps case study. *Transportation Research Procedia*, 11, 118 - 135.
670 doi: 10.1016/j.trpro.2015.12.011.
- 671 Beko, G., et al., 2013. Children's phthalate intakes and resultant cumulative exposures
672 estimated from urine compared with estimates from dust ingestion, inhalation and
673 dermal absorption in their homes and daycare centers. *PLoS One*, 8, (4), 1-18. doi:
674 10.1371/journal.pone.0062442.
- 675 Biesterbos, J. W., et al., 2013. Usage patterns of personal care products: important factors for
676 exposure assessment. *Food and Chemical Toxicology*, 55, 8-17. doi:
677 10.1016/j.fct.2012.11.014.
- 678 Bonnel, P., et al., 2015. Workshop synthesis: comparing and combining survey modes.
679 *Transportation Research Procedia*, 11, 108-117. doi: 10.1016/j.trpro.2015.12.010.
- 680 Bopp, S. K., et al., 2018. Current EU research activities on combined exposure to multiple
681 chemicals. *Environment International*, 120, 544-562. doi:
682 10.1016/j.envint.2018.07.037.
- 683 Cao, S., et al., 2016. Health risks of children's cumulative and aggregative exposure to metals
684 and metalloids in a typical urban environment in China. *Chemosphere*, 147, 404-11.
685 doi: 10.1016/j.chemosphere.2015.12.134.

686 Comiskey, D., et al., 2017. Integrating habits and practices data for soaps, cosmetics and air
687 care products into an existing aggregate exposure model. *Regulatory Toxicology and*
688 *Pharmacology*, 88, 144-156. doi: 10.1016/j.yrtph.2017.05.017.

689 Cowan-Ellsberry, C. E., Robison, S. H., 2009. Refining aggregate exposure: example using
690 parabens. *Regulatory Toxicology and Pharmacology*, 55, (3), 321-9. doi:
691 10.1016/j.yrtph.2009.08.004.

692 Delmaar, C., et al., 2015. Validation of an aggregate exposure model for substances in
693 consumer products: a case study of diethyl phthalate in personal care products. *Journal*
694 *of Exposure Science and Environmental Epidemiology*, 25, (3), 317-23. doi:
695 10.1038/jes.2014.68.

696 Dor, F., et al., Quantités de terre et poussières ingérées par un enfant de moins de 6 ans et
697 bioaccessibilité des polluants. Etat des connaissances et propositions. Institut de Veille
698 Sanitaire, InVS, Saint-Maurice, France, 2012, pp. 83.

699 EFSA, 2010. Scientific Opinion on Lead in Food. EFSA Panel on Contaminants in the Food
700 Chain (CONTAM). *EFSA Journal*, 8, (1570), 151p. doi: 10.2903/j.efsa.2010.1570.

701 EFSA, 2012. Guidance on the Use of Probabilistic Methodology for modelling Dietary
702 Exposure to Pesticide Residues. *EFSA Journal*, 10, (10). doi:
703 10.2903/j.efsa.2012.2839.

704 Etchevers, A., et al., 2014. Blood lead levels and risk factors in young children in France,
705 2008-2009. *International Journal of Hygiene and Environmental Health*, 217, (4-5),
706 528-37. doi: 10.1016/j.ijheh.2013.10.002.

707 Etchevers, A., et al., 2015. Environmental determinants of different blood levels in children:
708 A quantile analysis from a nationwide survey. *Environment International*, 74, 152-
709 159. doi: 10.1016/j.envint.2014.10.007.

710 Fantino, M., Gourmet, E., 2008. Nutrient intakes in 2005 by non-breast fed French children of
711 less than 36 months. *Archives de Pédiatrie*, 15, (4), 446-455. doi:
712 10.1016/j.arcped.2008.03.002.

713 Gabarron, M., et al., 2017. Soil or dust for health risk assessment studies in urban
714 environment. *Archives of Environmental Contamination and Toxicology*, 1-14. doi:
715 10.1007/s00244-017-0413-x.

716 Giovannangelo, M., et al., 2007. Variation of biocontaminant levels within and between homes
717 - the AIRALLERG study. *Journal of Exposure Science and Environmental*
718 *Epidemiology*, 17, 134-140. doi: 10.1038/sj.jes.7500482

719 Glorennec, P., et al., 2015. Exposition au plomb des enfants dans leur logement. *Projet*
720 *Plomb-Habitat (2008-2014) : principaux résultats, retombées et perspectives.*
721 *Environnement, Risques & Santé*, 14, (1), 28-37. doi: 10.1684/ers.2014.0751.

722 Glorennec, P., et al., 2012. French children's exposure to metals via ingestion of indoor dust,
723 outdoor playground dust and soil : Contamination data. *Environment International*, 45,
724 129-134. doi: 10.1016/j.envint.2012.04.010.

725 Glorennec, P., et al., 2016. Environmental and dietary exposure of young children to
726 inorganic trace elements. *Environment International*, 97, 28-36. doi:
727 10.1016/j.envint.2016.10.009.

728 Glorennec, P., et al., 2010. Identifying sources of lead exposure for children, with lead
729 concentrations and isotope ratios. *Journal of Occupational and Environmental*
730 *Hygiene*, 7, 253-260. doi: 10.1080/15459621003648281.

731 Gosens, I., et al., 2014. Aggregate exposure approaches for parabens in personal care
732 products: a case assessment for children between 0 and 3 years old. *Journal of*
733 *Exposure Science and Environmental Epidemiology*, 24, (2), 208-14. doi:
734 10.1038/jes.2013.33.

735 Guerin, T., et al., 2017. Levels of lead in foods from the first French total diet study on infants
736 and toddlers. *Food Chemistry*, 237, 849-856. doi: 10.1016/j.foodchem.2017.06.043.

737 Haas, C. N., 1999. On modeling correlated random variables in risk assessment. *Risk*
738 *Analysis*, 19, (6), 1205-1214. doi: 10.1023/A:1007047014741.

739 Hulin, M., et al., 2014. Assessment of infant exposure to food chemicals: the French Total
740 Diet Study design. *Food Additives and Contaminants: Part A*, 31, (7), 1226-39. doi:
741 10.1080/19440049.2014.921937.

742 Iman, R., Conover, W. A., 1982. A distribution-free approach to inducing rank correlation
743 among input variables. *Commun Stat: Simul Comput*, B11, (3), 311-334. doi.

744 Isaacs, K. K., et al., 2014. SHEDS-HT: An integrated probabilistic exposure model for
745 prioritizing exposures to chemicals with near-field and dietary sources. *Environmental*
746 *Science and Technology*, 48, 12750-12759. doi: dx.doi.org/10.1021/es502513w.

747 Jackson, C. H., 2011. Multi-state models for panel data: the msm package for R. *Journal of*
748 *Statistical Software*, 38, (8), 1-28. doi.

749 Karrer, C., et al., Submitted. Linking probabilistic exposure and pharmacokinetic modeling to
750 assess the cumulative risk from the bisphenols BPA, BPS, BPF, and BPAF for
751 Europeans. doi.

752 Kennedy, M. C., et al., 2012. BREAM: A probabilistic Bystander and Resident Exposure
753 Assessment Model of spray drift from an agricultural boom sprayer. *Computers and*
754 *Electronics in Agriculture*, 88, 63-71. doi: 10.1016/j.compag.2012.07.004.

755 Kennedy, M. C., et al., 2015a. A European model and case studies for aggregate exposure
756 assessment of pesticides. *Food and Chemical Toxicology*, 79, 32-44. doi:
757 10.1016/j.fct.2014.09.009.

758 Kennedy, M. C., et al., 2015b. New approaches to uncertainty analysis for use in aggregate
759 and cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79, 54-
760 64. doi: 10.1016/j.fct.2015.02.008.

761 Lucas, J.-P., et al., 2014. Source contributions of lead in residential floor dust and within-
762 home variability of dust lead loading. *Science of the Total Environment*, 470-471,
763 768-779. doi: 10.1016/j.scitotenv.2013.10.028.

764 Lucas, J.-P., et al., 2012. Lead contamination in French children's homes and environment.
765 *Environmental Research*, 116, 58-65. doi: 10.1016/j.envres.2012.04.005.

766 Lumley, T., 2004. Analysis of complex survey samples. *Journal of Statistical Software*, 9, (1),
767 1-19. doi.

768 Ma, Y.-Z., Berndsen, A., 2014. How to combine correlated data sets - A Bayesian
769 hyperparameter matrix method. *Astronomy and Computing*, 5, 45-56. doi:
770 10.1016/j.ascom.2014.04.005.

771 Millard, S. P., 2013. *EnvStats: An R package for environmental statistics*. Springer-Verlag
772 New York.

773 Moya, J., Philips, L., 2014. A review of soil and dust ingestion studies for children. *Journal of*
774 *Exposure Science and Environmental Epidemiology*, 24, 545-554. doi: 1559-0631/14.

775 Oulhote, Y., et al., 2013. Implications of different residential lead standards on children's
776 blood lead levels in France: predictions based on a national cross-sectional survey.
777 *International Journal of Hygiene and Environmental Health*, 216, (6), 743-50. doi:
778 10.1016/j.ijheh.2013.02.007.

779 Özkaynak, H., et al., 2011. Modeled estimates of soil and dust ingestion rates for children.
780 *Risk Analysis*, 31, (4), 592-608. doi: 10.1111/j.1539-6924.2010.01524.x.

781 Paustenbach, D. J., 2000. The practice of exposure assessment: a state-of-the-art review.
782 *Journal of Toxicology and Environmental Health, Part B Critical Reviews*, 3, (3), 179-
783 291. doi: 10.1080/10937400050045264.

784 Pelletier, M., et al., 2017. Indoor residential exposure to semivolatile organic compounds in
785 France. *Environment International*, 109, 81-88. doi: 10.1016/j.envint.2017.08.024.
786 Pouillot, R., et al., The mc2d package. In: R. p. v. 0.1-18, (Ed.), 2016.
787 R Core Team, R: A language and environment for statistical computing., R Foundation for
788 Statistical Computing, Vienna, Austria., 2017.
789 Safford, B., et al., 2015. Use of an aggregate exposure model to estimate consumer exposure
790 to fragrance ingredients in personal care and cosmetic products. *Regul Toxicol*
791 *Pharmacol*, 72, (3), 673-82. doi: 10.1016/j.yrtph.2015.05.017.
792 Stanek, E. J., et al., 2001. Biasing factors for simple soil ingestion estimates in mass balance
793 studies of soil ingestion. *Human and Ecological Risk Assessment*, 7, (2), 329-355. doi:
794 10.1080/20018091094394.
795 U.S. EPA, Exposure Factors Handbook 2011 Edition (final report). U.S. Environmental
796 Protection Agency, Washington, DC, 2011, pp. 1436.
797 van der Voet, H., et al., 2015. The MCRA model for probabilistic single-compound and
798 cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79, 5-12.
799 doi: 10.1016/j.fct.2014.10.014.
800 von Lindern, I., et al., 2016. Estimating children's soil/dust ingestion rates through
801 retrospective analyses of blood lead biomonitoring from the Bunker Hill superfund
802 site in Idaho. *Environmental Health Perspectives*, 124, (9), 1462-1470. doi:
803 [10.1289/ehp1510144](https://doi.org/10.1289/ehp1510144).
804 Wilson, R., et al., 2013. Revisiting dust and soil ingestion rates based on hand-to-mouth
805 transfer. *Human and Ecological Risk Assessment*, 19, (1), 158-188. doi:
806 10.1080/10807039.2012.685807.
807 Zartarian, V., et al., 2017. Children's Lead Exposure: A Multimedia Modeling Analysis to
808 Guide Public Health Decision-Making. *Environmental Health Perspectives*, 125, (9),
809 097009. doi: 10.1289/ehp1605.
810