**Supplementary Materials for**

**Association of Organic Food Consumption with Obesity in a Nationally Representative Sample**

Corentin J. Gosling1,2\*, Aurélie Goncalves3, Mickaël Ehrminger4 & Richard Valliant5

1Paris Nanterre University, DysCo Lab, France

2EA 4057, University of Paris, France

3UNIV. NIMES, EA7352 CHROME / APSY-V, F-30021 Nîmes Cedex, France

4Paris Saclay University, UVSQ, Simone Veil Health Sciences department

5Institute for Social Research, University of Michigan

\*Correspondence concerning this article should be addressed to Corentin J. Gosling, DysCo Lab, F9200 Paris Nanterre University, 200 Avenue de la République, 92000 Nanterre, France. Email: corentin.gosling@parisnanterre.fr

**This file includes:**

S1. Supplemental Text: Description of the sampling

S2. Supplemental Text: Description of the 24h-recall and the food propensity questionnaire

S3. Supplemental Text: Description of the physical activity levels

S4. Supplemental Analysis: Regression diagnostics

**S1. Supplemental Text: Description of the sampling**

The Third French Individual and National Food Consumption (INCA3) Survey 2014-2015 is a cross-sectional survey aiming at estimating food consumption and eating habits of individuals living in France. The study was carried out between February 2014 and September 2015 among a representative sample of individuals living in mainland France (excluding Corsica). A total of 5,855 individuals, divided into 2,698 0- to 17-year-old children and 3,157 18- to 79-year-old adults participated in the study. Individuals were selected according to a three-stage cluster sampling design (geographical units, households and individuals), based on the 2011 annual national census, with geographical stratification (region, size of urban area) in order to ensure national representativeness. Two independent random samples were drawn: a "Children" sample (0- to 17 year-old) and an "Adults" sample (18- to 79 year-old). Data related to various issues connected to food-related, nutritional and health risk/benefit assessment were collected: consumption of foods, drinks and food supplements, eating habits (eating occasions and places of consumption, consumption of foods grown or bred by the households, food production method, etc.), practices posing a potential health risk (food preparation and storage, consumption of raw foods of animal origin, etc.), knowledge and habits with regard to food. Data on physical activity and sedentarity as well as anthropometric and socio-demographic characteristics and standard of living were also collected. In order to ensure the national representativeness of the results presented, individuals and households weighting factors were estimated. The individual weighting factors were estimated separately for children and adults taking into account geographic and socio-economic variables. These weighting factors were systematically taken into account for analyses

**S2. Supplemental Text: Description of the 24h-recall and the food propensity questionnaire**

The dietary intake of the individuals was collected over 3 non-consecutive days (2 weekdays and 1 weekend day) spread over around 3 weeks. The 24h-recall method was used for individuals aged 15- to 79 year-old and the 24h-open-ended food record method (with a diary) for individuals aged 0- to 14 year-old. For the 3 selected days, individuals had to report their dietary intake by identifying all the foods and beverages consumed during day or night. They were asked to describe them in as much detail as possible and to quantify them using a picture book of food portions sizes and household measures. Regardless of the age, interviews were conducted by telephone, using the standardised and computerised GloboDiet software, by professional interviewers specifically trained in the methods and the software used. Among the 5,855 individuals included in the study, 4,114 (2,121 adults and 1,993 children) validated the DIET level of participation by responding at least to two dietary interviews.

**S3. Supplemental Text: Description of the physical activity levels**

In young children (10 years old or younger), physical activity has been defined as[[1]](#footnote-1):

* low if the number of days of physical activity is less than or equal to 2, and that the means of transport used to get to school is passive (car or public transit)
* high if the number of days of physical activity is greater than or equal to 5, and that the means of transport used to get to school is active (on foot, bicycle, scooter, etc.).
* moderate for any other intermediate situations.

For adolescents, (between 11 and 17 years of age), physical activity has been defined as:

* low if the adolescent participates in moderate or intense physical activity less than 5 times a week and participates in intense physical activity less than 3 times a week,
* high if the adolescent engages in moderate or vigorous physical activity on a daily basis, or engages in vigorous physical activity at least 5 times a week,
* moderate for any other intermediate situation.

For adults (18 years old or older), physical activity has been defined as:

* moderate if participants engage in intense physical activity at least 3 days per week for an average of at least 25min/day, engage in moderate physical activity at least 5 days per week for an average of at least 30min/day, or engage in moderate or intense physical activity at least 5 days per week, until a minimum of 600 Metabolic Equivalent of Task (METs)-minutes is reached.
* High is participants engage in intense physical activity at least 3 days per week, until a minimum of 1500 METs-minutes is reached, or engage in moderate or intense physical activity at least 1 day per week, until a minimum of 3000 METs-minutes is reached.
* Low for other situations

**S4. Supplemental Analysis: Regression diagnostics**

First, we performed Cook's distance, high leverages and standardized residuals analyses to determine whether there were any atypical observations that significantly influence the association of organic food consumption with BMI. To the best of our knowledge, diagnostics have not been implanted in common statistical software for logistic models fitted with complex survey data, like that of the INCA3 study. Therefore, only models with a continuous dependent variable (BMI) and without multiple imputations could have been analyzed. Therefore, all figures and results presented in this section are for linear regressions fitted with BMI as dependent variable, all a priori covariates as predictors and are run on participants with no missing value for these covariates (i.e., on the sample used for complete case analysis).

Table 8.A describes the coefficients for the regression models when the atypical points identified by each of the three methods were removed. All analyses were performed combining the ‘svydiags’ package (Valliant, 2018) with the ‘survey’ package in R (Lumley, 2020). The Figures 8.B, 8.C and 8.D depict the graphical results of these diagnostics.

 Second, we plot the standardized residuals against either the fitted values, the index of organic food consumption or the BMI (Figures 8.E, 8.F, 8.G). The plot 8.G revealed an interesting pattern in our two age groups, which could suggest a non-linear relationship between BMI and organic food consumption. This non-linear association was assessed in two successive analyses. In the first analysis, we added the square root of the index of organic food consumption as a predictor. In both age groups, the effect of the index of organic food consumption and its square root were non-significant. Then, we performed a Wald test assessing whether the joint effect of organic food consumption and its square root is equal to zero using the ‘survey’ package in R (Lumley, 2020). Results revealed a significant joint effect of these two predictors in adults, but not in children (Table 8.H). It should be noted that using the TMI instead of the BMI as the dependent variable in children leads to a significant joint effect of these predictors in the children group.

 Figure 8.F suggests that the variance of the residuals was larger for smaller values of organic food consumption. To investigate this in the second analysis, we weighted the adjusted model by the square root of the index of organic food consumption (in order to de-emphasize the influence of participants with a large value of this index). As reported in Table 8.I, this weighting improved model fit only in the children group, for which the *p-value* of the association between organic food consumption and BMI became marginally significant.

Table 8.A.

Coefficients for the adjusted regression models in which BMI was used as the dependent variable.

|  |  |
| --- | --- |
|  | BMI |
|  | *b* (SE) | *p* | Partial *r* | Method | N excluded |
| **Children and adolescents (N = 1458)** |
| Adjusted Model | -0.026 (0.013) | .045 | -0.05 | Cook’s distance | 0 |
| Adjusted Model | -0.028 (0.010) | .008 | -0.07 | High Leverage | 58 |
| Adjusted Model | -0.020 (0.011) | .069 | -0.05 | Standardized Residuals | 27 |
| **Adults (N = 1679)** |
| Adjusted Model | -0.070 (0.017) | <.001 | -0.10 | Cook’s distance | 0 |
| Adjusted Model | -0.083 (0.017) | < .001 | -0.12 | High Leverage | 84 |
| Adjusted Model | -0.075 (0.016) | < .001 | -0.11 | Standardized Residuals | 19 |

*Adjusted models are run with all a priori covariates, demographics (age and sex), SES (family income and education level), nutritional covariates (MD adherence, regime, dietary supplements and energy intake) and physical activity (physical activity and sedentarity). Participants with any missing value on the independent variable, dependent variable or covariates were excluded from all models. BMI was treated as a continuous variable (linear regression).*

****

Figure 8.B. Values of the modified Cook’s distance for each participant in the children (A) and adult group (B). Values of modified Cook’s distance are considered as large if greater than 3.

****

Figure 8.C. Hat values for each participant in the children (A) and adult group (B). Hat values are considered as large if larger than 3 times the mean of all hat values.

****

Figure 8.D. Standardized residuals values for each participant in the children (A) and adult group (B). Standardized residuals values are considered as large if their absolute values are larger than 3.



Figure 8.E. Scatterplot of the fitted values and the standardized residuals values for the children (A) and adult group (B).



Figure 8.F. Scatterplot of the index of organic food consumption and the standardized residuals values for the children (A) and adult group (B).



Figure 8.G. Scatterplot of the BMI and the standardized residuals values for the children (A) and adult group (B).

Table 8. H
Coefficients for the multiple regression models in which BMI is used as the dependent variables and the square root of the organic food consumption is included as an additional predictor.

|  |  |
| --- | --- |
|  | BMI |
|  | *b* (SE) | *p* | Predictor |
| **Children and adolescents (N = 1458)** |
| Adjusted Model  | -0.01 (0.046) | .84 | Index of organic food consumption |
| -0.077 (0.207) | .71 | Square root of the index of organic food consumption |
| Joint test | F = 2.06 | .13 |  |
| **Adults (N = 1679)** |
| Adjusted Model | -0.031 (0.052) | .56 | Index of organic food consumption |
| -0.18 (0.248) | .46 | Square root of the index of organic food consumption |
| Joint test | F = 8.66 | <.001 |  |

*Adjusted models are run with all a priori covariates, demographics (age and sex), SES (family income and education level), nutritional covariates (MD adherence, regime, dietary supplements and energy intake) and physical activity (physical activity and sedentarity) and the square root of the index of organic food consumption. Participants with any missing value on the independent variable, dependent variable or covariates were excluded from all models. BMI was treated as a continuous variable (linear regression).*

Table 8. I
Coefficients for the multiple regression models in which BMI is used as the dependent variables and the square root of the organic food consumption is used as weight.

|  |  |  |
| --- | --- | --- |
|  | BMI |  |
|  | *b* (SE) | *p* | Partial r | Weight | Model fit (AIC) |
| **Children and adolescents (N = 1458)** |  |
| Adjusted Model | -0.026 (0.013) | .045 | -0.05 | No weight | 7892 |
| Adjusted Model | -0.023 (0.013) | .066 | -0.05 | Square root of the index of organic food consumption | 7855 |
| **Adults (N = 1679)** |  |
| Adjusted Model | -0.070 (0.017) | < .001 | -0.10 | No weight | 10400 |
| Adjusted Model | -0.061 (0.018) | .001 | -0.08 | Square root of the index of organic food consumption | 10460 |

*Adjusted models are run with all a priori covariates, demographics (age and sex), SES (family income and education level), nutritional covariates (MD adherence, regime, dietary supplements and energy intake) and physical activity (physical activity and sedentarity). Participants with any missing value on the independent variable, dependent variable or covariates were excluded from all models. BMI was treated as a continuous variable (linear regression).*

1. Some adjustments were made for children who were out of school, sick or on holiday during the week before the interview: these can be retrieved in ANSES (2017), available at https://www.anses.fr/fr/system/files/NUT2014SA0234Ra.pdf. [↑](#footnote-ref-1)