Supplemental material

Appendix 1: Precisions and examples of ultra-processed foods according to the NOVA classification

All food and beverage items of the NutriNet-Santé composition table were categorized by a team of three trained dieticians into one of the four food groups in NOVA, a food classification system based on the extent and purpose of industrial food processing [1-3]. The whole classification was then reviewed by a committee composed of the three dietitians and five researchers, specialists in nutritional epidemiology. In case of uncertainty for a given food/beverage item, a consensus was reached among researchers based on the percentage of home-made and artisanal foods versus industrial brands reported by the participants.

The “ultra-processed foods” group of the NOVA classification is the primarily focus of this study. Examples of such products as well as examples of distinctions between ultra-processed products and products from other NOVA categories are provided below:

Examples of ultra-processed food according to the NOVA classification:

Carbonated drinks; sweet or savoury packaged snacks; ice-cream, chocolate, candies (confectionery); mass-produced packaged breads and buns; margarines and spreads; industrial cookies (biscuits), pastries, cakes, and cake mixes; breakfast ‘cereals’, ‘cereal’ and ‘energy’ bars; ‘energy’ drinks; flavoured milk drinks; cocoa drinks; sweet desserts made from fruit with added sugars, artificial flavours and texturizing agents; cooked seasoned vegetables with ready-made sauces; meat and chicken extracts and ‘instant’ sauces; ‘health’ and ‘slimming’ products such as powdered or ‘fortified’ meal and dish substitutes; ready to heat products including pre-prepared pies, pasta and pizza dishes; poultry and fish ‘nuggets’ and ‘sticks’, sausages, burgers, hot dogs, and other reconstituted meat products, and powdered and packaged ‘instant’ soups, noodles and desserts.

For instance, fruit compotes with only added sugar are considered as “processed foods”, while flavoured fruit desserts with added sugar, texturizing agents and colorants are considered as “ultra-processed foods”.

Regarding meats, salted-only red or white meats are considered as “processed foods” whereas smoked or cured meats with added nitrites and conservatives, such as sausages and ham are classified as “ultra-processed foods”.

Similarly, canned salted vegetables are considered as “processed foods” whereas industrial cooked or fried seasoned vegetables, marinated in industrial sauces with added flavourings are considered as “ultra-processed foods”.

Example of list of ingredients for an industrial Chicken and Leek flavour soup considered as “ultra-processed” according to the NOVA classification: *“Dried Glucose Syrup, Potato Starch, Flavourings, Salt, Leek Powder (3.6%), Dried Leek (3.5%), Onion Powder, Dried Carrot, Palm Oil, Dried Chicken (0.7%), Garlic Powder, Dried Parsley, Colour [Curcumin (contains MILK)], Ground Black Pepper, MILK Protein, Stabilisers (Dipotassium Phosphate, Trisodium Citrate)”*.

Appendix 2: Flow chart

118,290 participants included in NutriNet-Santé, until August 2015

→ 7,903 with prevalent cancer at baseline

110,387

→ 5,407 participants with less than two dietary records

104,980

104,980 participants included:

22821 (21.7%) men and 82159 (78.3%) women

Appendix 3: Method for deriving dietary patterns by principal component analysis and corresponding factor loadings

Dietary patterns were produced from principal-components analysis based on 20 predefined food groups, using the SAS ‘‘Proc Factor’’ procedure (SAS Institute Inc., Cary, North Carolina). This factor analysis forms linear combinations of the original food groups, thereby grouping together correlated variables. Coefficients defining these linear combinations are called factor loadings. A positive factor loading means that the food group is positively associated with the factor, whereas a negative loading reflects an inverse association with the factor. For interpreting the data, we considered foods with a loading coefficient under -0.25 or over 0.25. We rotated factors by orthogonal transformation using the SAS ‘‘Varimax’’ option to maximize the independence (orthogonality) of retained factors and obtain a simpler structure for easier interpretation. In determining the number of factors to retain, we considered eigenvalues greater than 1.25, the scree test (with values being retained at the break point between components with large eigenvalues and those with small eigenvalues on the scree plot), and the interpretability of the factors. For each subject, we calculated the factor score for each pattern by summing observed consumption from all food groups, weighted by the food group factor loadings. The factor score measures the conformity of an individual’s diet to the given pattern. Labeling was descriptive, based on foods most strongly associated with the dietary patterns. The healthy pattern (explaining 10.6% of the variance) was characterized by higher intakes of fruit, vegetables, soups and broths, unsweetened soft drinks and whole grains and lower sweetened soft drinks intake. The Western pattern (explaining 7.0% of the variance) was characterized by higher intakes of fat and sauces, alcohol, meat and starchy foods.

	Factor loadings	
	Healthy Pattern	Western Pattern
Alcoholic drinks	-.099552	0.284771
Breakfast cereals	0.079447	-.181769
Cakes and biscuits	-.197629	0.003444
Dairy products	0.066066	-.013702
Eggs	0.078582	0.043744
Fats and sauces	0.012600	0.544911
Fish and seafood	0.204373	0.100759
Fruit	0.354075	0.052298
Meat	-.188274	0.318483
Pasta and rice	-.212857	0.341941
Potatoes and tubers	-.029615	0.402694
Poultry	-.030137	0.064064
Processed meat	-.228028	0.207877
Pulses	0.192815	0.026104
Soups and broths	0.264233	0.227787
Sugar andconfectionery	-.088870	0.120660
Sweetened soft drinks	-.288870	-.007506
Unsweetened soft drinks	0.258563	0.152704
Vegetables	0.471255	0.231818
Whole grains	0.380881	-.043132

Appendix 4: Methodology and results of the mediation analysis

Mediation analyses were carried out according to the method proposed by Lange et al.[4] in order to evaluate the direct and indirect "effects" in the relationship between the exposure and the outcome, through nutritional mediators. Under the assumption of a causal relationship between quartiles of the proportion of ultra-processed food in the diet (=Exposure, quoted "A") and cancer risk (=Outcome, quoted "Y"), the aim was to estimate how much of this effect was mediated through various factors reflecting the nutritional quality of the diet. The latter factors (dietary intakes of sodium, total lipid, fatty acids, and carbohydrates, and Western-type dietary pattern) were considered as potential Mediators (quoted "M") in each model. The following covariates were considered as potential confounders (quoted "C"): age, sex, BMI, height, physical activity, smoking status, number of 24h-dietary records, alcohol intake, energy intake, family history of cancer, and educational level. To evaluate the direct effect and the indirect effect mediated by each nutritional factor, we applied a mediation analysis in the counterfactual framework. The mediation analyses were implemented according to the following steps for a categorical exposure:

(1) Construction of a new data set by repeating each observation in the original data set. This new variable A* corresponds to the value of the exposure relative to the indirect path. Each observation was repeated four times such that A* got to take all possible values of exposure (quartiles of ultra-processed).

(2) Fitting of a multinomial logistic regression applied to the new data set to estimate the association between ultra-processed food and cancer, conditioned on baseline confounders, and computing predicted values, first using the original variable A and then the new variable A*.

(3) Weighting (W) each observation calculated according to the following formula through applying the fitted models from steps 2 et 3 to the new dataset:

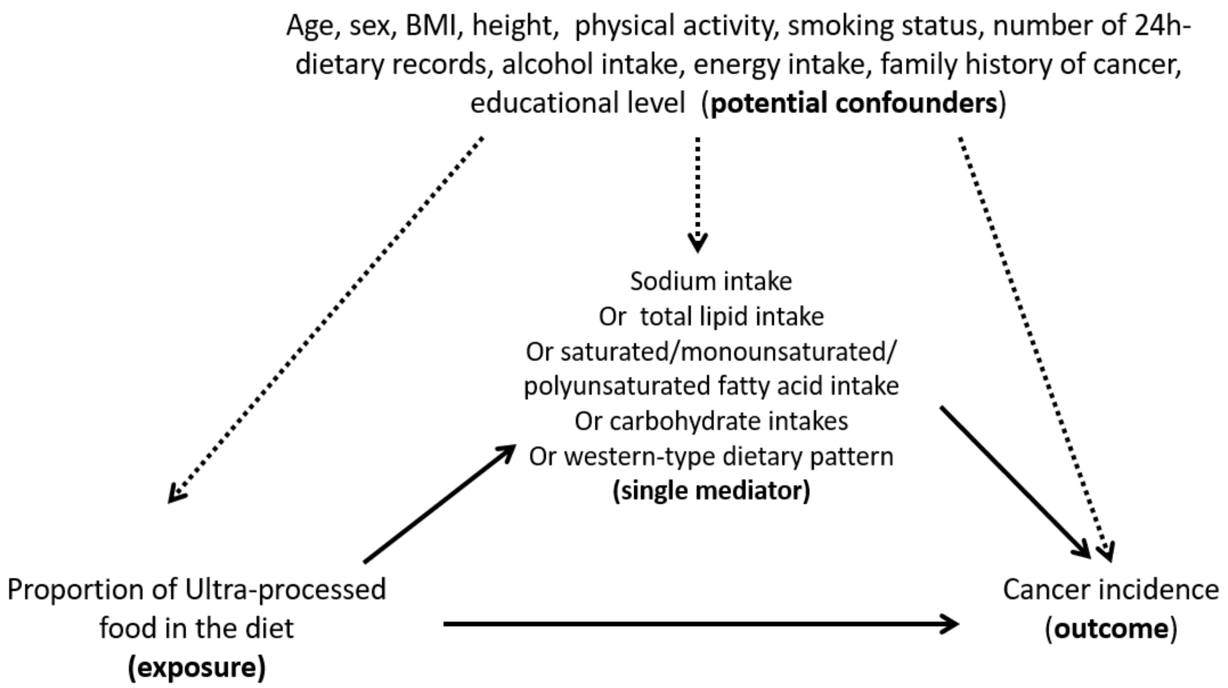
$$W_i = \frac{1}{P(A = A_i \mid C = C_i)} \frac{P(M = M_i | A = A_i^*, C = C_i)}{P(M = M_i | A = A_i, C = C_i)}$$

with A, the exposure, M, the mediator, C, the set of baseline confounders

(4) Fitting of a weighted Cox Marginal Structural Model (MSM) for direct and indirect effects controlling for baseline confounders, as the outcome corresponds to a survival time. The "Covsandwich" statement in SAS software allows getting robust standard errors.

(5) To evaluate how much of the total effect was due to the mediator effect, we calculated the 'proportion explained' by each single mediator as $(HR_{\text{total effect}} - HR_{\text{direct effect}}) / (HR_{\text{total effect}} - 1)$ where $HR_{\text{total effect}}$ and $HR_{\text{direct effect}}$ were respectively, the Hazard Ratios for total effect and for direct effect.

The figure below shows a conceptual model of the association between the proportion of ultra-processed foods in the diet and cancer risk, taking into account nutritional factors as potential single mediators:



Conceptual model of the association between ultra-processed food consumption and change in cancer risk taking into account nutritional factors as potential single mediator

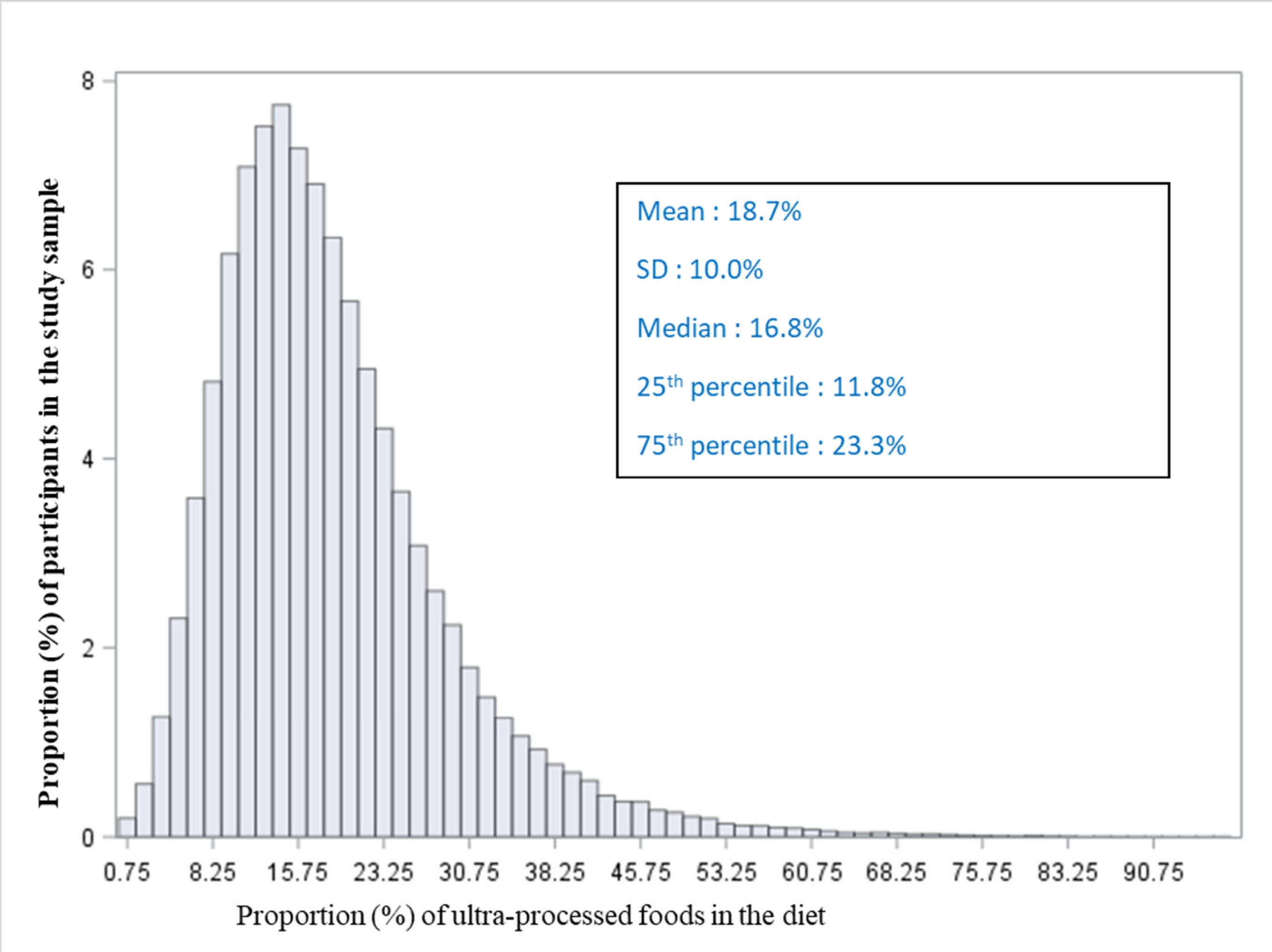
The table below shows the results of mediation analyses testing for a potential mediation by total lipid, carbohydrate, sodium, SFA, PUFA and MUFA intakes, and the Western dietary pattern of the association between ultra-processed food intake and cancer risk.

Table 1 – Hazard Ratios of direct, indirect and total effects and proportion of total effects mediated by several nutritional factors in the prospective associations between ultra-processed food and overall cancer risk, N=104980, NutriNet-Santé cohort, France, 2009-2017

	Tested nutritional mediators of the association between ultra-processed foods and overall cancer risk													
	Total lipids		Sodium		Carbohydrates		Western pattern		SFAs		PUFAs		MUFAs	
Effect	HR	p-value	HR	p-value	HR	p-value	HR	p-value	HR	p-value	HR	p-value	HR	p-value
Indirect effect	1.000	0.799	1.003	0.889	1.000	0.900	1.005	0.910	1.000	0.900	1.000	0.900	1.000	0.900
Direct effect	1.302	<0.0001	1.263	<0.0001	1.217	<0.0001	1.317	<0.0001	1.166	0.001	1.319	<0.0001	1.328	<0.0001
Total effect	1.302		1.267		1.217		1.324		1.166		1.319		1.328	
Proportion of the total effect mediated by the nutritional factor	0.00%		1.42%		0.00%		2.04%		0.00%		0.00%		0.00%	

SFAs: saturated fatty acids, PUFAs: poly-unsaturated fatty acids, MUFAs: mono-unsaturated fatty acids, HR: Hazard Ratio

Appendix 5: Distribution of the main exposure (proportion of ultra-processed food in the diet) in the study sample (N=104 980), NutriNet-Santé, France



Appendix 6: Associations between the quantity (g/d) of each ultra-processed food group and overall and breast cancer risks, from multivariable Cox proportional hazard models, NutriNet-Santé cohort, France, 2009 – 2017 (n=104,980)

	Continuous		
	HR[a,b]	95%CI	P-value
All cancers			
N for cases/non cases		2228/102752	
Starchy foods	1.01	(0.99-1.02)	0.4
Fruits and vegetables	1.00	(0.99-1.01)	0.2
Dairy products	1.01	(1.00-1.02)	0.05
Fats	1.07	(1.03-1.12)	0.002
Salty snacks	0.98	(0.93-1.02)	0.3
Meat, fish, eggs	1.01	(0.99-1.03)	0.4
Processed meat	0.99	(0.97-1.01)	0.5
Sugary products	1.01	(1.00-1.02)	0.03
Beverages	1.00	(1.00-1.01)	0.005
Breast Cancer			
N for cases/non cases		739/81420	
Starchy foods	1.00	(0.98-1.03)	0.7
Fruits and vegetables	1.01	(0.99-1.02)	0.3
Dairy products	1.01	(0.99-1.02)	0.3
Fats	1.06	(0.97-1.14)	0.2
Salty snacks	1.02	(0.95-1.10)	0.6
Meat, fish, eggs	1.01	(0.97-1.04)	0.8
Processed meat	0.98	(0.94-1.02)	0.4
Sugary products	1.02	(1.01-1.03)	0.006
Beverages	1.00	(0.99-1.01)	0.2

CI, confidence interval, HR, Hazard ratio

[a] adjusted for age (timescale), sex, energy intake without alcohol, number of 24h-dietary records, smoking status, educational level, physical activity, height, BMI, alcohol intake, and family history of cancers. Breast cancer models were additionally adjusted for menopausal status, hormonal treatment for menopause, oral contraception and number of children.

[b]HR for an increase of 10g of the quantity (in g/d) of each ultra-processed food group

References

(1) Monteiro CA, Cannon G, Levy RB, Moubarac JC, Jaime PC, Martins AP et al. NOVA. The star shines bright. *World Nutrition* 2016; 7(1-3):28-38.

(2) Monteiro CA, Cannon G, Moubarac JC, Levy RB, Louzada ML, Jaime PC. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutr* 2017;1-13.

(3) Moubarac JC, Parra DC, Cannon G, Monteiro CA. Food Classification Systems Based on Food Processing: Significance and Implications for Policies and Actions: A Systematic Literature Review and Assessment. *Curr Obes Rep* 2014; 3(2):256-272.

(4) Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol* 2012; 176(3):190-195.