European Journal of Applied Sciences – Vol. 11, No. 3
Publication Date: June 25, 2023
DOI:10.14738/aivp.113.14640.
Cornelli, U., Grossi, E., Recchia, M., Cestaro, B., Rondanelli, M., Cazzola, R., & Pistolesi, E. (2023). The Mediterranean Diet and Cancers in Italy: Stochastic and Non-Stochastic Analysis. European Journal of Applied Sciences, Vol - 11(3). 59-84.

SERVICES FOR SCIENCE AND EDUCATION

The Mediterranean Diet and Cancers in Italy: Stochastic and Non-Stochastic Analysis

Cornelli, Umberto
Loyola University School of Medicine, Chicago

Grossi, Enzo
Villa Santa Maria Foundation, Tavernerio, Italy

Recchia, Martino
Medistat Epidemiology and Clinical Biostatistics,
Mario Negri Institute Alumni Association, Milan, Italy

Cestaro, Benvenuto
University of Milan, Italy

Rondanelli, Mariangela
Department Public health, Experimental and
Forensic Medicine University of Pavia, Italy

Cazzola, Roberta
Department of Biochemical and Clinical Sciences,
University of Milan, Italy

Pistolesi, Elvira
University of Milan, Italy

ABSTRACT

Background: The relationship between foods and cancers has been documented in many epidemiological and cohort studies of the Mediterranean Diet (MeD). Objective: Correlation between food expenditure in Italy and the prevalence of nine cancers: breast, prostate, colon, pancreas, TBL (throat-bronchial-lung), bladder in males and females, uterus, and ovaries, with a comparison of differences between Northern and Southern Italy. Methods: The expenditure of 19500 Italian families on the 56 most sold food categories was taken from ISTAT (Italian National Institute of Statistics) records for 2016. These data were then correlated with the prevalence of death from the nine above mentioned cancers in 2020. Stochastic and non-stochastic analyses were used to determine the correlations between foods and cancers. Results: None of the cancers showed identical protective or causative food patterns. The data confirmed the causative role of wine and the protective role of pasta, flour and sugar. Surprisingly fish and olive oil were found to be causative for some cancers. Southern Italy showed significantly lower prevalence of cancer due to a lower consumption of causative foods and higher consumption of protective ones. Conclusion: MeD has changed in Italy. Some causative and protective foods were found for all the cancers, but none had the same pattern. However, foods can

be considered important co-factors in cancer development. Pasta, flour and sugar were shown to be protective for many cancers and, besides wine, some other foods, like olive oil and fish, were causative for some cancers. The correlation between food expenditure and disease seems to be a valid, simple research method.

Keywords: breast cancer, prostate cancer, colon cancer, pancreas cancer, TBL (throat-bronchial-lung) cancer, bladder cancer in males and females, uterus cancer, ovarian cancer, foods.

INTRODUCTION

The MeD was first defined by Ancel Keys following the "seven countries study" in which some countries with coasts on the Mediterranean Sea were involved in a twenty-year study [1, 2] to investigate the effect of eating habits on different diseases (cardiovascular and cancer).

The first description of the diet [3] included homemade minestrone, pasta served with tomato sauce and a sprinkle of cheese, only occasionally enriched with a few pieces of meat or served with a little local sea food, lots of bread, large quantities of fresh vegetables, moderate portions of meat and fish (approximately twice a week), some red wine, and fresh fruit for dessert. However, there was not sufficient material to give a proper definition of what the MeD was [4]. The MeD has changed during the years, and it would be more appropriate to take data from individual countries to correlate diet and diseases.

In more recent evaluations, the diet in Italy is characterized by a relatively high consumption of cereals, fruit, vegetables, legumes, fish, nuts and olive oil, together with a relatively low amount of dairy food, red meat and wine (5, 6]. A recent review showed that the MeD may still be protective for various cancers [7-9]. The high content of micronutrients and phytochemicals may be considered protective [10,11]. However, most of the time, these studies do not report the quantities of food involved.

Furthermore, many of the cohort studies were conducted on relatively small numbers of subjects using food intake questionnaires (FIQ). The epidemiological studies also use the same tools, but take larger samples to represent the whole population living in the various geographical areas and pool the data from all the countries. However, because FIQs generally do not allow proper quantification of foods and rely on individuals' memory, they are less accurate than other dietary assessment methods [12].

Unfortunately, at present, food education is mainly in the hands of advertising, TV cookery programmes and talk shows, which are rarely supported by experts in nutrition.

More than 75% of TV advertising in Italy regards food, and follows the very common mantra "eat what you like most, provided it is in the pyramid of what is produced".

The EFSA (European Food Safety Agency) usually gives information about the toxicity of some foods, but we really need a more systematic analysis of the correlation between foods and diseases. This is possible, and plenty of data are available to provide correct information.

This study takes a new approach. It considers the relationship between food expenditure and nine different cancers: breast, prostate, colon (colorectal), pancreas, TBL (throat-bronchial-lung), bladder cancer in males and females, uterine and ovarian cancers.

The expenditure of 19500 families on 56 food categories in 2016 was considered. The families lived in 540 municipalities and formed a representative sample of the population of Italy. This expenditure was correlated (linear correlation) with the prevalence of the cancers in 2020, and, where necessary, the food quantities/year were derived from the cost/kg.

MATERIAL AND METHODS

The prevalence of death from cancer (per 10^4 inhabitants) in Italy was taken from the 2020 ISTAT records [13]. The food expenditure was taken from public data recorded by ISTAT [14] using the CAPI (computer-assisted personal interview) system. This system is considered the most reliable, complete and meticulous questionnaire on food expenditure. It is based on 14-day records compiled with technical assistance [15]. The expenditure of 19500 families in 540 municipalities in the 20 regions of Italy on food purchased from local stores or produced by households was calculated for 2016. The sample covers 6.3% of all municipalities, which ISTAT considers representative of each region of Italy. According to the 2016 ISTAT calculations, the 19500 families consisted of approximately 45000 people (each family was composed of an average of 2.3 members). In total, there were 19500 adult women (aged > 24 and < 45 years) and 19500 adult men (aged > 24 and <45). The remaining subjects were between 0 to 14 years old.

The following geographic areas were considered: Northern Italy (Aosta Valley, Piedmont, Liguria, Lombardy, Veneto, Friuli-Venezia Giulia, Trentino-Alto Adige, Emilia Romagna and Tuscany), Southern Italy (Lazio, Umbria, Marche, Abruzzo, Campania, Molise, Basilicata, Apulia, Calabria and the islands of Sicily and Sardinia). Piedmont and Aosta Valley were taken together as one region only (Piedmont VA). When necessary, the expenditure was transformed into quantities based on the cost/Kg. Since the costs of the foods were not the same throughout Italy, the values in terms of Kg were corrected by assuming that prices were 15% lower in the southern regions.

Statistical Analysis

The correlations were calculated by taking the expenditure/year for each food category and the prevalence of death from cancer (per 10^4 inhabitants). Only values of r^2 higher than 0.456 or lower than -0.456 were considered significant. The difference in the prevalence of death between northern and southern regions was calculated using the Man Whitney test.

SAS Institute's JMP Pro 14 (2019) software was used for the analysis.

For the non-stochastic analysis, a Euclidean minimum spanning tree (MST) was drawn up for the 56 food categories. The MST considered cancers in terms of high and low prevalence. The food categories close to the “low prevalence” node were considered protective, while those close to the “high prevalence” node were considered causative.

Details regarding MSTs are provided in the appendix.

RESULTS

The expenditure on the 56 food categories and the relative quantities has an identical linear correlation with the prevalence's of cancer when the total expenditure/cost per Kg ratio is calculated to give the total quantities/year.

Neutral, Protective, And Causative Food Correlations

The correlations with food categories in terms of expenditure/year are shown in Figures 1 to 4 for neutral, protective, causative and causative/protective foods respectively. In particular, Figure 4 shows the food categories which may be causative for some cancers and protective for others. Figure 1 shows the neutral food categories.

Food category	Type of cancer in 2020								
	Breast	Prostate	Colon	Pancreas	TBL	Bladder M	Bladder F	Uterus	Ovarian
Pork	-0,4383	-0.3506	-0.4294	-0.2850	-0.1268	0.2454	0.0054	-0.2800	-0.0345
Horse meat	0.1267	-0.1020	0.1251	-0.0497	-0.0318	-0.0083	-0.0260	-0.0055	-0.1329
Poultry	0.0034	0.1534	0.3328	-0.3029	0.4025	0.4427	0.0894	0.1529	0.1132
Canned meat	0.2130	-0.0998	-0.0730	0.1701	0.3930	0.1279	0.0102	0.0871	0.1382
Rice	0.1233	-0.1880	-0.1402	-0.0337	0.0627	-0.0308	-0.1927	-0.2726	-0.2456
Fruit preserves	0.1177	-0.1285	-0.3858	0.3058	-0.0410	-0.3911	-0.1233	-0.1551	-0.0736
Salt	0.3184	-0.0909	-0.1175	0.2075	0.1798	-0.1776	-0.2751	0.0648	-0.1392
Soup	0.1505	-0.1489	-0.2510	0.1191	0.1808	-0.0760	-0.0934	0.0660	-0.1309
Seed oil	-0.0239	-0.2053	-0.2375	-0.1152	-0.0685	-0.0126	-0.1687	-0.0418	-0.4008
Lard	0.1944	0.1737	0.0124	0.2698	0.2277	-0.0431	0.2062	0.0302	0.4387
Coffee	0.3657	0.0189	0.3474	-0.0214	0.4177	0.3096	0.1845	-0.0312	0.1322
Tea	-0.0795	-0.4234	-0.2063	-0.0256	0.0835	-0.0354	0.1622	-0.1995	-0.1392
Milk	-0.1080	-0.2792	-0.1484	-0.3735	0.2145	-0.0038	-0.0613	-0.0148	-0.2976
Cheese	0.4284	0.1472	-0.0559	0.2759	0.0630	-0.4397	0.0217	0.1322	0.1395
Citrus fruit	0.3160	-0.0451	-0.2512	0.4399	0.2661	-0.3446	0.1807	0.1098	0.4468
Bananas	0.0169	-0.1134	0.1468	-0.0553	0.2685	-0.0325	-0.0753	-0.2240	0.0374
Legumes F/F [a]	-0.1933	-0.3374	-0.0274	-0.4465	0.1645	0.2621	-0.2373	-0.3453	-0.2331
Vegetables, dried	0.2492	0.0039	-0.2234	0.4293	0.1055	0.0148	0.1567	-0.2603	0.3066
Grapes & strawberries	0.1961	-0.0107	-0.3064	0.3696	0.0298	-0.0132	0.0763	-0.2436	0.2005
Fruit, dried	0.4048	0.0891	-0.1222	0.3749	0.0184	-0.3290	0.2344	0.1348	0.3404
Fruit, canned	0.4209	0.1476	0.0993	0.5136	0.3402	-0.0670	0.2735	0.0837	0.4186
Potatoes	-0.0878	-0.0738	0.1249	-0.3964	0.0616	0.2149	-0.2419	-0.2898	-0.2614
Tomatoes, canned	-0.0540	-0.3672	-0.0094	-0.1063	0.1833	0.2261	0.0546	-0.3567	-0.0823
Mineral water	0.3584	0.1183	0.2617	0.1084	0.4403	0.3291	-0.0121	0.0752	0.2823
Beer	0.0188	0.1018	0.0428	0.0202	-0.2266	-0.4005	-0.3069	-0.1192	-0.2370

Figure 1. Linear correlations between food expenditure/year and cancers: neutral foods, i.e., foods with non-significant r^2 values (±0.456 cut-off)

a = F/F fresh or frozen

Figure 2 shows the protective food categories.

Food category	Type of cancer in 2020								
	Breast	Prostate	Colon	Pancreas	TBL	Bladder M	Bladder F	Uterus	Ovarian
Lamb	-0.5425	-0.1393	-0.1342	-0.4713	-0.4424	0.2908	-0.1064	-0.3826	-0.1957
Pasta	-0.5349	-0.4113	-0.3339	-0.3946	-0.1638	0.2456	0.2444	-0.2682	0.2462
Flour	-0.6621	-0.3392	-0.5216	-0.7009	-0.5986	-0.0213	-0.4963	-0.2819	-0.4120
Sugar	-0.4991	-0.1004	-0.0489	-0.8283	-0.4407	0.2306	-0.4939	-0.2480	-0.5571
Eggs	-0.2995	-0.1609	-0.0530	-0.4539	0.0899	0.3013	-0.0732	-0.1744	-0.1471
Butter	0.2801	-0.0888	-0.3318	0.2834	-0.0088	-0.4771	-0.0894	-0.0511	-0.1467
Margarine	-0.3840	-0.2162	-0.2458	-0.5218	-0.4243	-0.0869	-0.5692	-0.3726	-0.4931
Fish, canned	0.2437	0.0385	-0.2030	0.4519	-0.0581	-0.4595	0.3481	0.0335	0.4235
Legumes, dried	-0.3324	-0.4250	-0.2805	-0.5999	-0.1358	0.3097	-0.2509	-0.3211	-0.4395
Fruit juice	-0.0124	-0.2142	-0.2546	0.0412	-0.0093	-0.5925	-0.2557	-0.1452	-0.1287
NAD [a]	0.3169	-0.0224	-0.0059	0.3751	0.1077	-0.6058	-0.2258	0.1095	-0.1336
Spritis	0.1665	-0.0700	-0.4292	0.3558	-0.0725	-0.4614	0.0337	-0.1957	0.3205

Figure 2. Linear correlations between food expenditure/year and cancers: protective food categories, i.e., foods with significant r^2 values (-0.456 cut-off)

a = non-alcoholic drinks; the protective food categories for each cancer are highlighted in green.

Figure 3 shows the causative food categories.

Food category	Type of cancer in 2020								
	Breast	Prostate	Colon	Pancreas	TBL	Bladder M	Bladder F	Uterus	Ovarian
Beef	0.3568	-0.1799	-0.1846	0.5493	0.4477	-0.3725	0.2257	0.0977	0.2581
Game	0.1496	-0.1956	-0.1649	0.1107	0.3044	0.0885	0.4782	0.0300	0.3809
Processed meat	0.3021	-0.0186	-0.2322	0.5249	0.2450	-0.1734	0.3503	0.2522	0.4504
Bread & breadsticks	0.5699	0.1759	0.2674	0.4001	0.4396	-0.1680	0.1994	0.1427	0.1923
Biscuits	0.3622	-0.0589	-0.0189	0.1798	0.4414	0.2084	0.1870	-0.0979	0.2325
Pastries	0.3555	0.0119	-0.1001	0.4724	0.1302	-0.4182	-0.0586	0.1635	0.0596
Ice cream	0.5627	0.1216	0.1689	0.5992	0.4497	-0.3850	0.2488	0.2007	0.4551
Fish F/ F[a]	-0.1315	0.0985	0.4459	-0.3749	0.2177	0.6460	-0.0491	-0.0369	-0.1431
Olive oil	0.4615	0.0801	0.4965	0.0922	0.7129	0.4041	0.1982	0.3192	0.1412
Powdered milk	0.4380	0.1334	-0.1686	0.6200	0.2009	-0.3464	-0.0497	0.2533	0.2976
Apples	0.3860	0.0401	0.4644	0.2562	0.6882	0.3619	0.4959	0.1975	0.5136
Pears	-0.1331	-0.0842	0.2408	-0.4054	0.2125	0.4734	-0.2360	-0.1953	-0.0741
Fruit, frozen	0.4438	0.3063	0.1410	0.3429	0.3439	-0.1097	0.2571	-0.1752	0.4848
Vegetables F [b]	0.6383	0.0997	0.1700	0.7543	0.6057	-0.2673	0.4232	0.2423	0.3921
Wine	0.6069	0.1776	0.0998	0.8309	0.4442	-0.3692	0.5742	0.4752	0.6576

Figure 3. Linear correlations between food expenditure/year and cancers: causative food categories, i.e., foods with significant r^2 values (+0.456 cut-off)

a = fresh or frozen; b= fresh; the causative food categories for each cancer are highlighted in red.

Figure 4 shows the mixed causative/protective food categories.

Food category	Type of cancer in 2020								
	Breast	Prostate	Colon	Pancreas	TBL	Bladder M	Bladder F	Uterus	Ovarian
Veal	-0.2201	0.2426	0.2650	-0.5237	-0.1053	0.4578	-0.2479	-0.1251	-0.1437
Crustaceans	-0.3312	-0.4715	-0.0783	-0.4860	0.1405	0.5767	-0.0982	-0.3877	-0.3362
Yogurt	0.6091	0.1711	0.0742	0.7208	0.3440	-0.5532	0.3686	0.3184	0.3257
Tomatoes, fresh	-0.1859	-0.0688	0.1683	-0.4919	-0.0562	0.5264	-0.0218	-0.1921	-0.2021

Figure 4. Linear correlations between food expenditure/year and cancers: protective/causative food categories, i.e., foods with significant r^2 values (± 0.456 cut-off)

The causative food categories for each cancer are highlighted in red; the protective food categories for each cancer are highlighted in green.

Total Food Quantities

Table 1 shows the total quantities of the 56 food categories/year. The values are the expenditure/cost/Kg ratio.

Table 1. Quantities in Kg/year derived from total expenditure.

Food category	Kg/year	Food category	Kg/year	Food category	Kg/year	Food category	Kg/year
Milk	209.4	Potatoes	38.4	Pears	17.0	Tea	4.8
Pasta	129.1	Yogurt	38.3	Fruit preserves	16.5	Legumes, dried	4.6
Bread & breadsticks	112.1	Fish F/F [c]	36.3	Grapes & strawberries	15.6	Soup	3.0
Salt	85.0	Apples	35.8	Coffee	13.0	Spirits	2.9
Mineral Water	83.1	Olive oil	33.7	Legumes F/F [c]	13.0	Fruit, dried	2.5
NAD [b]	79.6	Citrus fruit	31.3	Ice cream	11.7	Lamb	2.0
Sugar	56.9	Tomatoes, canned	28.8	Fruit juice	11.3	Fish, canned	2.0
Flour	53.6	Cheese	27.1	Beef	10.9	Meat, canned	1.5
Poultry	49.4	Processed meat	25.6	Seed oil	10.6	Margarine	1.5
Wine	44.7	Rice	21.9	Fruit, canned	9.1	Fruit, frozen	1.4
Vegetables, fresh	42.9	Eggs	21.8	Vegetables, dried	8.3	Horse meat	1.0
Beer	41.5	Veal	19.9	Pastries	7.2	Powdered milk	1.0
Biscuits	39.9	Bananas	19.8	Game	6.4	Crustaceans	0.5
Tomatoes, fresh	38.9	Pork	17.1	Butter	4.9	Lard	0.4

b = non-alcoholic drinks; c = F/F = fresh/frozen.

There were 25 neutral, 15 causative and 12 protective categories; 4 categories were found to be either protective or causative.

In total, the subjects ate 396 kg/year of causative foods and 348 Kg/year of protective foods. The three main foods with values a lot higher than the others were: milk, pasta, and bread and breadsticks, which were neutral, protective and causative respectively (see Figures 1, 2 and 3).

Among the meats, the most eaten was poultry, followed by fish and processed meat.

The most common drinks were mineral water, followed by NAD, wine and beer.

Salt, sugar and flour were among the top ten food categories. Salt was used mostly for boiling pasta, rice and vegetables and cleaning (sometimes for electric dish washing). Only part was used for seasoning food.

The most eaten fruit was apples and citrus fruits.

In terms of fats, olive oil was used three times more than seed oil.

The amounts of potatoes, fresh vegetables, tomatoes were similar, and legumes were used much less than fresh vegetables. There is no correlation between the quantities in Kg/year and the number of cancers, although the categories at the lowest consumption levels (< 5 kg/year) were less causative than categories at higher consumption levels (> 40 Kg/year).

Minimum Spanning Tree (MTS)

The non-stochastic evaluation was carried out using a Euclidean MST (see Appendix).
Cancers were mapped in terms of high and low prevalence. All the values higher than the median belonged to the high prevalence category, while the median and values lower than the median belonged to the low prevalence category.

Figure 5 shows the clusters identified by the MST and the relative connections (nodes) between food categories and cancers.

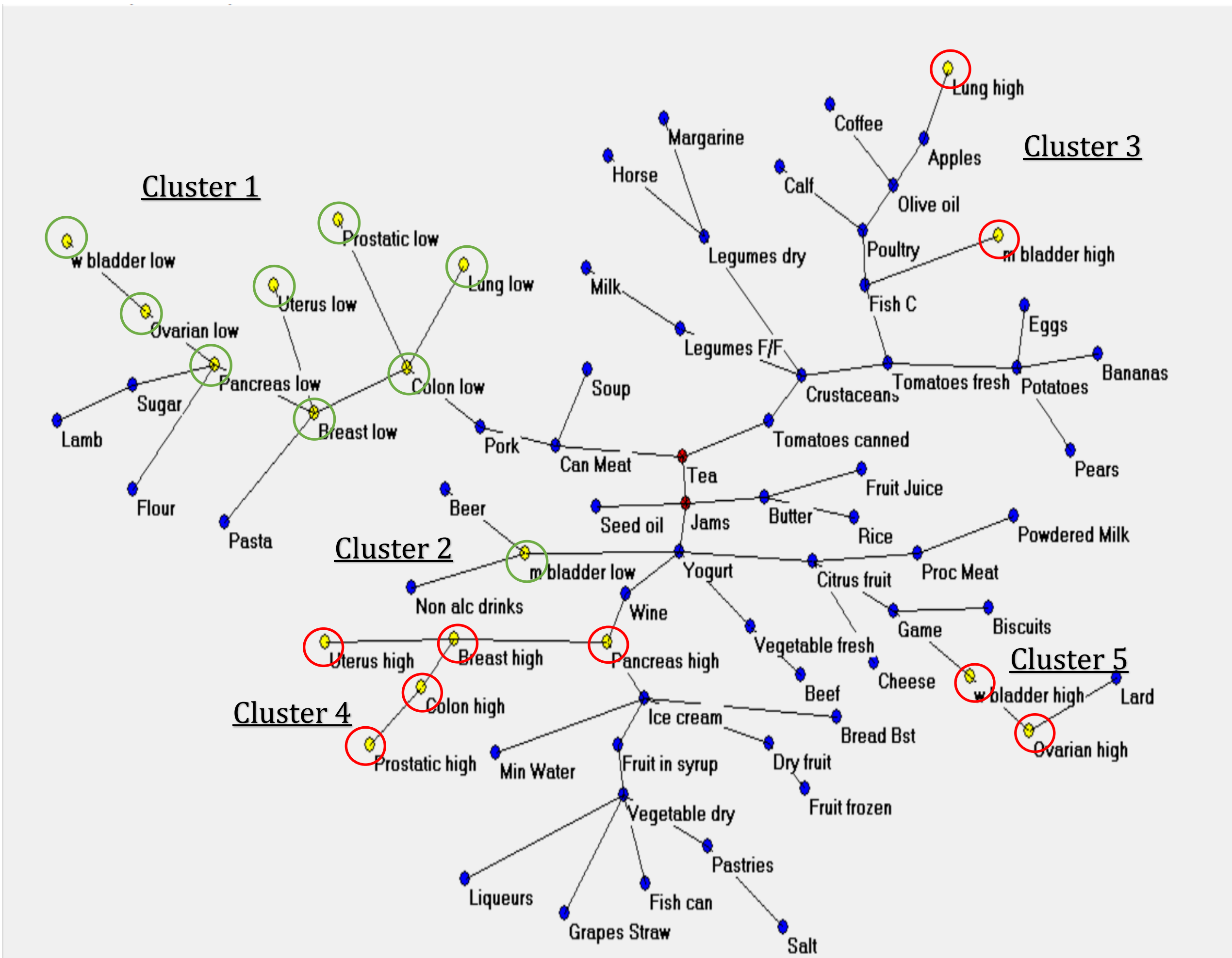

Figure 5. Layout of low and high prevalence of each cancer with the relative close food categories

Key: green circles mark low cancer prevalence, and red circles mark high prevalence.

The MST isolates five clusters: two corresponding to low prevalence; and three corresponding to high prevalence.

Figure 6 shows the food categories close to the low cancer prevalence (green plus signs), which can be considered protective. Figure 6 also shows the food categories close to the high prevalence (red plus signs) which can be considered causative.

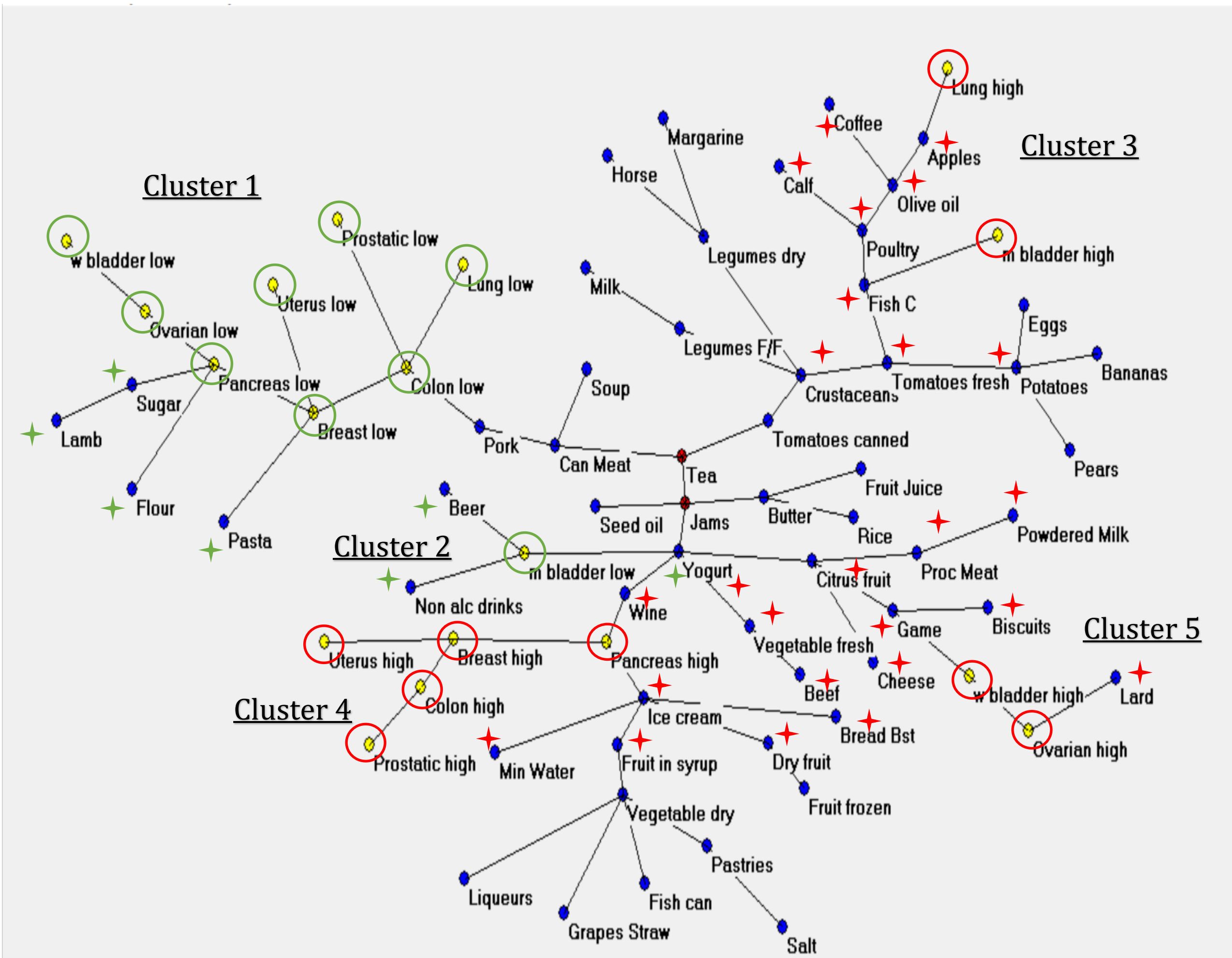

Figure 6. Layout of low and high prevalence of each cancer with the relative close food categories

Key: green circles mark low cancer prevalence, and the food categories close to the nodes are considered protective (green plus signs); the red circles mark high cancer prevalence, and the food categories close to the nodes are considered causative (red plus signs). Yogurt is the only food category with both green and red plus signs.

The protective food categories identified by the MST were fewer than those identified in the stochastic analysis (7 instead of 14). Only the protective food categories identified both by the stochastic and non-stochastic analyses were considered (Figure 7). The nodes close to cancer which were not significantly correlated in the stochastic analysis were not considered variables in common between the two analyses. The protective variables in common between the two analyses in the MST were concentrated in Clusters 1 and 2 (Figure 6).

There were more causative foods in the MST than in the stochastic analysis (25 instead of 15). As for protective foods, only the food categories in common between the two analyses were considered (Figure 8). They are found in clusters 3, 4 and 5.

Figure 7 shows the protective food categories in common between the stochastic and non-stochastic analyses.

Food category	**Type of cancer in 2020**								
	Breast	**Prostate**	**Colon**	**Pancreas**	**TBL**	**Bladder M**	**Bladder F**	**Uterus**	**Ovarian**
Lamb	+	+							+
Pasta	+	+		+	+		+	+	
Flour	+	+							
Sugar	+	+	+	+	+		+	+	+
Yogurt						+			
NAD						+			

Figure 7. Protective food categories in common between stochastic and non-stochastic analyses according to cancer

The most common protective food categories were sugar (protective in nine types of cancers), followed by pasta (protective in six cancers). They are found in clusters 1 and 2.

Figure 8 shows the causative food categories in the stochastic and non-stochastic analyses.

Food category	**Type of cancer in 2020**								
	Breast	**Prostate**	**Colon**	**Pancreas**	**TBL**	**Bladder M**	**Bladder F**	**Uterus**	**Ovarian**
Game							+		
Fish F/ F[a]					+	+			
Olive oil					+	+			
Yogurt	+			+					
Wine	+	+	+	+				+	

Figure 8. Causative food categories in common between stochastic and non-stochastic analyses according to cancer

The causative foods in common between the two analyses are found in clusters 3, 4, and 5. Wine was seen to be an independent causative food category in five cancers. Olive oil and fish F/F were causative for two cancers (TBL and bladder cancer in males), and lastly game was seen to be causative only for bladder cancer in females.

Cancer Prevalence From 2006 To 2020

Table 2 shows the prevalence of the nine cancers for the period from 2006 to 2020.

Table 2. Prevalence per 10^4 inhabitants of the nine cancers from 2006 to 2020

Year	Type of cancer								
	Breast	Prostate	Colon	Pancreas	TBL	Bladder M	Bladder F	Uterus	Ovarian
2006	3.82	2.64	3.10	1.64	5.61	1.46	0.38	0.80	0.98
2007	3.94	2.64	3.13	1.66	5.62	1.54	0.37	0.79	1.08
2008	4.01	2.61	3.14	1.60	5.61	1.53	0.39	0.80	1.05
2009	3.99	2.61	3.13	1.64	5.64	1.52	0.38	0.82	1.10
2010	3.95	2.61	3.18	1.58	5.67	1.56	0.40	0.79	1.04
2011	3.90	2.62	3.21	1.69	5.68	1.54	0.39	0.81	1.05
2012	3.89	2.52	3.22	1.69	5.61	1.55	0.41	0.79	1.05
2013	3.83	2.46	3.11	1.62	5.54	1.50	0.39	0.81	1.06
2014	3.89	2.43	3.06	1.62	5.47	1.48	0.40	0.80	1.00
2015	3.93	2.44	3.12	1.59	5.57	1.50	0.39	0.81	1.02
2016	4.03	2.55	3.22	1.98	5.56	1.65	0.41	0.84	1.04
2017	4.11	2.61	3.20	2.04	5.58	1.65	0.45	0.86	1.07
2018	4.20	2.61	3.20	2.05	5.58	1.59	0.43	0.86	1.09
2019	4.15	2.62	3.24	2.13	5.48	1.62	0.43	0.85	1.07
2020	4.29	2.72	3.18	2.17	5.41	1.62	0.45	0.86	1.07

Comparing the 2006 data with the 2020 data, only the prevalence of TBL cancer decreased slightly, while the prevalence of all the other cancers increased (U Mann-Whitney $p < 0.05$).

Comparison Between Geographical Areas

The MST based on food expenditure and the nine cancers shows that the pattern of the Italian regions is highly consistent with their geographical location. This means that Northern Italy Cluster 1 is well separated from Southern Italy, apart from Lazio and Campania which are geographically part of the South, but in the calculations are considered as lying within Cluster 2. The two clusters are characterized by different food patterns.

The concept of geometrical structure is connected with the concept of distance. Specifically, to preserve the geometrical structure of the original data after applying the mapping algorithm, the elements that used to be close in the original space have to be close in the sub-space too. This also means that distant elements in the original space still need to be separated in the target space. If it is important to preserve the specific metric of the original space, the projection technique should aim at isometry between the original space and the resulting map.

To achieve this, we have projected it onto a lower dimensional space, in much the same way as we did for food and cancers, on the basis of a simple transposition of the dataset. We discovered that our technique was successfully able to cluster regions according to their geography to a surprising level of accuracy, thus validating our analysis in terms of robustness.

Figure 8 shows the MST of the geographical distribution of food categories.

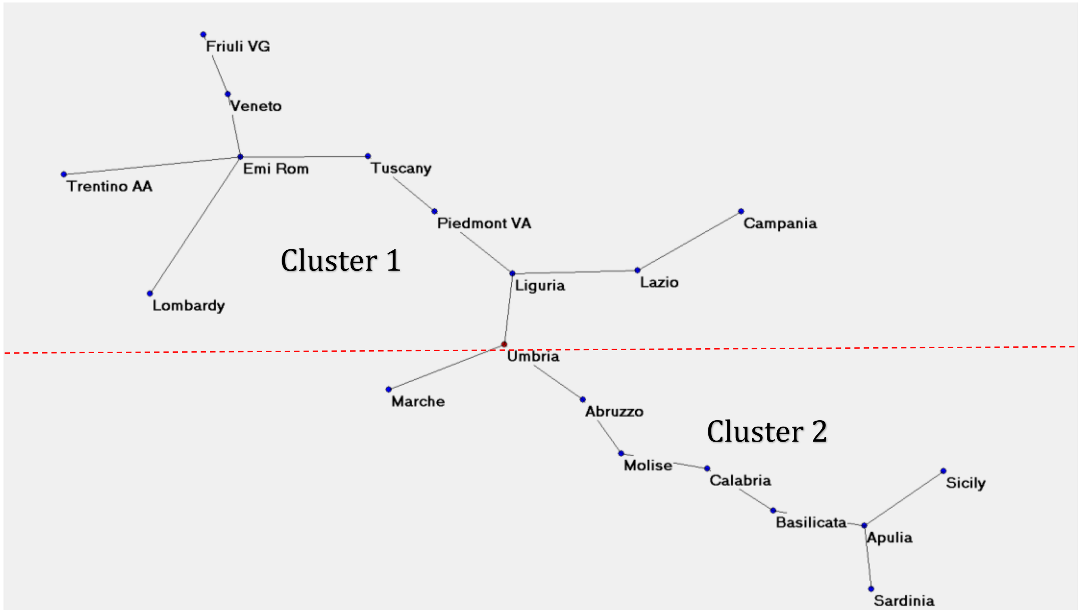

Figure 8. MST that considers the Euclidean distance between the food patterns in each Italian region

Key: the red node is the mathematical centre of the MST and corresponds to Umbria, which is the actual Euclidean centre of Italy from a geographical point of view. The dotted red line separates the two clusters.

The concept of geometrical structure is connected with the concept of distance. Specifically, to preserve the geometrical structure of the original data after applying the mapping algorithm, the elements that used to be close in the original space have to be close in the sub-space too. This also means that distant elements in the original space still need to be separated in the target space. If it is important to preserve the specific metric of the original space, the projection technique should aim at isometry between the original space and the resulting map. To achieve this, we have projected the regions onto a lower dimensional space, in much the same way as we did for food and cancers, on the basis of a simple transposition of the dataset. We discovered that our technique was successfully able to cluster regions according to their geography to a surprising level of accuracy, thus validating our analysis in terms of robustness.

Table 3 shows the prevalence of cancer in Northern and Southern Italy.

Table 3. Prevalence of cancer per 10^4 inhabitants in the regions of Italy for the year 2020

Region	Type of cancer								
	Breast	Prostate	Colon	Pancreas	TBL	Bladder M	Bladder W	Uterus	Ovarian
Northern Italy	4.31	2.69	3.44	2.29	5.88	1.59	0.45	0.88	1.16
Southern Italy	3.63	2.61	3.26	1.79	4.88	1.68	0.42	0.79	0.93
U Man Whitney	<0.05	ns	ns	<0.05	<0.05	ns	ns	ns	<0.05

TBL: throat-bronchial-lung; M = males; F = females.

The prevalence of cancers is higher in the northern regions, than in the southern regions, apart from bladder cancer in males, which is slightly lower, although this difference is not statistically significant.

Twenty-six food categories were significantly different when the Northern and Southern regions were compared.

Twenty-six foods categories were significantly different comparing the Northern and Southern Regions.

The results are summarized in Table 4.

Table 4. Significantly higher (protective) or lower (causative) in South of Italy compared to North of Italy Regions: p value of Mann Whitney U test for cancer prevalence

Cancer	Foods categories Protective higher	Foods categories Protective lower	Foods categories Causative higher	Foods categories Causative lower	P
Breast	Lamb, Pasta, Flour, Sugar	None	None	Yogurt, Wine	< 0.05
Prostate	Lamb, Pasta, Flour, Sugar	None	None	Wine	NS
Colon	Sugar	None	None	Wine	NS
Pancreas	Pasta, Sugar	None	None	Yogurt	< 0.05
TBL	Pasta, Sugar	None	Fish	None	< 0.05
Bladder M	None	Yogurt	Fish	None	NS
Bladder F	Pasta, Sugar	None	None	Game	NS
Uterus	Pasta, Sugar	None	None	Wine	NS
Ovarian	Lamb, Sugar	None	None	None	< 0.05

Southern Italy regions are characterized by a significant higher expenditure of four protective food categories represented by lamb, pasta, flour and sugar.

As regards the causative categories, the expenditure was higher for fish F/F only, and lower for wine the remaining categories were not different. Expenditure on yogurt (causative for breast cancer and protective for bladder cancer in males) was significantly lower.

Causative variables such as olive oil and NAD were not significantly different in Southern compared to Northern regions. A higher consumption of pasta/sugar or sugar only has no effect on prostate, uterus and bladder cancer in males and females. This means that factors other than pasta and sugar are decisive for cancer development. Data do not change in the case of 15 % average cost reduction.

DISCUSSION

This study considers the Italian diet as it was in 2016 based on food expenditure.

The sample includes 540 municipalities (6.3% of the total of 7904), which ISTAT considers the standard figure to measure all demographic, economic, social and health variables in Italy. Food expenditure is also one of the standard measurements.

This study has several limitations, and we will try to identify those which may affect the results. The number of cases analysed was 19500 for each gender, aged between 24 and 54. There was also a percentage of much younger family members: children between 0 and 14 years old, who accounted for about 14% of the total.

These younger people are exposed to foods which are not completely covered by the family expenditure, because they eat some of their meals at school. In the end, the sample corresponds to about 86% of the total figure recorded by ISTAT. Nevertheless, the number of subjects in this study is at least one order of magnitude larger than in the usual prospective studies and represents the population of Italy.

Some of the most advanced, industrialized regions of Europe are found in Italy (such as Lombardy and Veneto), as well as some of the most rural (such as Molise and Basilicata).

Therefore, this data cannot be considered as valid for other countries that follow the MeD.

- There are a lot of foods available on the market, and we took the 56 categories most sold in Italy in 2016. These categories contain more than one type of food, and each of the 540 municipalities considered may prefer several different foods despite them all belonging to the same category.
  Furthermore, products sold in one region may be different from those sold in other regions, and sometimes foods are imported from countries outside Italy.
- The prices of the food categories were not the same in every region, and some foods may cost less than average, particularly in Southern Italy. The impact of these differences may be important, and despite some correction (a 15% cost reduction for some food categories in Southern Italy), it is hard to calculate the precise amount of food consumed. The quantity of food in terms of kg/year was calculated using the expenditure/cost per kg ratio.
- Only linear correlations were calculated, and this may not be the case for some of the food categories analysed [16]. The curves were carefully analysed visually, to pick up outliers and non-linear correlations.
- Foods may be purchased by one person and eaten by another. Since this is possible, this data has to be considered with caution.
- A common major bias in all epidemiological studies based on food categories is the differences within the categories. For example, cheese can be different in each municipality, and even meat, vegetables and fruit.
  Furthermore, production quality control is not usually taken into account.
  Italy is in charge of food safety in Europe, and the food quality control here is very efficient, also in terms of the relative compositions and biological/non-biological origin.

However, checks in rural areas may be more problematic and may fall short of the required standards.

In the case of meat sold in stores, the origin, rearing and slaughterhouse are usually shown on the labels. In the case of dairy, poultry, pasta, rice, bread, oil, and processed meat, the same information is also on the labels, although not complete for some aspects which may not be controllable.

These checks are almost impossible in the case of self-produced foods (e.g., processed meat, fruit and vegetables). However, the families that produce their own food are a very small percentage of the total figure and have a minimal effect on the overall picture.

- In terms of statistical evaluation, classic epidemiological studies are based on a large number of replications corresponding to the number of subjects studied (usually > 1000). The power of the analyses used to determine the effects (in terms of α and β errors) is very high. Consequently, a 5% higher risk of any disease ends up being statistically significant. In other words, the *true values* of these analyses are the risks. In this study, there were only 19 replications (the number of Italian regions), and the power of the stochastic analysis is much lower than it would be in epidemiological or cohort studies.

  In these conditions, a non-stochastic analysis (MST) gives more information about the relationship between food categories and the prevalence of cancer. Combining this with the stochastic analysis may allow a more precise evaluation.

  Another strength of this study is the analysis performed by applying an MST to a transposed dataset with regions as variables and foods as records.
- The food expenditure data refers to 2016 and the cancers were calculated in 2020.

  Four years may be a short evaluation period, because many cancers take longer to terminate in death. However, considering the prevalence data from 2006 to 2020, all the cancers were highly correlated, and this suggests that these correlations will also be valid in the future. This is an important aspect which may reduce the error caused by the presence of subjects aged < 24 years in the sample.

  One intriguing aspect emerges from the mortality data, and concerns COVID-19, which started in January 2020. The number of deaths from cancer in 2019 does not seem to be very different from that in 2020. This suggests that the effect of COVID-19 was not very large.
- We used prevalence of death as the variable to correlate instead of prevalence of disease because death is the soundest variable for any disease. The distinction between death due to the disease and death with the disease is a bias found in all studies concerning prevalence data, which stems from the records available.
- The diet correlates from year to year. However, some differences were evident in 2016 compared to the past, and this aspect may represent a bias in the future too.

  Although the various regions in Italy keep as much as possible to traditional ways of using and preparing foods, in the long run, this may change, as it appears to have done between 2004 and 2016 [6].
- There are several variables responsible for cancers, and some are unknown. In this study, causative variables, such as smoking, body weight, sedentariness and pollution were not considered and may be important in the case of differences between regions.

  Despite the several limitations of this study, some of the information is very interesting.

General Considerations

The Current Diet in Italy:

The MeD is usually reported to be an anti-inflammatory, antioxidant diet based on minimally processed whole grain bread, other cereals and legumes, extra virgin olive oil (EVOO), and low dairy food and wine consumption [5]. In the cohort studies done in the 1960s [6], the key foods in Italy were cereals, followed by vegetables, animal by-products (cheese, milk and eggs) and fruit.

The data taken in 2016 were quite different, and show that milk is the most common food, followed by pasta and bread.

Poultry and fish were found to be the main source of meat, followed by veal and processed meat, while potatoes, fresh vegetables with large quantities of tomatoes were the most common foods for side dishes. Legume consumption was relatively low, at about 30% of fresh vegetables. Olive oil is the main source of fat and is three times more common than seed oil.

The prevalent beverages were mineral water and non-alcoholic drinks, followed by wine and beer.

Among the animal by-products, the most consumed was yogurt, followed by cheese and eggs. Fruit consumption mainly consisted of apples, citrus fruit, bananas and pears.

Should it still be considered a typical MeD?

Our data show that the diet has changed in Italy, although these changes may have taken many years to occur. However, the consumption of pasta, sugar and flour (protective food categories), and fish and wine (causative food categories) are not decreasing; they are either stable or increasing.

Food Categories and Cancers:

Stochastic and non-stochastic analysis (MST) showed that only 11 food categories may be protective or causative, and those used in relatively small quantities (< 5 kg/year, equivalent to < 100g/week) amount to 25% of the total.

In terms of positive or causative activity, apart from lamb, only quantities > 5 kg had an effect on cancer. With the exception of lamb, all the other ten causative or protective food categories were consumed in quantities > 6 kg/year, and 9 exceed 21 kg/year. In other words, quantity seems to be important for protective and causative effects.

Protective Food Categories:

Twelve protective categories were identified by the stochastic analysis. These were many more than those that emerged from the MST, and - in the end - there were only six categories in common according to each cancer (Table 7).

Surprisingly, sugar was found to be protective in eight cancers out of nine.

It is commonly believed that “sugar fuels cancer” [17], because the metabolism of cancer cells involves the consumption of large amounts of glucose, and subjects suffering from type-2 diabetes have been shown to be at high risk of pancreatic and uterine cancer [18]. Therefore, dietary restriction of carbohydrates and pharmacological agents able to reduce insulin production or inhibit insulin signals have been exploited as potential therapeutic methods for cancer prevention or treatment. Nevertheless, meta-analyses of prospective studies have widely reported the lack of association between cancer incidence and carbohydrate intake, even when factors indicative of food ability to increase post prandial glucose concentration, such as glycemic index and glycemic load have been considered [19]. In addition, the World Cancer Research Fund and American Institute of Cancer Research concluded in its 2018 report that there was not sufficient evidence for an association between sugar and cancer [20]. More recently, sugar-sweetened beverages have been associated with poorer overall survival in bladder cancer [21] and a prospective study suggested that sugar may be a risk factor for cancer, in particular for breast cancer, regardless of weight gain and weight status. However, Table 2 of this study [22] shows the following data (see our Table 5).

Table 5. Quartiles of additional sugar intake. Number of participants, incidence of cases of cancer and p values of the trends; taken from Table 2 of [22]

	Quartiles								
Cancer	**Q1 Part [a]**	**Q1 Cases [b]**	**Q2 Part**	**Q2 cases**	**Q3 Part**	**Q3 cases**	**Q4 Part**	**Q4 cases**	**P trend**
Breast	19,935	203	19,936	226	19,936	182	19,935	172	0.02
Prostate	5384	108	5384	91	5385	75	5384	49	0.3
Colorectal	25,319	72	25320	54	25,321	40	25,319	16	0.03
Lung	25,319	46	25,320	36	25,321	23	25,319	17	0.9

A = participants; b = cases

It is hard to believe that sugar consumption increases the risk of breast cancer, since the number of participants was the same, the period was almost identical, and the number of cases was higher in Q2 only, while it was lower in Q3 and Q4 compared to Q1. It seems that sugar consumption reduces the incidence of all four cancers or, at worse, is neutral.

However, the average daily quantity of sugar consumed in our study was 67.2 g/day, compared to 89.7 g/day in the Nutrinet-Santé cohort [22], and this difference may have an effect on the results.

As far as our results are concerned, the sugar consumed in Southern Italy was 27.6% higher than in Northern Italy (65.6 g/day and 51.4 g/day respectively: U Man Whitney $p<0.01$) and the prevalence of breast cancer was significantly lower (3.63 and 4.31 respectively: U Man Whitney $p<0.01$). The same was true for pancreas, TBL, and ovarian cancers (see Table 3).

Although the prevalence of the other cancers (colon, uterus, and bladder cancer in males and females) was lower, the difference was not statistically significant. This suggests that we may have to consider other variables (e.g., insulin resistance, environment, pollution, lifestyle, age at first pregnancy etc.) which are important co-factors, along with genetic variables. Some randomized Mendelian studies suggest that genetic variables may determine the risk for most cancers [18].

The other most common protective food category was pasta (in six cancers out of nine).

The available literature [23] tends to see pasta as having neutral activity in the case of breast cancer, and causative activity in the case of colorectal cancer in women.

The latter study was conducted in Italy, and although the daily amount of pasta was not reported, the highest quintiles should be similar or even lower than the average daily quantity of 153 g/day in this study.

Pasta (white or integral) is a healthy food. The carbohydrate and protein content remains almost intact since it is prepared at very low temperature. It is only boiled before use. It is also free from any carcinogenic compounds (e.g., acrolein) which may stem from burning during preparation.

Lamb was shown to be protective for breast, prostate and ovarian cancers. Two aspects should however be considered. The first is the quantity/week (calculated considering the expenditure/year in relation to the cost per Kg). It is lowest among the meats (< 20 g/week, specifically 408 g/week for poultry and 90 g/week for beef), and low values may cause correlations to lose power. The second is that lamb consumption was much higher than average in the regions where the prevalence of breast and ovarian cancer were lower than average (Abruzzo, Molise and Basilicata).

The effects of other protective food categories, such as yogurt and non-alcoholic drinks, were limited to bladder cancer.

The protective effect of non-alcoholic drinks (NAD) has no clear explanation unless the sugar content is considered. The total quantity/year of NAD was very close to that of mineral water, and its effect was limited to bladder cancer in males only, where sugar consumption was not found to be protective. However, daily consumption in the NAD category is about 100 ml. This amounts to 8.5 g of sugar/day, which may have a positive impact.

Dairy Products:

Milk is the most consumed food in Italy, in a quantity that is about 60% higher than pasta and bread.

According to these data, milk seems to be very important in the Italian MeD (209.4 Kg/year). It was seen to be completely neutral for all the cancers considered in this study. Milk has never been reported to be an important part of any MeD. It may therefore be part of a change in diet during the years.

According to the EFSA, milk has to be checked for aflatoxin M1 content (limit of 50 ng/L [24]), while - to our knowledge - this measurement is not required for yogurt.

Yogurt was found to have a dual effect: protective for bladder cancer in males and causative for breast and prostate cancers. Similar dualism has been described by other authors [25].

The presence of M1 has been documented in Italy [26, 27] and stricter checks should be carried out on yogurt. Moreover, yogurt is a fermented food known to have a high probiotic content, such as *Lactobacillus* and *Bifidobacterium*, which are not considered free from side effects, such as gene transfer, excessive immune stimulation, and microbiota alterations. In other words, it may have a positive effect on cancer development for some subjects, while it may have the opposite effect for others.

Yogurt also showed a dual effect in our study, since it was causative for breast and pancreas cancers (one of the strongest correlations, r = 0.7208), and protective for bladder cancer in males.

In a previous meta-analysis study [28], yogurt and milk were not associated with an increase in breast cancer, while yogurt was protective in another study [29].

However, previous studies [21] have shown that more than one food may be involved in cancers, and experimental models have shown the antitumoral effects of various components of fruits, vegetables, and soy products. It may be that different combinations of foods end up giving contradictory results. The evidence that bladder cancer in females has a completely different food pattern than in males, suggests that foods may only be a concomitant cause among several causative factors that produce cancer and that these are different in males and females.

Causative Food Categories:
Five causative food categories were found to be in common between the stochastic and non-stochastic analyses, depending on the cancer.

Wine was the category most frequently found to be causative (5 cancers). However, no data were available for white and red wine.

Alcohol is commonly found to be a cause of cancer in many epidemiological and cohort studies, and this may confirm the validity of this study. The most probable reason is that the alcohol contained in wine is a precursor of carcinogenetic acetaldehyde [30-32].

Apart from mineral water, wine was the most consumed drink in Italy (44.7 Kg/year or 45.2 L/year), followed by beer (41.5 Kg/year) and then spirits (2.9 Kg/year). In terms of alcohol, it is evident that the highest consumption derives from drinking wine (causative), followed by beer, and then spirits (neutral).

The conclusion is that “the dose makes the poison”, exactly as Paracelsus said.

Fish:
In general, fish is considered a healthy product and protective for many cancers [33], and pescetarians (also vegetarians) have been shown to be at lower overall risk for cancer, in particular as regards colon cancer [34].

This was not what we found, since fish was causative for TBL and bladder cancer in males.

It should be considered that about 30% of fishery products in Italy are farmed fish, and the way they are produced is not highly publicized. Fish farms use medications, such as antibiotics, to prevent or treat diseases which can affect fish, and they also use several chemicals to maintain the water quality in the tanks, nets or cages. These consist of oxidants, coagulants, osmoregulatory, algicides, herbicides, disinfectants, anaesthetics, agricultural pesticides, and also hormones [35-36].

Furthermore, the products used in aquaculture are not listed among the mandatory information required on the labelling under the EU regulations.

Olive Oil:

The causative effect of olive oil in two cancers (TBL and bladder in males) was unexpected. However, it was not possible to distinguish between extra virgin olive oil (EVOO) and other olive oils from the expenditure data.

This difference may be important. Common labelling does not mention the antioxidant content, which is very high in EVOO, and the low degree of acidity required by the labelling can easily be achieved through various forms of correction, which may turn a very poor-quality olive oil into EVOO. On the contrary, seed oil, which has a much lower antioxidant content than EVOO, was completely neutral for all cancers.

Although the cost of EVOO should be much higher than that of seed oils, sometimes EVOO is sold at very low prices, which is not compatible with good production standards.

Recently, the EFSA has reviewed the maximum levels of MOSH (mineral oil saturated hydrocarbons) and MOAH (mineral oil aromatic hydrocarbons) in EVOO, which were fixed at between 1-15 mg/Kg in 2018. These substances are considered mutagenic/carcinogenic, and the current levels in EVOO are not considered safe. The EFSA will shortly lower the allowed limits.

However, regardless of their origin (olives or seeds), oils are fats, and inappropriate consumption can lead to inflammation, irrespective of their composition in terms of mono/polyunsaturated fatty acids.

Moreover, unsaturated oil metabolism generates DNA adducts, which can be mutagenic/carcinogenic [37, 38]. One more aspect of EVOO is important: a lot of food adulteration in Italy concerns olive oil and strict controls are necessary.

Meat and Game:

Meat and processed meat were found to be causative of many cancers, and the literature is extensive [39, 41]. None of the meats in our study were found to be causative, apart from game in the case of bladder cancer in women. There is no explanation for this. The average consumption/week is about 53 g and after lamb (protective), game was the least consumed. As for lamb, the small quantities consumed may cause it to lose its discriminating capacity.

The real surprise was that none of the other meats were generally found to be causative. On the contrary, pork was found to be protective in the cases of colon and breast cancers in the non-stochastic analysis since the correlations were negative (-0.438 and -0.429 respectively) and close to the statistically significant cut-off (-0.456).

Difference Between Northern and Southern Regions

The finding that the prevalence of cancer in Northern Italy is higher than in the Southern regions cannot be denied. It is evident that Southern and Northern Italy have different eating habits in both stochastic and non-stochastic analyses. Twenty-six of the 56 food categories were significantly different, and the quantities of the four protective foods consumed (lamb, pasta, flour and sugar) were higher in Southern Italy. Consumption of wine (causative) was also significantly lower than in Northern Italy. Yogurt consumption, which was found to be causative or protective (pancreas and bladder cancer in males respectively), was significantly lower, while the intake of fish (causative) was higher. All the remaining food categories were not significantly different. However, cancers have many causes, from genetic predisposition to the environment, and foods may only be considered co-factors.

CONCLUSIONS

The relationship between food expenditure and cancer was analysed in an attempt to find a method to obtain immediate, valid information on foods which should be commonly consumed. Comparing the food pattern of the 1960s with the most recent 2016 pattern, it is evident that the diet in Italy has changed, and it is doubtful if it can still be considered a MeD.

Milk has become by far the most common food category, and besides olive oil, bread, pasta, fresh vegetables and fruit, many other foods are emerging, such as processed meat, cheese and eggs.

The stochastic analysis showed that pasta, flour and sugar were protective for some cancers and never causative, as is instead the case of lamb.

Wine, on the other hand, was causative in many cancers, and surprisingly so were olive oil and fish for some cancers.

Our data do not correspond to the common findings based on epidemiological or cohort studies. For this reason, more detailed analysis is necessary based on a larger dataset.

One common element in all the studies, and in our study too, is that foods can be important co-factors in cancer development.

We must not forget that excessive weight and the consequent inflammatory condition are important causes of cancer, and these conditions are determined by diet, regardless of whether it is based on causative or protective food categories.

In the end, while awaiting more precise evaluation by the public health system, our suggestion is that diets should be based on small quantities of all foods, and a wide variety of different food categories is very helpful to live longer and healthier lives.

Acknowledgements

We are grateful to COOP supermarkets (Italy) for providing the data and their help, to ISTAT for providing the original data and death records, and lastly to the WHO for providing the ASDR data on breast cancer deaths. Without the data provided by these three bodies, this study would not have been possible.

We are also grateful to the 26 students of the Food Toxicology course at the Fondazione Istituto Tecnico Superiore per le Tecnologie della Vita, Bergamo, Italy, who retrieved the data provided by COOP supermarkets.

Authors

UC conceived the research and wrote the text, and MR and EG were responsible for the stochastic and non-stochastic analyses respectively. BC, MR, CR, EP, controlled the nutritional data and the text.

All the authors agreed with the data analysis and with the text prepared.

Conflict of interest

There were no conflicts of interest.

Funding

There was no funding.

Ethical Committee

No ethical committee was required for this study.

Data Availability

All the data are available upon reasonable request from the corresponding author:
Umberto Cornelli: Piazza Novelli 5 20129 Milan - Italy: email; ucornelli@gmail.com

References

[1] Keys A. Keys M. How to eat well and stay well the Mediterranean way. New York: Doubleday. 1975.

[2] Noach A. Trustell AS. There are many Mediterranean diets. Asia Pacfic J Clin Nutr 2001;10 (1): 2-9.

[3] Nestle M. Mediterranean diets: historical and research overview. Am J Clin Nutr 1995; 61: 1313S-1320S

[4] Ferro-Luzzi A. Branca F. Mediterranean diet. Italian style: prototype of a healthy diet. Am J Clin Nutr 1995; 61: 1338S-1345S.

[5] Esposito K. Maiorino MI. Bellastella G et al. A journey into a Mediterranean diet and type 2 diabetes: a systematic review with meta-analyses. BMJ 2015: 5:3008222 doi: 10.1136/bmjopen-2015-008222.

[6] Alberti-Fidanza A. Fidanza F. Mediterranean adequacy index of Italian Diets. Public Health Nutr 2004;7: 937-941.

[7] Maroni P. Bendinelli P. Fulgenzi A. Ferraretto A. Mediterranean diet food component as possible adjuvant therapies to counteract breast and prostate cancer progression to bone metastasis. Biomolecules 2021; 11: 1336: 1-18.

[8] Di Maso M. Dal Maso L. Augustin LSA et al. Adherence to the Mediterranean diet and mortality after breast cancer. Nutrients 2020; 12: 3649-3654.

[9] Mentella MC. Scladaferri F. Ricci C Gasbarrini A. Miggiano GAD. Cancer and Mediterranean diet: a review. Nutrients 2019: 11: 2059; doi:10.3390/nu11092059.

[10] Detopoulou P. Demopoulos C A. Antonopoulou S. Micronutrients. phytochemicals and Mediterranean diet: a potential protective role against COVID-19 through modulation of PAF actions and metabolism. Nutrients 2021. 13.462. https//doi.org/10.390/nu13020642.

[11] Mokbel K. Mockbel K. Chemoprevention of breast cancer with vitamins and micronutrients: a concise review. In vivo 2019; 33: 983-997.

[12] Yikiung P, Dodd KW, Thompson FE et al. Comparison of self-reported dietary intake from automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018; 107 (1):80-93

[13] ISTAT 2019 Mortalità per territorio di evento- Italy (Mortality by event territory).

[14] https://www.istat.it/it/archivio/71980.

[15] Le 1000 tavole degli Italiani; italiani.coop/le-1000-tavole-italiane.

[16] Kazemi A. Barati-Boldaji R. Soltani S et al. Intake of various foods groups and risk of breast cancer: a systematic review and dose response meta-analysis of perspective studies. Adv Nutr 2021; 12: 809-849.

[17] Penson RT. Sugar fuels cancer. Cancer 2009; 115 5 : 918-921.

[18] Yuan S, Kar S, Carter P et al. Is type 2 diabetes casually associated with cancer risk? Evidence from a two-sample Mendelian Randomization study. Diabeters 2020; 69: 1588-1956.

[19] Song M. Sugar intake and cancer risk: when uncertainty meets biological plausibility. Am J Clin Nutr; 112 (5): 1155-1156.

[20] World Cancer Research Fund/American Institute of Cancer Research. Diet, nutrition, physical activity and cancer: a global perspective, continuous update project report 2018. Recommendations and public health and policy implication. London: WCRF/AICR; 2018.

[21] Leeming RC, Karagas MR, Gilbert-Diamond D et al. Diet quality and survival in a population-based bladder cancer. Nutr Cancer 2022; 74 (7): 2400-2411.

[22] Debras C, Chazelas E, Srour B et al. Total and added sugar intakes, sugar types, and cancer risk: results from prospective NutrientNet-Santé cohort. Am J Clin Nutr 2020; 112: 1267-1279.

[23] Augustin LSA, Malerba S, Lugo A et al. Association of bread and pasta with the risk of cancer of breast and colorectum. Annals Oncol 2013; 24(12): 3094-3099.

[24] EFSA Panel on Contaminants in the Food Chain (CONTAM). Risk assessment of aflatoxin in food. EFSA Journal 2020 doi: 10.2903/j.efsa.2020.6040.

[25] Zhang K, Dai H, liang W et al. Fermented dairy foods and risk of cancer. Int J Cancer 2018; 144: 2099-2018.

[26] EFSA Reg UE 165/2010 GU CEE L50/8 27/02/2010.

[27] Galvano F. Pietri A. Bertuzzi T et al. Maternal dietary habit and mycotoxin occurrence in human mature milk. Mol Nutr Food Res 2008; 52: 496-501.

[28] Lu Chen MM. Min Li MM. Hao Li MM. Milk and yogurt intake and breast cancer risk a meta-analysis. Medicine 2019; 98: 12(e14900).

[29] McCann SE. Hays J. Baumgart CW et al. Usual consumption of specific dairy foods is associated with breast cancer in the Roswell Park Cancer Institute Data Bank and BioRepository. Curr Dev Nutr 2017;1:1-6.

[30] Boffetta P, Hashile M. Alcohol and cancer. Lancet Oncology 2006; 7. 149-146.

[31] Seitz MK, Tehard B, Matsuaki S, Yokoyama A et al. Alcohol and cancer. Alcohol 2001: 25: 137S--143S.

[32] Druesne-Pecollo M, Tehard B, Mallet Y et al. Alcohol and genetic polymorphism: effect on risk of alcohol-related cancers. Lancet Oncology 2009; 10:173-180.

[33] Fernandez E. Chatenoud L. La Vecchia C et al. Fish consumption and cancer risk. Am J Clin Nutr 1999; 70 (1): 85-90.

[34] Parra-Soto S. Ahumada D. Peterman-Roha F et al. Association of meat, vegetarian, pescatarian and fish-poultry diets with 19 cancer sites and all cancers: finding from UK Biobank prospective cohort study and meta-analysis. BMC Medicine 2022; 20:79 ;doi.org/10.1186/s12916-022-02257-9.

[35] Wang Q, Li Z, Gui JF et al. Paradigm changes in freshwater acquacultutre practice in China: moving toward achieving environmental integrity and sustainability. Ambio 2018; 47 (4): 410-426.

[36] Boyd CE, McNevin AA. Chemicals in acquaculture practice in China. 2014 doi.10112/9781118857195.ch9.

[37] Chung FL. Pan J. Choudhury S et al. Formation of trans-4-hydroxy-2-nonenal-and other enal-derived cyclic DNA adducts from omega-3 and omega-6 polyunsaturated fatty acids and their role in DNA repair and human p53 gene mutation. Mutat Res 2003; 531 (1-2): 25-36.

[38] Basu AK. DNA damage, mutagenesis and cancer. Molecular Science 2018; 19: 970: doi: 10.3390/ijms19040970.

[39] Diallo A. Deschasaux M. Latino-Mertel P et al. Red and processed meat intake and cancer risk: results from prospective NutriNet-Santè cohort study. Int J Cancer 2018; 142 (2): 230-237.

[40] Farvid MS. Sidahmed E. Spence ND et al. Consumption of red meat and processed meat and cancer incidence: a systematic review and meta-analysis of prospective studies. Eur J Epidemiol 2021; 36 (9): 937-951.

[41] Huang Y. Cao D. Chen Z et al. Red processed meat consumption and cancer outcome: umbrella review. Food Chem 2021; 356: 129697; doi10.1016/j.foodchem.2021.129697.

APPENDIX

Minimum Spanning Tree

A minimum spanning tree (MST) is the spanning tree of a connected undirected graph. It connects all the vertices and minimizes the total weight of its edges.

The MST algorithm, described by the Czech scientist Otakar Boruvka in 1926, originally aimed to optimize the planning of electricity connections between cities. It was later refined by Kruskal with a specific deterministic algorithm.

The main advantage of the MST algorithm is that it provides an overview of the variables together and makes it very easy to understand clustering through links that directly connect variables that are very close to each other. The importance of the variables in the graph is related to their number of links. Hubs may be defined as the variables with the maximum number of connections in the graph. The clustering distance between two variables is related to their degree of separation.

The MST algorithm does not map all the associations found in the data, but creates only the most parsimonious pattern of multi-dimensional similarities between the nodes in the system that still ensures that all nodes are spanned, i.e., that all the nodes are connected to at least one other node, and that there is only a single path between two nodes (no loops). The more similar two nodes are, the smaller the distance between them (the closer they are). In the computation of a MST, each connection between nodes is evaluated against all the other connections between all the nodes in the graph by optimizing the square matrix of distances between the variables. Every link that causes a loop to appear in the graph is eliminated, irrespective of its strength of association, leading to the simplest, most consistent graph-theoretic representation of a phenomenon.

The minimum spanning tree problem is defined as follows: find an acyclic subset T of E that connects all of the vertices in the graph and whose total weight is minimized, where the total weight is given by:

$$d(T) = \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} d_{i,j}, \forall d_{i,j}$$

T is called the spanning tree, and MST is the T which minimizes the sum of edge weights:

$$Mst = Min\{d(T_k)\}$$

Given an undirected graph G, representing a matrix d of distances, with V vertices, completely linked to each other, the total number of their edges (E) is:

$$E = \frac{V \cdot (V-1)}{2}$$

and the number of its possible trees is:

$$T = V^{V-2}$$

Kruskal (1956) discovered an algorithm that determines the MST of any undirected graph in a quadratic number of steps, in the worst case.

Conceptually, an MST represents the energy minimizing state of a structure. In fact, if we consider the atomic elements of a structure as vertices of a graph and the strength between them as the weight of each edge linking a pair of vertices, the MST represents the minimum energy required for all elements of the structure to maintain their cohesion. In a closed system, all the components tend to minimize total energy. So the MST, in specific situations, can represent the most probable state towards which a system tends.

Metric functions used to obtain the distance matrix to filter with MST methods
We used the Euclidean metric function:

The Euclidean distance between variables is easy to generate. It is necessary, first of all, to scale between 0 and 1 and then to transpose the matrix of the assigned dataset.

$$d_{i,j}^{[E]} = \sqrt{\sum_{k=1}^{M}(x_{i,k} - x_{j,k})^2}; \quad i, j \in [1,2,..N]; \quad x \in [0,1].$$

where :

$d_{i,j}^{[E]} =$ Euclidean distance among any couple of variables;

x_i = Value of any Record scaled between 0 and 1;

$N =$ Number of Variables of the assigned dataset;

$M =$ Number of Records of the assigned dataset;

At this point, the assigned dataset is transformed into an undirected weighted graph where MST methods can be applied.