

**Appendix 12:**  
**CLUSTER ANALYSIS FOR GROUPING OF FOOD CATEGORIES**

## APPENDIX 12: CLUSTER ANALYSIS FOR GROUPING OF FOOD CATEGORIES

The results of the uncertainty analysis of the risk assessment were summarized by a cluster analysis of food categories. The similarity between categories was evaluated for the predicted number of cases of listeriosis expressed as the risk per serving and per annum. Cluster analysis is a descriptive statistical technique by which a set of objects are partitioned or classified into subsets according to some measure of similarity between objects<sup>1</sup>. Typically, this partitioning is defined to generate hierarchical subsets of the objects to be classified. A single level of disjoint partitioning, without any sub-partitioning of the objects within the primary clusters, is a special case of the more general objective of obtaining a hierarchical classification.

The use of a cluster analysis to summarize the results of the *L. monocytogenes* risk assessment provides a means to convey the implications of the uncertainty analysis of the rankings of food categories which is, in some sense, more informative than statistical point null hypothesis tests of differences in the location of the distribution of ranks across food categories (e.g., as provided by Kruskal-Wallis test or sign test). Testing for differences in location (e.g., the median) of the uncertainty distributions of risk rankings, according to either risk per serving or cases per annum, does not incorporate any consideration of whether or not the differences obtained are meaningful on a practical level.

Although the possibility exists that the elicitation and specification of the variability and uncertainty of the model could result in two or more pairings of food categories with identical distributions for either risk per serving or expected cases per annum, this is very unlikely and small differences in the location of rank distributions are expected. In this event, statistical analysis of the output of the simulation based on use of point null hypothesis tests to define differences between food categories is likely to result in categorizing all such (small) differences as significant (i.e., provided that the output of the simulation is sufficiently large). While composite rather than point null hypotheses could be used to define practical or meaningful differences between the risk rankings of different food categories (e.g., by equivalence testing methods), the application of these methods is not readily available. Consequently, a cluster analysis approach was adopted as an alternative.

Central to any cluster analysis is the specification of a definition of similarity, or conversely dissimilarity, between the objects to be classified<sup>1</sup>. With respect to a cluster analysis of risk ranking of the food categories, the “objects” to be classified are the uncertainty distributions (of risk per serving and expected cases per annum) and thus a classification requires a definition of the “distance” or dissimilarity between any two such distributions. The measure of similarity adopted here for the cluster analysis was defined

---

<sup>1</sup> Jain A.K., Murty M.N. and Flynn P.J. (1999). Data Clustering: A review. *ACM Computing Surveys* 31(3), pg 264-323.

by the degree to which any two uncertainty distributions overlap. If, for two food categories, the uncertainty distribution of their risk rankings were identical then the distributions would overlap maximally and it would be reasonable to infer that they are two food categories that should be judged to be similar in risk ranking. Conversely, if the risk rank distributions of two food categories did not overlap at all then it would be reasonable to infer that they are very dissimilar foods in regard to risk ranking.

Based on this intuitive notion of distance between two distributions the following measure of dissimilarity was used:

$$\text{distance}(A, B) = \Pr(\text{rank}(A) > \text{rank}(B))$$

where A and B denote any two food categories, and rank() denotes their rank distributions (according to either risk per serving or expected number of cases per annum). Thus, if the rank of food category A is higher than that of food category B with a high probability of belief (i.e., according to their uncertainty distributions) then A and B would be considered sufficiently dissimilar to belong in different clusters. A level of 90% probability of belief that the rank of one food category was higher than another was chosen as a cut-off value for classifying any two distributions as dissimilar. That is to say, any two food categories A and B were considered to be of different risk category (or cluster) if:

$$\text{distance}(A, B) > 0.90$$

Obviously, both the definition of distance used and what constitutes a “significant” distance based on the definition are subjective. With respect to the latter, this is not intrinsically different from the specification of confidence levels in frequentist-based hypothesis testing. A level of 0.05 is common by convention but it is a subjective choice nonetheless and other significance levels can and often have been advocated. With respect to the former, we note that the chosen measure of distance is not the only one that could be made. Also, it is a pseudo-distance measure because it does not satisfy all properties of distance measure proper; specifically it is not a symmetric function of the argument. However, other more sophisticated information-theoretic measures of the distance between two distributions such as the Kullback-Leibler divergence are computationally difficult and also do not satisfy all of the properties of a distance measure per se (i.e., they are quasi- or pseudo-distances).

Given the chosen definition of distance between two distributions and the cut-off probability value for significant distance, all food categories were compared in a pairwise fashion. Based on these comparisons a partitioning of the food categories into disjoint subsets of similar risk (either by risk per serving or cases per annum) was obtained by defining clusters in the ordering of food categories from highest median rank to lowest median rank. Specifically, the food categories were ranked according their median rank and then partitions were formed by taking the first cluster as being the largest set of ordered food categories (starting from the first) for which all pairwise comparisons of food categories within the set were equivalent based on the definition of

significant distance between their respective uncertainty distributions. This process was repeated with all of the remaining food categories until each food category was assigned to one (and only one) cluster. If, for any given food category, there was no other food category that was similar, based on the definition, then that single food category was taken to form a cluster of one.

The results of the calculations of dissimilarity (or distance) between the twenty-three food categories are shown in Tables A12-1 and A12-2 based on the simulation output of the uncertainty distributions of mean risk per serving and expected number of cases per annum, respectively ( $n = 4,000$  uncertainty samples or iterations). Based on these calculations the results of clustering the food categories according to either per serving risk or cases per annum are shown in Table A13-3. The sensitivity of the results to different specification of cut-off values for belief that one food category ranks higher than another, and is therefore dissimilar, is shown in Table A12-4. A level of 90% probability was chosen here as a reasonable summarization in order to obtain a relatively small number of clusters. At the 90% cut-off value there is a high degree of belief that, based on the uncertainty distributions, the foods in one cluster are of appreciably higher risk than those foods in any lower ranked cluster. While there are differences in risk rankings of food categories within any given cluster we are not “confident at a 90% level” that the differences are practically significant given all the attendant uncertainties that have been incorporated into the assessment.

**Table A12-1. Probabilities<sup>1</sup> (over uncertainty) that food categories rank higher (or lower) than other food categories based on the mean risk per serving.**

	DM	FNR	P	UM	SS	CR	HFD	SUC	PM	FSC	FR	PF	RS	F	DFS	SSC	SRC	V	DS	IC	PC	CD	HC	
DM	0.0%	50.6%	65.8%	84.9%	82.8%	94.7%	98.5%	95.1%	97.4%	100.0%	100.0%	99.1%	100.0%	95.8%	99.6%	99.9%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FNR	49.5%	0.0%	71.8%	86.5%	84.7%	96.3%	98.3%	96.8%	97.8%	99.9%	100.0%	99.1%	100.0%	96.1%	99.8%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
P	34.2%	28.2%	0.0%	77.4%	76.6%	90.1%	96.0%	91.0%	96.0%	99.9%	100.0%	98.7%	100.0%	93.1%	99.2%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UM	15.1%	13.6%	22.6%	0.0%	49.9%	56.0%	68.9%	69.2%	80.1%	92.0%	94.9%	91.0%	96.5%	84.4%	92.2%	97.0%	95.3%	98.4%	98.3%	100.0%	99.9%	99.5%	99.8%	99.8%
SS	17.3%	15.3%	23.5%	50.2%	0.0%	52.8%	66.6%	69.1%	81.7%	95.1%	99.8%	92.0%	99.5%	84.8%	93.6%	98.8%	97.0%	100.0%	99.3%	100.0%	100.0%	100.0%	100.0%	100.0%
CR	5.4%	3.7%	10.0%	44.0%	47.2%	0.0%	71.0%	68.0%	84.6%	97.6%	99.8%	93.3%	99.7%	83.5%	94.4%	99.4%	97.3%	100.0%	99.6%	100.0%	100.0%	100.0%	100.0%	100.0%
HFD	1.5%	1.8%	4.0%	31.2%	33.5%	29.0%	0.0%	57.9%	74.9%	94.6%	98.8%	87.9%	99.3%	79.5%	91.1%	98.3%	95.6%	99.9%	99.1%	100.0%	100.0%	99.9%	100.0%	100.0%
SUC	4.9%	3.2%	9.0%	30.8%	30.9%	32.0%	42.1%	0.0%	57.2%	78.1%	85.2%	78.5%	87.3%	73.4%	81.6%	87.9%	84.8%	90.7%	91.7%	97.2%	97.0%	96.7%	98.3%	98.3%
PM	2.6%	2.2%	4.0%	19.9%	18.4%	15.5%	25.1%	42.8%	0.0%	80.0%	93.2%	78.3%	96.6%	75.0%	84.1%	95.8%	88.9%	99.2%	97.0%	99.8%	100.0%	99.4%	100.0%	100.0%
FSC	0.0%	0.2%	0.1%	8.1%	5.0%	2.5%	5.4%	21.9%	20.0%	0.0%	66.6%	63.1%	78.2%	61.8%	68.4%	83.2%	75.1%	88.6%	89.2%	98.0%	98.4%	95.7%	98.5%	98.5%
FR	0.0%	0.0%	0.0%	5.2%	0.2%	0.3%	1.2%	14.9%	6.8%	33.4%	0.0%	57.4%	77.1%	58.0%	62.8%	82.3%	70.2%	86.5%	86.6%	99.2%	99.5%	96.0%	99.0%	99.0%
PF	1.0%	0.9%	1.4%	9.1%	8.0%	6.7%	12.1%	21.5%	21.7%	36.9%	42.6%	0.0%	50.6%	48.4%	50.7%	57.5%	57.6%	62.6%	69.8%	84.5%	83.7%	83.3%	91.1%	91.1%
RS	0.0%	0.0%	0.0%	3.6%	0.5%	0.4%	0.7%	12.7%	3.4%	21.8%	23.0%	49.5%	0.0%	50.3%	51.8%	65.9%	60.2%	73.1%	78.0%	96.8%	97.8%	91.7%	96.7%	96.7%
F	4.2%	3.9%	6.9%	15.6%	15.2%	16.6%	20.5%	26.6%	25.0%	38.2%	42.0%	51.6%	49.7%	0.0%	50.5%	57.2%	57.8%	60.9%	69.5%	84.5%	84.3%	83.2%	91.3%	91.3%
DFS	0.4%	0.2%	0.8%	7.8%	6.4%	5.6%	8.9%	18.4%	15.9%	31.6%	37.2%	49.3%	48.2%	49.5%	0.0%	58.1%	58.3%	64.4%	71.8%	88.8%	89.1%	85.8%	92.7%	92.7%
SSC	0.1%	0.0%	0.0%	3.0%	1.2%	0.6%	1.7%	12.1%	4.2%	16.9%	17.7%	42.5%	34.2%	42.9%	42.0%	0.0%	50.5%	58.5%	69.0%	89.7%	90.4%	84.7%	92.3%	92.3%
SRC	0.1%	0.1%	0.1%	4.7%	3.0%	2.8%	4.5%	15.2%	11.2%	24.9%	29.8%	42.4%	39.8%	42.2%	41.7%	49.5%	0.0%	55.3%	63.1%	80.7%	80.9%	79.3%	88.5%	88.5%
V	0.0%	0.0%	0.0%	1.6%	0.0%	0.0%	0.1%	9.3%	0.9%	11.4%	13.5%	37.4%	26.9%	39.1%	35.7%	41.5%	44.7%	0.0%	63.0%	86.1%	85.8%	81.1%	91.8%	91.8%
DS	0.0%	0.0%	0.1%	1.8%	0.7%	0.5%	0.9%	8.3%	3.1%	10.9%	13.4%	30.2%	22.0%	30.6%	28.2%	31.0%	36.9%	37.0%	0.0%	72.7%	72.3%	71.2%	85.1%	85.1%
IC	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	2.9%	0.2%	2.0%	0.8%	15.5%	3.3%	15.5%	11.3%	10.4%	19.3%	14.0%	27.4%	0.0%	50.9%	53.0%	70.5%	70.5%
PC	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	3.1%	0.0%	1.6%	0.5%	16.3%	2.2%	15.7%	11.0%	9.6%	19.1%	14.2%	27.8%	49.1%	0.0%	51.9%	69.4%	69.4%
CD	0.0%	0.0%	0.0%	0.6%	0.1%	0.1%	0.2%	3.3%	0.6%	4.3%	4.0%	16.7%	8.3%	16.8%	14.2%	15.3%	20.8%	18.9%	28.8%	47.1%	48.2%	0.0%	65.9%	65.9%
HC	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	1.7%	0.1%	1.5%	1.1%	8.9%	3.3%	8.7%	7.3%	7.7%	11.5%	8.2%	14.9%	29.5%	30.7%	34.1%	0.0%	0.0%

<sup>1</sup> Probabilities are defined as Prob(rank(A) > rank(B)) where A is the food category identified in the row labels and B is the food category identified in the column labels (based on 4,000 uncertainty iterations of the model).

<i>LEGEND</i>	
DM = Deli Meats	RS = Raw Seafood
FNR = Frankfurters (not reheated)	F = Fruits
P = Pâté and Meat Spreads	DFS = Dry/Semi-Dry Fermented Sausages
UM = Unpasteurized Fluid Milk	SSC = Semi-soft Cheese
SS = Smoked Seafood	SRC = Soft Ripened Cheese
CR = Cooked Ready-To-Eat Crustaceans	V = Vegetables
HFD = High Fat and Other Dairy Products	DS = Deli-type Salads
SUC = Soft Unripened Cheese	IC = Ice Cream and Frozen Dairy Products
PM = Pasteurized Fluid Milk	PC = Processed Cheese
FSC = Fresh Soft Cheese	CD = Cultured Milk Products
FR = Frankfurters (reheated)	HC = Hard Cheese
PF = Preserved Fish	

**Table A12-2. Probabilities<sup>1</sup> (over uncertainty) that food categories rank higher (or lower) than other food categories based on the number of cases per annum.**

	DM	PM	HFD	FNR	SUC	P	CR	UM	SS	F	FR	V	DFS	FSC	SSC	SRC	DS	RS	PF	IC	PC	CD	HC
DM	0.0%	91.9%	98.5%	99.6%	99.8%	100.0%	100.0%	99.8%	99.6%	92.4%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PM	8.1%	0.0%	60.3%	75.1%	83.8%	96.0%	98.0%	93.8%	94.5%	77.9%	100.0%	99.2%	99.1%	100.0%	99.9%	99.6%	99.5%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%
HFD	1.5%	39.7%	0.0%	69.7%	80.4%	95.3%	97.9%	93.1%	94.0%	75.4%	99.8%	99.1%	99.0%	100.0%	99.8%	99.8%	99.4%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FNR	0.4%	24.9%	30.3%	0.0%	70.2%	92.0%	95.8%	87.2%	91.0%	72.5%	99.3%	98.5%	98.0%	100.0%	99.7%	99.6%	98.9%	100.0%	99.8%	100.0%	100.0%	100.0%	100.0%
SUC	0.2%	16.2%	19.7%	29.9%	0.0%	60.6%	64.9%	60.8%	69.8%	59.5%	83.8%	78.8%	85.8%	91.1%	90.3%	88.4%	88.2%	92.8%	92.9%	95.6%	95.7%	96.0%	97.7%
P	0.1%	4.0%	4.7%	8.0%	39.4%	0.0%	57.2%	54.0%	69.8%	59.6%	94.0%	82.2%	90.1%	100.0%	98.2%	93.6%	93.2%	100.0%	98.9%	99.7%	100.0%	99.6%	100.0%
CR	0.0%	2.0%	2.2%	4.2%	35.1%	42.8%	0.0%	48.7%	65.8%	57.0%	91.8%	79.3%	89.2%	99.9%	98.0%	92.2%	92.5%	100.0%	98.5%	99.6%	100.0%	99.2%	99.9%
UM	0.3%	6.2%	6.9%	12.8%	39.2%	46.0%	51.3%	0.0%	61.3%	56.5%	80.5%	76.2%	85.7%	96.6%	93.9%	89.3%	89.4%	97.7%	95.3%	98.4%	98.7%	97.4%	99.0%
SS	0.4%	5.5%	6.0%	9.0%	30.2%	30.2%	34.3%	38.7%	0.0%	54.0%	72.7%	71.0%	82.2%	98.7%	94.7%	87.0%	88.8%	99.5%	95.4%	99.5%	99.8%	97.6%	99.3%
F	7.6%	22.1%	24.7%	27.6%	40.5%	40.5%	43.0%	43.5%	46.0%	0.0%	54.7%	57.6%	69.7%	75.9%	74.6%	74.5%	76.8%	82.3%	79.9%	89.7%	89.4%	89.1%	94.2%
FR	0.0%	0.1%	0.2%	0.7%	16.2%	6.0%	8.2%	19.5%	27.3%	45.4%	0.0%	58.5%	75.2%	89.8%	88.9%	78.4%	82.4%	98.4%	88.4%	98.7%	99.3%	95.7%	98.5%
V	0.0%	0.9%	0.9%	1.6%	21.3%	17.8%	20.7%	23.8%	29.0%	42.5%	41.6%	0.0%	66.4%	78.6%	77.6%	72.9%	76.7%	89.5%	81.2%	92.8%	94.7%	91.0%	96.4%
DFS	0.0%	1.0%	1.0%	2.0%	14.2%	9.9%	10.8%	14.4%	17.9%	30.3%	24.9%	33.7%	0.0%	59.2%	58.0%	57.9%	58.9%	67.6%	66.2%	78.2%	79.6%	79.7%	88.3%
FSC	0.0%	0.0%	0.0%	0.0%	8.9%	0.0%	0.1%	3.4%	1.3%	24.1%	10.2%	21.5%	40.8%	0.0%	50.1%	50.1%	51.4%	67.1%	60.1%	76.2%	78.7%	77.1%	87.2%
SSC	0.0%	0.2%	0.2%	0.3%	9.7%	1.8%	2.0%	6.2%	5.3%	25.4%	11.1%	22.4%	42.0%	49.9%	0.0%	50.1%	53.1%	66.7%	60.4%	77.5%	78.5%	76.8%	86.8%
SRC	0.0%	0.4%	0.2%	0.4%	11.6%	6.4%	7.8%	10.7%	13.1%	25.6%	21.6%	27.1%	42.1%	49.9%	49.9%	0.0%	50.6%	58.6%	60.1%	69.0%	70.5%	72.6%	81.3%
DS	0.0%	0.5%	0.6%	1.1%	11.8%	6.8%	7.5%	10.6%	11.2%	23.2%	17.6%	23.3%	41.1%	48.6%	47.0%	49.4%	0.0%	53.3%	58.8%	71.9%	72.8%	74.5%	86.3%
RS	0.0%	0.0%	0.0%	0.0%	7.2%	0.0%	0.0%	2.3%	0.5%	17.7%	1.6%	10.5%	32.5%	33.0%	33.3%	41.5%	46.7%	0.0%	52.6%	69.8%	72.0%	72.8%	84.9%
PF	0.0%	0.1%	0.0%	0.2%	7.1%	1.1%	1.5%	4.7%	4.6%	20.1%	11.7%	18.8%	33.8%	39.9%	39.6%	39.9%	41.2%	47.5%	0.0%	59.0%	60.7%	62.0%	71.4%
IC	0.0%	0.0%	0.0%	0.1%	4.4%	0.3%	0.4%	1.7%	0.5%	10.3%	1.3%	7.2%	21.8%	23.8%	22.5%	31.0%	28.2%	30.2%	41.1%	0.0%	53.0%	59.8%	74.9%
PC	0.0%	0.0%	0.0%	0.0%	4.3%	0.0%	0.0%	1.3%	0.2%	10.6%	0.7%	5.3%	20.4%	21.3%	21.5%	29.5%	27.3%	28.0%	39.3%	47.0%	0.0%	57.0%	72.6%
CD	0.0%	0.0%	0.0%	0.0%	4.0%	0.4%	0.9%	2.7%	2.4%	10.9%	4.3%	9.0%	20.3%	22.9%	23.2%	27.4%	25.6%	27.2%	38.1%	40.2%	43.0%	0.0%	62.0%
HC	0.0%	0.0%	0.0%	0.0%	2.3%	0.0%	0.2%	1.1%	0.7%	5.8%	1.6%	3.6%	11.7%	12.9%	13.3%	18.7%	13.8%	15.1%	28.6%	25.1%	27.5%	38.0%	0.0%

<sup>1</sup> Probabilities are defined as Prob(rank(A) > rank(B)) where A is the food category identified in the row labels and B is the food category identified in the column labels (based on 4,000 uncertainty iterations of the model).

<i>LEGEND</i>	
DM = Deli Meats	DFS = Dry/Semi-Dry Fermented Sausages
PM = Pasteurized Fluid Milk	FSC = Fresh Soft Cheese
HFD = High Fat and Other Dairy Products	SSC = Semi-soft Cheese
FNR = Frankfurters (not reheated)	SRC = Soft Ripened Cheese
SUC = Soft Unripened Cheese	DS = Deli-type Salads
P = Pâté and Meat Spreads	RS = Raw Seafood
CR = Cooked Ready-To-Eat Crustaceans	PF = Preserved Fish
UM = Unpasteurized Fluid Milk	IC = Ice Cream and Frozen Dairy Products
SS = Smoked Seafood	PC = Processed Cheese
F = Fruits	CD = Cultured Milk Products
FR = Frankfurters (reheated)	HC = Hard Cheese
V = Vegetables	

Table A12-3. Clustering of Similar Food Categories Based on the Uncertainty Distribution of Relative Risk Ranking on Per Serving and Per Annum Basis.

Cluster	Risk per Serving	Risk per Annum
Cluster 1	<ul style="list-style-type: none"> <li>Deli Meats</li> <li>Frankfurters, not reheated</li> <li>Pâté and Meat Spreads</li> <li>Unpasteurized Fluid Milk</li> <li>Smoked Seafood</li> </ul>	Deli Meats
Cluster 2	<ul style="list-style-type: none"> <li>Cooked RTE Crustaceans</li> <li>High Fat and Other Dairy Products</li> <li>Pasteurized Fluid Milk</li> <li>Soft Unripened Cheese</li> </ul>	<ul style="list-style-type: none"> <li>High Fat and Other Dairy Products</li> <li>Frankfurters, not reheated</li> <li>Pasteurized Fluid Milk</li> <li>Soft Unripened Cheese</li> </ul>
Cluster 3	<ul style="list-style-type: none"> <li>Deli-type Salads</li> <li>Dry/Semi-dry Fermented Sausages</li> <li>Fresh Soft Cheese</li> <li>Frankfurters, reheated</li> <li>Fruits</li> <li>Preserved Fish</li> <li>Raw Seafood</li> <li>Semi-soft Cheese</li> <li>Soft Ripened Cheese</li> <li>Vegetables</li> </ul>	<ul style="list-style-type: none"> <li>Cooked RTE Crustaceans</li> <li>Fruits</li> <li>Pâté and Meat Spreads</li> <li>Unpasteurized Fluid Milk</li> <li>Smoked Seafood</li> </ul>
Cluster 4	<ul style="list-style-type: none"> <li>Cultured Milk Products</li> <li>Ice Cream and Frozen Dairy Products</li> <li>Processed Cheese</li> <li>Hard Cheese</li> </ul>	<ul style="list-style-type: none"> <li>Deli-type Salads</li> <li>Dry/Semi-dry Fermented Sausages</li> <li>Frankfurters, reheated</li> <li>Fresh Soft Cheese</li> <li>Semi-Soft Cheese</li> <li>Soft Ripened Cheese</li> <li>Vegetables</li> </ul>
Cluster 5	Not Applicable	<ul style="list-style-type: none"> <li>Cultured Milk Products</li> <li>Hard Cheese</li> <li>Ice Cream and Frozen Dairy Products</li> <li>Preserved Fish</li> <li>Processed Cheese</li> <li>Raw Seafood</li> </ul>

Table A12-4. Sensitivity of clustering procedure to the cut-off probability used to define similar versus dissimilar food categories.

Measure for ranking	Cut-off probability (distance) for defining any two categories as dissimilar	Total # of pairwise comparisons for which food categories are not judged dissimilar <sup>1</sup>	# of distinct disjoint clusters <sup>2</sup> of similarly ranked food categories
Risk per serving	0.95	139	4
Risk per serving	0.90	116	4
Risk per serving	0.75	61	7
Cases per annum	0.95	149	4
Cases per annum	0.90	124	5
Cases per annum	0.75	69	7

<sup>1</sup>There are a total of 276 pairwise comparisons of 23 food types; two food categories were considered dissimilar if  $\Pr(\text{rank}(A) > \text{rank}(B)) >$  the cut-off probability value where A is the food with higher mean rank and B is the food with lower mean rank

<sup>2</sup>A cluster is defined here as a collection of food categories for which  $\Pr(\text{rank}(A) > \text{rank}(B)) <$  cut-off probability value for any pair (A,B) in the cluster; each food is assigned to only one cluster and therefore clusters are disjoint.